УДК 519.237.5

ПОСТРОЕНИЕ МОДЕЛЕЙ МАССОВОЙ ОЦЕНКИ ОБЪЕКТОВ НЕДВИЖИМОСТИ С УЧЕТОМ ПРОСТРАНСТВЕННОЙ ЗАВИСИМОСТИ

А.В. Беляева, Е.А. Гребенюк

Предложен подход к построению массовой оценки объектов недвижимости, учитывающий основные факторы местоположения объектов, влияющие на цену: расстояние до центров влияния и близость расстояний между объектами. Построены модели, реализующие этот подход, и исследованы их свойства. Показано, что совместный учет обоих факторов значимо улучшает качество модели.

Ключевые слова: пространственная корреляция, тест Морана, тест множителей Лагранжа, пространственная авторегрессионная модель, центр влияния.

ВВЕДЕНИЕ

Компьютерные методы массовой оценки объектов недвижимости активно развиваются по пути учета местоположения объектов. Среди множества факторов, влияющих на цену недвижимости (площадь квартиры, число комнат, категория дома, площадь участка и пр.) один из главных — местоположение объектов. В разработанной Дж.К. Эккертом [1] методологии, в соответствии с которой в настоящее время оценивается стоимость объектов недвижимости во многих странах, местоположение учитывается посредством выделения в пространстве «центров влияния», по мере приближения к которым цена недвижимости возрастает (центры позитивного влияния) или убывает (центры негативного влияния). При оценивании стоимости объекта расстояние до центров влияния вводится как экзогенная переменная в (линейную либо нелинейную) регрессионную модель.

Другое активно развивающееся направление учета местоположения заключается в применении пространственных корреляционных моделей [2, 3]. В их основе лежит гипотеза, что характеристики близко расположенных объектов модели связаны между собой, причем эта связь ослабевает с увеличением расстояния между объектами. В моделях с пространственной корреляцией учитывается расстояние между объектами, но не отражаются такие воздействия на цену, как наличие поблизости загрязняющих окружающую среду объектов, шумных магистралей, кладбищ и пр., а также лесопарковых зон, центров коммуникаций и пр. Центр влияния не полностью определяет влияние местоположения объекта на цену, поскольку на моделируемом участке может присутствовать ряд центров влияния, расположенных в разных точках и различно влияющих на объект, и не все из них могут быть идентифицированы. Поэтому учет местоположения только путем выделения центров влияния может оказаться не полным.

В настоящей статье предложен подход, позволяющий совокупно оценивать влияние обоих факторов местоположения объекта: малое различие цен близких между собой объектов и зависимость цен от расстояния от объектов до центров влияния. Показано, что предлагаемый подход позволяет улучшить качество оценивания объектов. Структура моделей совпадает со структурой пространственных регрессионных моделей, рассмотренных в работах [2, 4]. Для методологии построения этих моделей характерны свои особенности как при выборе конкретной модели оценивания, так и при диагностике построенных моделей и выборе наиболее подходящего варианта. Рассмотрены основные этапы построения моделей, спецификации и выбора варианта, приведены результаты моделирования.

1. ПОСТРОЕНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ МАССОВОЙ ОЦЕНКИ НЕДВИЖИМОСТИ С УЧЕТОМ ФАКТОРА МЕСТОПОЛОЖЕНИЯ ОБЪЕКТА

1.1. Постановка задачи

Имеется набор объектов недвижимости $Z = (z_1, ..., z_n)$, представляющий собой выборку из объектов жилого массива (города, района, поселка). Каждый объект определяется вектором параметров: ценой, площадью, числом комнат, категорией дома, площадью участка и пр., а также координатами. Требуется построить регрессионную модель, позволяющую выполнять компьютеризированную массовую оценку цен объектов с учетом фактора местоположения объекта.

Здесь рассматриваются линейные регрессионные модели. Поскольку выбор независимых регрессоров до некоторой степени ограничен имеющейся в распоряжении статистика информацией, то повысить эффективность модели при построении массовой оценки можно путем оптимального учета информации о зависимости цены от расположения объекта в пространстве. Поэтому в качестве критериев качества модели выбираются:

— среднеквадратичная ошибка оценивания (сумма квадратов отклонений модельных значений цен от фактических);

— эффективность учета местоположения объекта, определяемая сравнением моделей, построенных с учетом пространственного фактора, со стандартными регрессионными моделями.

Местоположение объекта будем учитывать двумя способами: учитывать влияние на цену расстояния до центров влияния и, в случае наличия пространственной зависимости между объектами, учитывать влияние на цену объекта характеристик близко расположенных объектов.

Центр влияния — это объект, не принадлежащий множеству оцениваемых объектов Z и оказывающий позитивное или негативное влияние на стоимость объектов оценивания. К таким центрам могут относиться крупный деловой район города, лесопарк, торговый центр, близость к которым повышает привлекательность объекта недвижимости (центр положительного влияния) или промышленный район с неблагоприятной экологической обстановкой (центр отрицательного влияния).

Идея, лежащая в основе учета пространственной зависимости, заключается в том, что при оценивании характеристик объекта учитываются значения этих же характеристик близко расположенных объектов. Для учета этой зависимости используются пространственные модели [2]. При их построении стандартные процедуры регрессионного анализа имеют свою специфику. Выбор структуры модели, помимо определения набора регрессоров, включает в себя выбор типа модели (SAR, SEM, GSM) и выбор пространственных матриц. При оценивании пространственных моделей обычным методом наименьших квадратов (МНК) получают смещенные и несостоятельные оценки, поэтому для оценивания применяются методы максимального правдоподобия, инструментальной переменной (IV) и обобщенный метод моментов (GMM) [3, 5, 6].

Построение модели с учетом пространственного фактора осуществляется:

— определением центров влияния;

— проверкой наличия пространственной корреляции;

— оцениванием различных вариантов моделей и их диагностикой;

— сравнением вариантов и выбором модели.

1.2. Учет расстояний до центров влияния

Предлагаемая процедура учета расстояний до центров влияния несколько отличается от процедуры, используемой Эккертом [1], и состоит из следующих шагов.

1. Если известны координаты центра влияния, то вычисляется вектор расстояний от центра влияния до каждого из объектов множества Z.

2. Если эта информация отсутствует, то координаты центра влияния определяются по выборочным данным. Стандартным МНК строится регрессия цен объекта на множество независимых переменных (координаты не включаются). Анализируется зависимость остатков регрессии от координат: строится трехмерный график зависимости остатков регрессионной модели от координат x и y и контурный график уровней остатков res. На рис. 1 и 2 приведены примеры построения трех-

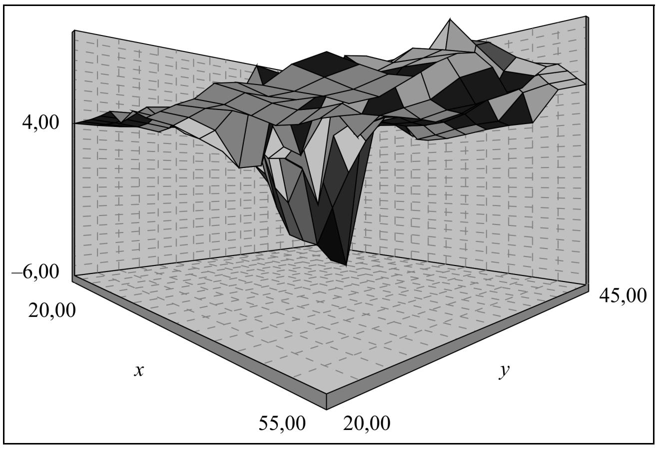

Рис. 1. Трехмерный график остатков

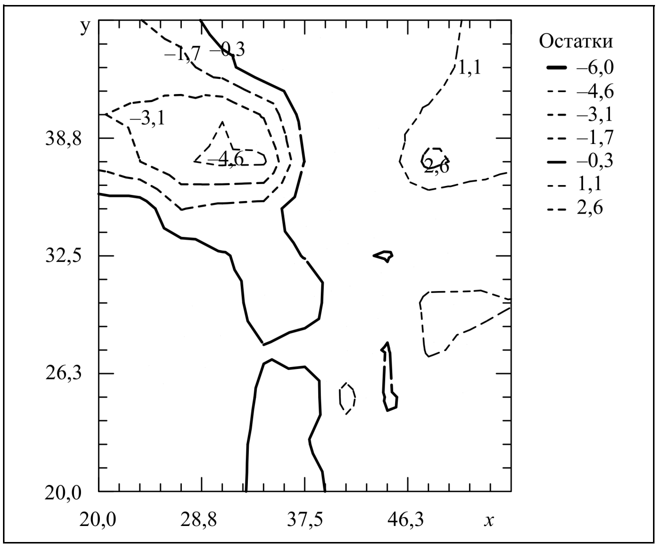

Рис. 2. Плоскость уровней остатков

мерного графика остатков и контурного графика его уровней[1]. С помощью анализа пространственного графика и контурного графика выделяются участки, на которых остатки значительно отличаются от нуля. Координаты центров этих участков определяются как координаты центров влияния.

3. Для нахождения значений регрессора, учитывающего зависимость цен от расстояния до центров влияния, строится диаграмма рассеяния для переменных *res* и *d*, где *res* — остатки регрессионной модели: разность фактических цен объектов Y и цен Y^M, рассчитанных по регрессионной модели, не учитывающей пространственный фактор, d — вектор расстояний объектов до центров влияния. С помощью диаграммы выбирается вид функциональной зависимости $res = F = f(ad) + b$, где a и b — параметры функции, подбираемые методом наименьших квадратов. Вектор значений F добавляется в регрессионную модель в качестве регрессора. Если регрессор при оценивании модели оказывается значимым и не увеличивает величины критериев Акаике и Шварца, то он вводится в модель.

Предложенная процедура позволяет учитывать фактор центров влияния при оценивании объектов, не входящих в обучаемую выборку для построения модели.

1.3. Описание класса пространственных моделей

Если местоположение не имеет значения для рассматриваемой группы объектов, то их цена может быть описана стандартной регрессионной моделью:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \qquad (1)$$

где Y — n-мерный вектор цен объектов X — $(n\times k)$-матрица объясняющих переменных, β — k-мерный вектор коэффициентов, ε — n-мерный вектор возмущений, σ^2 — дисперсия возмущений, n — число оцениваемых объектов.

Если местоположение объекта важно для анализа цены объекта, то для описания этой зависимости применяют различные способы «взвешивания» цен: $Y = \rho WY$, где Y — n-мерный вектор характеристик объектов ρ — константа, W — $(n\times n)$-матрица пространственной зависимости: элемент w_{ij} принимает не нулевое значение, если объекты i и j близкие в смысле выбранной метрики, и нулевые — в противном случае. Для улучшения свойств оценок используют умножение элементов матрицы на различные весовые коэффициенты с последующей стандартизацией. Стандартизация матрицы выполняется таким образом, чтобы сумма элементов в строке равнялась единице. Будем рассматривать следующие типы пространственных моделей [2].

- Модель с пространственной корреляцией в объясняемой переменной (spatial autoregressive — SAR-модель):

$$Y = \rho Wy + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \qquad (2)$$

где ρ — коэффициент пространственной авторегрессии, W — пространственная матрица.

- Модель с пространственной автокорреляцией в возмущениях (spatial error model — SEM-модель):

$$Y = X\beta + u,\ u = \lambda Wu + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \quad (3)$$

где u — n-мерный вектор возмущений, λ — коэффициент пространственной корреляции возмущений, описывающий пространственную корреляцию в ошибках модели.

- Общая модель, включающая в себя пространственные корреляции в объясняемых переменных и в возмущениях (general spatial model — GSM-модель):

$$Y = \rho W_1 Y + X\beta + u,\ u = \lambda W_2 u + \varepsilon,$$
$$\varepsilon \sim N(0, \sigma^2 I_n). \qquad (4)$$

Пространственные матрицы W_1 и W_2 могут как различаться, так и быть равны между собой, причем в последнем случае могут возникнуть проблемы при оценивании модели.

При построении пространственных матриц, используемых в моделях (2) — (4), учитывается принцип: близко расположенные объекты имеют близ-

[1] Графики построены с помощью пакета NCSS (Number Cruncher Statistical System).

кие характеристики. *Алгоритм построения матрицы W.*

1. Элемент w_{ij} пространственной матрицы W полагается равным единице, если объекты i и j являются близкими в смысле выбранной метрики, и нулю в противном случае.

Один из наиболее распространенных алгоритмов построения пространственных матриц заключается в определении близких объектов на основе триангуляции Делоне. Триангуляция набора точек на плоскости состоит в соединении всех точек непересекающимися отрезками таким образом, чтобы новых отрезков уже нельзя было добавить без пересечения с имеющимися. Если в каждой точке плоскости (X, Y) задана координата Z, то триангуляция Делоне обеспечивает наилучшую аппроксимацию поверхности (X, Y, Z). Объекты являются близкими к объекту i, если они соединены с ним отрезками в результате триангуляции Делоне.

Другой способ формирования пространственных матриц состоит в выделении для каждой из точек плоскости (X, Y) ее ближайших в смысле евклидовой метрики точек.

2. Значения расстояний между близкими точками учитываются заменой единиц в пространственной матрице весами: чем ближе расположены точки i и j, тем больше вес соответствующего элемента матрицы. Примеры назначения весов как функций расстояния d_{ij}:

$$w_{ij} = \begin{cases} d_{ij}^{-\alpha}, \text{ если } w_{ij} \neq 0, \\ 0, \text{ если } w_{ij} = 0, \end{cases}$$

или

$$w_{ij} = \begin{cases} \exp(-\alpha d_{ij}), \text{ если } w_{ij} \neq 0, \alpha > 0, \\ 0, \text{ если } w_{ij} = 0. \end{cases}$$

3. После выбора весов матрицу стандартизируют таким образом, чтобы сумма весов в каждой строке равнялась единице.

1.4. Проверка наличия пространственной автокорреляции

Для обоснования перехода от стандартной регрессионной модели к моделям с пространственной корреляцией и определения их вида разработано большое число тестов [2, 7, 8].

Тест Морана — один из широко применяемых. Он предназначен для проверки наличия пространственной зависимости, определяемой матрицей W. Статистика теста рассчитывается по формуле

$$I = (N/S)\frac{(Y-\bar{Y})W(Y-\bar{Y})}{(Y-\bar{Y})(Y-\bar{Y})},$$

где $\bar{Y}$ — среднее значение Y, N — размерность вектора Y, W — матрица пространственной зависимости, $S = \sum_i \sum_j w_{ij}$ — сумма элементов матрицы W. Если матрица W стандартизирована таким образом, что сумма элементов каждой строки равняется единице, то статистика принимает вид:

$$I = \frac{(Y-\bar{Y})W(Y-\bar{Y})}{(Y-\bar{Y})(Y-\bar{Y})}.$$

В работе [3] обобщена статистика Морана и разработан тест для проверки пространственной корреляции для линейной регрессионной модели (2). Показано, что статистика

$$z = \frac{I - E(I)}{\sqrt{\mathrm{Var}(I)}}$$

при выполнении нулевой гипотезы H_0: «пространственная корреляция отсутствует» имеет стандартное нормальное распределение. Значения статистики зависят от пространственной матрицы W: при различных моделях пространственных матриц тест Морана может давать противоречивые результаты. Этот тест не определяет характер пространственной зависимости, но лишь указывает на ее существование.

Тест множителей Лагранжа (МЛ) [7, 8] при исследовании на наличие пространственной автокорреляции в данных не требует оценки пространственной модели, достаточно оценить модель в условиях нулевой гипотезы H_0: «данные описываются регрессионной моделью (1)». Такие тесты могут также различать пространственную корреляцию в объясняемой переменной и пространственную корреляцию в ошибке модели. Статистика

$$LM_{err} = (ne^TWe/e^Te)^2(\mathrm{tr}(W^TW + W^2))^{-1}, \quad (5)$$

где e — вектор ошибок модели (1), используется для проверки гипотезы H_0 против альтернативной гипотезы H_1: «данные описываются регрессионной моделью (3)».

Статистика МЛ-теста для проверки пространственной корреляции в объясняющей переменной имеет вид:

$$LM_{lag} = (ne^TWe/e^Te)^2(n(WXb)^TM(WXb)/(e^Te) + \\ + \mathrm{tr}(W^TW + W^2))^{-1}, \quad (6)$$

где b — МНК-оценка β в модели (1), $M = I - X(X'X)^{-1}X'$. Эта статистика проверяет гипотезу H_0 против альтернативной гипотезы H_1: «данные описываются регрессионной моделью (2)». Обе статистики (5) и (6) асимптотически распределены как $\xi^2(1)$. Если обе статистики значимы, то возникает ситуация неопределенности, для разрешения которой в работе [7] предложено выбирать ту модель, для которой соответствующая статистика имеет более высокий уровень значимости. Однако эти рекомендации работают только в том случае, если уровни значимости статистик существенно различаются.

1.5. Оценка качества и сравнение вариантов

Одним из основных вопросов остается оценка качества построенной модели. В силу структуры пространственных моделей и применяемых методов оценивания остатки могут не иметь нулевого математического ожидания, следовательно, разложения дисперсии на дисперсию, объясненную регрессией, и дисперсию остатков не существует, поэтому критерий R^2 теряет смысл. Для оценивания пространственных моделей выберем набор критериев:

— критерии Акаике и Шварца — оценивают среднее значение ошибки модели с учетом штрафов за каждый включаемый регрессор;

— квадрат коэффициента корреляции между наблюдаемыми и предсказанными по модели значениями (Corr2);

— максимум функции правдоподобия (МП);

— оценка смещения среднего остатков относительно нуля;

— оценка приведенной суммы квадратов остатков (Std).

2. ОПИСАНИЕ ПРОЦЕССА ПОСТРОЕНИЯ МОДЕЛИ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

Для построения модели по экспериментальным данным рассмотрим пример Анселина [2], в котором определяется влияние пространственного фактора, доходов домохозяйств и стоимости жилья на уровень криминала (число ограблений на тысячу домохозяйств). Этот пример рассматривался многими исследователями для иллюстрации полученных результатов. Все вычисления проводились с помощью стандартных программ пакета MATLAB и набора программ, предназначенного для оценивания пространственных моделей и включающего в себя алгоритмы оценивания и построения пространственных матриц [4].

Определение центров влияния. Уравнение регрессии уровня криминала на размер доходов и стоимость жилья имеет вид:

$$Y = 68{,}619 - 1{,}597x_1 - 0{,}274x_2 + \varepsilon, \qquad (7)$$

где Y — уровень криминала, x_1 — доходности домохозяйств, x_2 — стоимость жилья, ε — ошибка модели. Анализируется 49 административных единиц, каждая из которых, помимо указанных переменных, определяется парой географических координат. В результате построения трехмерного графика отношений фактических значений цены к модельным в пространстве координат и контурного графика уровней отношений, определим координаты центра влияния: $x = 38{,}29$; $y = 30{,}35$. Построим диаграмму рассеяния переменных $d = \{d_i\}$, $i = 1, ..., n$, где d_i — расстояние от i-го объекта до центра влияния, и res — остатков регрессии (7) и определим вид зависимости res от d: $res_i = \exp(-d_i)$.

Построим регрессионную модель, добавив в нее в качестве регрессора вектор $x_3 = \{\exp(-d_i)\}$, $i = 1, ..., n$. Уравнение регрессии уровня криминала на размер доходов, стоимость жилья и расстояние до центра влияния имеет вид:

$$Y = 65{,}76 - 1{,}61x_1 - 0{,}229x_2 + 35{,}92x_3. \qquad (8)$$

Характеристики моделей (7) и (8) представлены в табл. 1, где AIC — критерий Акаике, SIC — критерий Шварца, Corr2 — квадрат коэффициента корреляции между наблюдаемыми и предсказанными по модели значениями, Std — среднеквадратичное отклонение остатков. Проверка по критерию Жака—Бера указывает, что остатки модели (7) соответствуют нормальному распределению: вероятность ошибиться, отвергнув нулевую гипотезу, $p_{value} = 0{,}253$; гипотеза о нормальности остатков модели (8) отвергается, $p_{value} = 0{,}016$. По результатам, представленным в табл. 1, значения критериев в регрессии, построенной с учетом расстояния до центров, выше, чем в обычной. Однако более высокие значения критериев Акаике и Шварца указывают на то, что эти улучшения слишком незначительны по сравнению с платой за улучшение — добавлением новой переменной.

Выбор структуры пространственных матриц. При построении пространственных моделей будем пользоваться матрицами смежности различной структуры: W — матрица, построенная на основе триангуляции Делоне; W_2 и W_4 — матрицы с двумя и четырьмя ближайшими соседями, соответственно. Весовые коэффициенты матриц W, W_2 и W_4 полагаем равными единице и проводим стандартизацию матриц по строкам.

Таблица 1

Сравнение качества моделей

Критерии	R^2	AIC	SIC	Corr2	Std
Модель (7)	0,533	4,996	5,154	0,552	11,194
Модель (8)	0,628	5,059	5,278	0,651	9,885

Проверка наличия пространственной корреляции. Проверим наличие пространственной корреляции, применив тест Морана. Результаты представлены в табл. 2.

Таблица 2

Проверка наличия пространственной корреляции тестом Морана

Модель	Матрица модели	W	W_2	W_4
(7)	z p	4,0149 $5,8341 \cdot 10^{-5}$	3,1125 0,0019	4,7344 $2,1972 \cdot 10^{-6}$
(8)	z p	3,2838 0,001	2,3372 0,0194	4,2225 $2,4165 \cdot 10^{-5}$

где z — значения модифицированной статистики Морана [3], p — уровень значимости. Результаты применения теста Морана указывают на наличие пространственной корреляции (включение информации о центрах влияния, практически, не меняет результатов теста), но тип корреляции не определен.

Для определения типа пространственной корреляции проведем тесты множителей Лагранжа. Проверим наличие пространственной корреляции в остатках регрессионных моделей (7) и (8), используя статистику (5); результаты приведены в табл. 3, где Lmerr — статистика, определяемая выражением (5).

Таблица 3

Проверка наличия пространственной корреляции в остатках тестом МЛ

Модель	Матрица модели	W	W_2	W_4
(7)	Lmerr p	11,073 0,0009	7,506 0,0061	15,903 $6,67 \cdot 10^{-5}$
(8)	Lmerr p	6,468 0,011	3,596 0,058	11,736 $6,132 \cdot 10^{-4}$

Тесты показывают наличие пространственной корреляции в остатках модели, однако добавление нового регрессора несколько повышает вероятность ошибки отвержения нулевой гипотезы об отсутствии корреляции в остатках.

Проверим наличие пространственной корреляции в лагах регрессионной модели с использованием статистики (6); результаты приведены в табл. 4, где Lmlag — статистика, определяемая выражением (6).

Таблица 4

Проверка наличия пространственной корреляции в лагах тестом МЛ

Модель	Матрица модели	W	W_2	W_4
(7)	Lmlag p	12,858 0,0003	13,816 0,0002	18,513 $1,69 \cdot 10^{-5}$
(8)	Lmlag p	10,128 0,0015	10,659 0,0011	14,96 $1,098 \cdot 10^{-4}$

Тесты показывают наличие пространственной корреляции в лагах, добавление нового регрессора мало влияет на статистику и вероятность ошибки.

По результатам тестов пространственная корреляция присутствует как в остатках, так и в лагах модели, причем включение информации о расстояниях до центра влияния не устраняет и, практически, не снижает уровень пространственной корреляции. Таким образом, по результатам тестов на наличие пространственной зависимости вид пространственной корреляции однозначно определить не удается.

Оценивание различных вариантов моделей и их диагностика. В силу неопределенности информации о типе пространственной корреляции для выбора наилучшей модели проверим все возможные варианты и построим пространственные модели (2)—(4). Оценки качества построенных SAR-моделей, учитывающих пространственную корреляцию в объясняемой переменной, приведены в табл. 5.

В первых трех столбцах таблицы приведены характеристики пространственных моделей, построенных без учета центров влияния, в следующих трех столбцах — моделей с добавленным регрессором x_3, в первой строке таблицы указана пространственная матрица, используемая для построения модели. Как видно из данных, представленных в таблице, добавление информации о центрах влияния позволяет улучшить качество модели по всем критериям. Из всех построенных SAR-моделей наилучшими характеристиками обладает модель в 4-м столбце (с матрицей W).

Оценим качество построенных SEM-моделей, учитывающих пространственную корреляцию в остатках. Результаты приведены в табл. 6.

Как и для SAR-модели, в первых трех столбцах таблицы приведены характеристики пространственных моделей, построенных без учета центров влияния, в следующих трех столбцах — моделей с добавленным регрессором x_3. Добавление нового регрессора практически не улучшает качество модели. Из всего набора построенных моделей опять же модель с матрицей W и с дополнительным регрессором обладает лучшими характеристиками, но она уступает модели с SAR-структурой.

Рассмотрим общую пространственную модель, предполагающую наличие пространственной корреляции как в лагах, так и в остатках, для различных вариантов моделей, различающиеся между собой комбинацией пространственных матриц. Так же, как и в предыдущих случаях, сравним два варианта: с учетом (табл. 7) и без учета информации до центров влияния (табл. 8).

Оптимальная модель — с дополнительным регрессором x_3 и комбинацией матриц $W — W_2$ (5-й столбец табл. 8). Ей незначительно уступает модель с SAR-структурой. Таким образом, по результатам экспериментальной проверки наилучшими характеристиками обладает GSM-модель с учетом расстояний до центра влияния.

Сравнение вариантов и выбор модели. В целом значения критериев Corr2, Std, МП в моделях с дополнительным регрессором x_3 выше, чем в моделях с добавленным регрессором, при этом, значения критериев Акаике и Шварца не увеличиваются, что говорит об эффективности использования дополнительного фактора.

Для проверки значимости эффекта учета расстояний до центра влияния используем диспер-

Таблица 5

Оценки качества SAR-моделей

Критерий	Матрица модели					
	W	W_2	W_4	W	W_2	W_4
Акаике	4,621	4,733	4,732	4,518	4,61	4,554
Шварца	4,776	4,888	4,887	4,711	4,8	4,747
Corr2	0,7303	0,6994	0,6988	0,7658	0,7439	0,7461
МП	−160,05	−163,15	−162,30	−155,98	−158,45	−157,66
Std	8,705	9,196	9,198	8,102	8,475	8,435

Таблица 6

Оценки качества SEM-моделей

Критерий	Матрица модели					
	W	W_2	W_4	W	W_2	W_4
Акаике	4,577	4,684	4,493	4,663	4,77	4,578
Шварца	4,731	4,838	4,648	4,856	4,96	4,771
Corr2	0,7297	0,6868	0,7274	0,7555	0,7173	0,7667
МП	−163,21	−165,49	−162,31	−159,75	−166,88	−157,88
Std	8,99	9,521	8,86	8,4697	8,995	8,167

Таблица 7

Оценки качества GSM-моделей без учета информации о расстоянии до центров влияния

Критерий	Матрица модели					
	$W_2 - W_4$	$W_4 - W_2$	$W - W_4$	$W_4 - W$	$W - W_2$	$W_2 - W$
Акаике	4,755	4,539	4,679	4,806	4,647	4,812
Шварц	4,948	4,694	4,872	5,000	4,84	5,005
Corr2	0,7196	0,7174	0,7382	0,7029	0,7459	0,7070
МП	−161,452	−161,179	−159,185	−161,866	−158,96	−162,514
Std	8,914	8,918	8,578	9,161	8,458	9,155

Таблица 8

Оценки качества GSM-моделей с учетом информации о расстоянии до центров влияния

Критерий	Матрица модели					
	$W_2 - W_4$	$W_4 - W_2$	$W - W_4$	$W_4 - W$	$W - W_2$	$W_2 - W$
Акаике	4,675	4,681	4,598	4,727	4,596	4,74
Шварц	4,906	4,913	4,829	4,958	4,827	4,972
Corr2	0,7613	0,7583	0,7765	0,7474	0,7782	0,7436
МП	−156,974	−156,8445	−154,936	−157,429	−155,17	−158,476
Std	8,203	8,234	7,917	8,418	7,892	8,48

сионный анализ. Будем рассматривать значения каждого из критериев Corr2, Std, МП (признаков) в моделях, построенных без дополнительного регрессора с соответствующими значениями в моделях с дополнительным регрессором. Каждый из признаков имеет $p = 2$ уровней и q наблюдений на каждом уровне. ($q = 3$ для SAR- и SEM-моделей и $q = 6$ для GSM модели). Для каждого из признаков вычислим факторную (порождаемую воздействием фактора) и остаточную (обусловленную случайными причинами) дисперсии и проверим значимость различия дисперсий по критерию Фишера—Снедекора. Если различие между этими дисперсиями значимо, то фактор оказывает существенное влияние на признак, следовательно, средние наблюдаемых значений на каждом уровне различаются значимо.

Значения критерия Фишера—Снедекора и критические значения приведены в табл. 9.

Таблица 9

Значения критерия Фишера – Снедекора

Критерий	Модель		
	SAR	SEM	GSM
Corr2	11,49	2,431	18,436
МП	14,43	0,557	26,96
Std	11,84	3,382	18,722
Акаике	7,644	1,19	0,545
Шварц	4,654	2,53	0,2
Критические значения	7,71	7,71	4,94

Критические значения критерия Фишера — Снедекора с уровнем значимости 0,05 равны: 7,71 (для SAR- и SEM-моделей), 4,94 (для GSM-модели). Из результатов, приведенных в табл. 9, следует, что увеличение значений критериев Corr2, Std и МП значимо в случае SAR- и GSM-моделей, а значения критериев Акаике и Шварца не увеличиваются, что свидетельствует об улучшении модели.

ЗАКЛЮЧЕНИЕ

Предложен подход к построению компьютерной массовой оценки объектов недвижимости, позволяющий учесть совместное влияние двух пространственных факторов: расстояния до центров влияния (объектов, не являющихся объектами выборки, близость к которым увеличивает или уменьшает стоимость оцениваемых объектов) и взаимного расположения объектов друг относительно друга. Предложенный подход был реализован в виде пространственной авторегрессионной модели, в которой близость объектов друг к другу описывается пространственной матрицей, построенной по координатам объектов. Для учета расстояний до центров влияния в модель включается дополнительный регрессор, значения которого вычисляются в виде функции от координат объекта и остатков (разностей между фактическими и модельными значениями) от линейной регрессионной модели, построенной без учета пространственных факторов.

Методика реализации предложенного подхода продемонстрирована на примере анализа пространственных данных. Эксперимент показал значимое улучшение качества оценки при введении в модель двух факторов учета пространственной зависимости, что показывает эффективность предложенного подхода.

Дальнейшее развитие подхода, учитывающего различные виды пространственной зависимости, заключается в исследовании и разработке моделей, позволяющих учитывать неравномерность влияния пространственного фактора: выделение локальных участков, на которых резко изменяется направление его воздействия.

Поскольку процедура построения моделей формализована на всех этапах и реализована в виде компьютерной программы, то предложенные модели и методика могут быть применены для построения компьютерной массовой оценки совокупности объектов недвижимости, расположенных на обширной территории, в целях налогообложения.

ЛИТЕРАТУРА

1. *Eckert J.K., O'Connor P.M., Chamberlain C.* Computer-Assisted Real Estate Appraisal. A California Savings and Loan Case Study // The Appraisal Journal. — 1993. — October. — P. 524—532.
2. *Anselin L.* Spatial Econometrics: Methods and Models. — Dordrecht: KluWer Academic, 1988. — P. 284.
3. *Cliff A., Ord J.K.* Spatial Processes: Model, and Application. — London: Pion, 1981. — P. 327.
4. *LeSage James P.* Spatial Econometrics. — 1999. — URL: www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf (дата обращения 6.11.2013).
5. *Kelejian H.H., Prucha I.* A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Mode l // International Economic Review. — 1999. — Vol. 40, iss. 2. — P. 509—533.
6. *Kelejian H.H., Prucha I.R., Yuzefovich E.* Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results // Advances in Econometrics. — 2004. — Vol. 18. — P. 163—198.
7. *Anselin L., Raymond J.G., Florax M.* Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results., New Directions in Spatial Econometrics (L. Anselin and Raymond J.G.M. Florax, eds). — Berlin: Springer, 1995. — P. 21—74.
8. *Anselin L., Rey S.* Properties of Tests for Spatial Dependence in Linear Regression Models // Geographical Analysis. — 1991. — Vol. 23, N 2. — P. 112—131.

Статья представлена к публикации членом редколлегии Р.М. Нижегородцевым.

Беляева Анна Валерьевна — аспирант, ✉ BelyaevaAV@gmail.com,

Гребенюк Елена Алексеевна — д-р техн. наук,
вед. науч. сотрудник, ☎ (495) 334-46-40, ✉ lngrebenuk@rambler.ru,

Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва.