

УДК 519.237.5
ББК 22.172

УЧЕТ ПРОСТРАНСТВЕННЫХ ФАКТОРОВ В МАССОВОЙ ОЦЕНКЕ ОБЪЕКТОВ НЕДВИЖИМОСТИ: СРАВНЕНИЕ ЭФФЕКТИВНОСТИ РАЗЛИЧНЫХ МЕТОДОВ

Беляева А. В.¹

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)*

Исследуется эффективность методов компьютерной массовой оценки (КМО) объектов недвижимости, использующих информацию о местоположении объекта. Это исследование является актуальным в силу того, что до построения модели не всегда может быть определена однозначно структура модели (тип пространственной автокорреляции, число ближайших соседей). Предложена методика оценки эффективности различных методов КМО, реализованная посредством имитационной модели. Приведены результаты исследования: 1) области эффективного применения методов в зависимости от типа пространственной зависимости в исходных данных; 2) результаты сравнения робастности методов. Методы различаются способами учета информации о пространственном расположении объекта.

Ключевые слова: компьютерная массовая оценка объектов недвижимости, пространственные авторегрессионные модели, центр влияния, имитационная модель, чувствительность методов.

¹ Анна Валерьевна Беляева, аспирант, (belyaevaav@gmail.com).

1. Введение

Массовая оценка объектов недвижимости – это оценка стоимости объектов недвижимости, которая проводится систематически на определенную дату, и при этом используются стандартные процедуры статистического анализа. Главное отличие массовой оценки от экспертной – работа не с отдельными объектами, а со статистическими данными, описывающими совокупность объектов.

Направление исследования зависимости стоимости объекта от его пространственных характеристик развивалось в последние три десятилетия в работах зарубежных ученых. Многочисленные эксперименты [1–9] показали, что рынок недвижимости наиболее качественно описывается с помощью моделей, учитывающих местоположения объекта недвижимости.

Методы построения пространственных моделей, учитывающие зависимость цены объекта от его расположения, можно разделить на три группы:

- методы, учитывающие при формировании цены объекта расстояние от него до центров влияния (объектов, которые не являются объектами выборки, но оказывают позитивное или негативное влияние на стоимость объектов недвижимости) [7];
- методы, учитывающие наличие пространственной автокорреляции между характеристиками недвижимости [3, 5];
- методы, основанные на разделении области оценивания на однородные зоны и анализе поведения цен в каждой из зон [2].

В работе [1] был предложен метод компьютерной массовой оценки, который одновременно учитывает влияние двух пространственных факторов на цену недвижимости: расстояние до центров влияния и пространственную автокорреляцию.

В настоящей статье средствами математического моделирования проведен сравнительный анализ эффективности методов КМО, учитывающих влияние пространственного фактора различными способами, включая метод оценивания, не учитывающий пространственный фактор. Для этого формируются обучающая и контрольная выборки объектов недвижимости (с ценами и характеристиками), построенные для различных форм про-

странственной зависимости в данных (с помощью имитационной модели), а затем применяется следующий алгоритм:

– по обучающей выборке для каждого из наборов смоделированных ситуаций методами М1–М4, о которых сказано ниже, вычисляются параметры модели оценивания;

– сравнивается качество моделей оценивания, построенных разными методами (для оценивания и сравнения прогнозных характеристик моделей используется контрольная выборка) и проверяется чувствительность моделей оценивания к параметрам, определяющим структуру модели.

2. Методика оценки эффективности КМО

Процедура оценки эффективности методов построения КМО, используемая в настоящей статье, включает следующие основные этапы.

Этап 1. Выбор методов для сравнения их эффективности. Для сравнения были выбраны:

– метод М1 линейной регрессии, модель оценивания:

$$(1) \quad Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n,$$

где (X_1, \dots, X_n) – непространственные характеристики объекта недвижимости, β_i – численные коэффициенты при непространственных характеристиках объекта, показывающие вклад соответствующих характеристик в стоимость;

– метод М2, учитывающий расстояние до центров влияния; модель оценивания совпадает с моделью (1), в которую добавляется дополнительное число регрессоров, равное числу центров влияния, оказывающих воздействие на цену объекта; значения регрессоров являются функциями от расстояний до центров влияния;

– метод М3 пространственных моделей. Рассматриваются

а) модели с пространственной корреляцией в лагах:

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

где y – $(n \times 1)$ -вектор зависимых переменных; X – $(n \times k)$ -матрица непространственных характеристик объекта недвижимости; ρ – коэффициент пространственной зависимости; β – $(k \times 1)$ -вектор коэффициентов при непростран-

ственных характеристиках объекта; ξ – случайная составляющая; W – $(n \times n)$ -матрица учета пространственных взаимодействий, элементы w_{ij} , $i, j = 1, \dots, n$, которой формируются по правилу:

$$w_{ij} = \begin{cases} 1, & \text{объекты } i \text{ и } j \text{ смежны в нектором смысле,} \\ 0, & \text{объекты } i \text{ и } j \text{ не смежны в этом же смысле;} \end{cases}$$

б) модели с пространственной корреляцией в ошибках:

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

где u – $(n \times 1)$ -вектор возмущений, λ – коэффициент пространственной корреляции в ошибках модели. Остальные обозначения и параметры совпадают с обозначениями и параметрами модели (2).

Смежность во всех экспериментах определяется по правилу p (в смысле евклидовой метрики) ближайших соседей, где p задается на основе анализа характера пространственной зависимости, но практически определяется экспериментальным путем – сравнением вариантов с разными значениями p .

– метод М4, предложенный в [1], модели оценивания которого совпадают с моделями (2) и (3), в которые добавляется дополнительное число регрессоров X , равное числу центров влияния, оказывающих воздействие на цену объекта.

Выбранные методы различаются только способом учета пространственного фактора: не учитывают (М1), учитывают только один из них (М2, М3), учитывают оба (М4). Для их реализации могут быть использованы стандартные программы, реализованные в большинстве статистических пакетов.

Этап 2. Построение моделей цен объектов, определяемых непространственными и пространственными факторами. На формирование цен, помимо таких факторов как число комнат, этаж, тип стен, общая площадь, площадь кухни и пр., оказывают влияние пространственные факторы: центры влияния и пространственная корреляция между характеристиками объектов. Модели цен различаются по числу и направленности центров влияния, по типу пространственной корреляции (в лагах или в ошибках), по силе воздействия пространственного фактора

(сильное, слабое, отсутствует), по виду матриц пространственных взаимодействий.

Этап 3. Для каждой модели формирования цен методами М1–М4, выбранными на шаге 1, вычисляется набор критериев, оценивающих качество модели, построенной по обучающей выборке и качество прогнозирования.

В случае пространственных моделей вопрос выбора критериев оценки качества моделей не является тривиальным – в силу структуры пространственных моделей и используемых методов оценивания остатки могут не иметь нулевого математического ожидания и, следовательно, разложения дисперсии на объясняемую дисперсию и дисперсию остатков не существует, поэтому критерий R^2 теряет смысл. Применяются следующие критерии:

- критерий Акаике (K_1)

$$AIC = \frac{2k}{n} - 2\ln(L),$$

где k – число параметров модели, L – максимизированное значение функции правдоподобия модели, n – объем выборки, по которой строилась модель;

- квадрат коэффициента корреляции между наблюдаемыми и предсказанными по модели значениями (K_2)

$$corr = \frac{\sum (Y_i - \bar{Y}_i)(Y_i^m - \bar{Y}_m)}{\sqrt{\sum (Y_i - \bar{Y}_i)^2 \sum (Y_i^m - \bar{Y}_m)^2}},$$

где

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \bar{Y}_m = \frac{1}{n} \sum_{i=1}^n Y_i^m$$

средние значения, Y_i – цены объектов выборки, Y_i^m – модельные значения цен объектов выборки;

- оценка величины смещения среднего остатков относительно нуля (K_3)

$$meanres = \frac{\sum_{i=1}^n res_i}{n},$$

где res_i – остатки модели;

- среднеквадратичное отклонение остатков (K_4)

$$sumolsar = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (res_i - \overline{res})^2},$$

– сумма квадратов ошибок прогноза цен контрольной выборки, построенного в результате анализа обучающей выборки (качество прогнозирования) (K_5)

$$forecast = \sqrt{\sum_{i=1}^n (Y_i - Y_i^m)^2},$$

где Y_i – цены объектов контрольной выборки, Y_i^m – значения цен объектов контрольной выборки, вычисленные по моделям, построенным по обучающей выборке.

Одним из важнейших критериев качества построенной модели является точность прогнозирования цен объектов. Необходимость в прогнозировании возникает тогда, когда нам необходимо оценить объекты, информация по сделкам о которых в нашей базе данных отсутствует на исследуемую дату. Для построения прогноза предлагается следующий алгоритм:

1. По критериям K_1 – K_4 строится модель (2) или (3) по обучающей выборке.
2. Для каждого из r объектов контрольной выборки вычисляется регрессор:

$$dci_i = \frac{\alpha}{d_i},$$

где коэффициент α определяется по экспериментальным данным минимизацией функции

$$I = \sum_{i=1}^n \left[res_i - \frac{\alpha}{d_i} \right],$$

res_i – остатки регрессии, d_i – расстояние от i -го объекта до центра влияния, $i = 1, 2, \dots, n$ и определяется в результате решения уравнения:

$$\frac{\partial I}{\partial \alpha} = \sum_{i=1}^n \left[res_i - \frac{\alpha}{d_i} \right] / d_i = 0.$$

3. Для каждого из r объектов контрольной выборки выбирается p ближайших соседей из обучающей выборки размера n , где p соответствует числу ближайших соседей построенной модели и строится $r \times n$ матрица W_{pc} пространственных взаимодействий объектов контрольной выборки согласно алгоритму, описанному в [1].

4. Прогнозное значение цены \hat{Y}_k , $k = 1, 2, \dots, r$, вычисляется по формулам:

$$\hat{Y}_k = \rho W_{pc} Y_l + X_k \beta \text{ для модели (2),}$$

$$\hat{Y}_k = X_k \beta + \lambda W_{pc} (Y_l - X_l \beta) \text{ для модели (3),}$$

где X_k – значения регрессоров контрольной выборки, $Y_l = (y_1, \dots, y_n)$ и X_l – значения цен и регрессоров обучающей выборки, коэффициенты β , λ , ρ оценены по обучающей выборке.

По результатам анализа значений критериев для каждой модели формирования цен определяется метод, являющийся наиболее эффективным для данной модели.

Этап 4. В результате анализа поведения критериев для различных моделей формирования цен формируются группы моделей (множества смоделированных наборов цен ниже называются Ситуациями) таким образом, чтобы в них попали модели формирования цен, для которых проверяемые методы имеют одинаковое ранжирование по эффективности. В результате для каждой Ситуации определяется наиболее эффективный метод.

3. Построение имитационной модели

Для проведения сравнительного анализа эффективности методов учета пространственных факторов была построена имитационная модель, включающая различные варианты воздействий пространственного фактора на формирование цен (возможные варианты взаимодействия получены на основании многочисленных экспериментов на данных реального рынка недвижимости [1, 4–9]).

Построение модели включает следующие шаги:

1. Генерация объектов. На плоскости $[0, 100] \times [0, 100]$ генерируются координаты 150 объектов (из них 100 – объекты обу-

чающей выборки, 50 – объекты контрольной выборки). Выборка объектов разбивается на обучающую и контрольную выборки следующим образом: из 150 сгенерированных объектов объект, номер которого кратен трем, попадает в контрольную выборку, остальные объекты – в обучающую.

2. Генерация регрессоров. Для каждого из 150 объектов генерируется вектор независимых переменных $\{x_1, x_2\}$, первая компонента – из равномерного распределения на отрезке $[0, 5]$, вторая – из равномерного распределения на отрезке $[10, 13]$. Как и в п.1, выборка регрессоров разбивается на контрольную и обучающую выборки.

3. Определение координат центров влияния и расчет регрессоров, учитывающих расстояния от объектов до центра влияния. В рамках проводимых экспериментов рассматривалось два варианта по наличию центров влияния. Один центр положительного влияния, в качестве центра была выбрана точка с координатами $(35,15; 60,86)$; и три центра влияния различной направленности: два положительной с координатами $(35,15; 60,86)$ и $(43,62; 20,30)$ и один отрицательной с координатами $(74,94; 29,55)$.

В случае одного центра влияния для каждой пары координат объекта $x_i, y_i, i = 1, \dots, 150$, вычисляется переменная d_i – евклидово расстояние от объекта до центра влияния; для учета воздействия центра влияния на цену выбирается функция вида

$$(4) \quad f(d_i) = \exp(-\alpha d_i),$$

где параметр α определяет радиус действия центра влияния, и вычисляются значения регрессора $X_3 = \{x_{31}, x_{32}, \dots, x_{3150}\}$, где $x_{3i} = f(d_i) = \exp(-\alpha d_i)$.

В случае нескольких центров влияния для каждого из них рассчитываются расстояния d_i^1, d_i^2, d_i^3 , где d_i^k – расстояние от i -го объекта до k -го центра влияния и по формуле (4) вычисляются значения регрессоров X_3, X_4, X_5 . Регрессионная матрица имеет вид $\{I, X_1, X_2, X_3\}$ в случае одного центра влияния, $\{I, X_1, X_2, X_3, X_4, X_5\}$ – в случае трех центров влияния.

4. Формирование матриц пространственных взаимодействий для методов М3 и М4. Построение пространственных матриц i ближайших соседей W_i размера 100×100 для объектов обучаю-

щей выборки, где $i = 2, 4, 8, 12$. Для объектов контрольной выборки строятся матрицы W_{ic} размера 50×100 , в которых для каждого объекта контрольной выборки выбирается i , $i = 2, 4, 8, 12$, ближайших соседей, из обучающей выборки.

Данные пунктов 1–4 одинаковы для всех проводимых экспериментов.

5. Формирование моделей цен. Строятся две модели цен: цены с пространственной зависимостью в лагах и цены с пространственной зависимостью в остатках.

Для генерации цен с пространственной зависимостью в лагах используется модель (2), цен с пространственной корреляцией в ошибках – модель (3).

Матрицы пространственных взаимодействий строятся по алгоритму нахождения p ближайших соседей. Формируются модели цен с $p = 2, 4, 8, 12$.

Параметры модели:

$\beta = \{\beta_0, \beta_1, \beta_2, \beta_3\}$ для экспериментов с одним центром влияния, где: $\beta_0 = 0,3$, $\beta_1 = \beta_2 = 20$, $\beta_3 = 50$;

$\beta = \{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ для случая трех центров влияния: $\beta_0 = 0,3$, $\beta_1 = \beta_2 = 20$, $\beta_3 = \beta_4 = 50$, $\beta_5 = -20$.

Параметры p и λ в зависимости от эксперимента, могут принимать значения 0,8 (сильная зависимость) или 0,2 (слабая зависимость).

При расчете регрессоров X_3 и X_3 , X_4 , X_5 по формуле (4) параметр α выбирается равным $\alpha = 0,04$, дисперсия случайной составляющей $\sigma = 1$.

Таким образом, было сформировано 32 набора данных, различающихся:

- числом и расположением центров влияния;
- формой пространственной корреляции в данных (в лагах и в ошибках);
- числом ближайших соседей пространственной матрицы;
- силой воздействия пространственного фактора (сильное и слабое воздействие).

4. Структура эксперимента и проведение анализа эффективности методов

Множество смоделированных наборов цен было разделено на 4 группы, которые дальше мы будем называть Ситуациями. Было выделено четыре Ситуации:

Ситуация 1. Сильная пространственная автокорреляция между ценами;

Ситуация 2. Слабая пространственная автокорреляция между ценами;

Ситуация 3. Сильная пространственная автокорреляция между факторами, неучтенными регрессионной моделью;

Ситуация 4. Слабая пространственная автокорреляция между факторами, неучтенными регрессионной моделью.

В Ситуациях 1 и 2 цены формируются с использованием модели (2).

В Ситуациях 3 и 4 цены формируются с использованием модели (3).

В каждой Ситуации построены модели цен с одним (положительным) центром влияния и модели с тремя центрами влияния различной направленности (два положительной направленности, один – отрицательной).

Для каждого из двух вариантов состава центров влияния сформированы 4 модели цен, различающихся числом p ближайших соседей.

Эксперимент включает: построение модели формирования цен, построение модели КМО по методам М1, М2, М3, М4, вычисление значений критериев K_1-K_5 для каждой из построенных моделей.

Каждый эксперимент повторяется 500 раз с различными значениями случайной составляющей модели цен. После окончания эксперимента рассчитываются средние значения критериев, которые используются для анализа сравнительной эффективности методов в Ситуациях 1–4.

В качестве примера на рис. 1 и рис. 2 представлены графики среднеквадратичной ошибки прогноза для различных Ситуаций: сплошной линией для случая одного центра влияния, пунктирной линией – для случая трех центров влияния. По оси ординат

указано значение среднеквадратичной ошибки прогноза, по оси абсцисс отложены названия методов, для которых оно было получено в экспериментах рассматриваемой Ситуации. Названия методов повторяются для каждого из четырех экспериментов (с двумя, четырьмя, восьмью и двенадцатью ближайшими соседями).

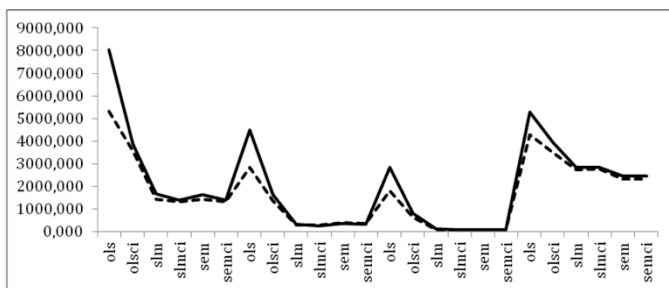


Рис. 1. Среднеквадратичная ошибка прогноза для Ситуации 1

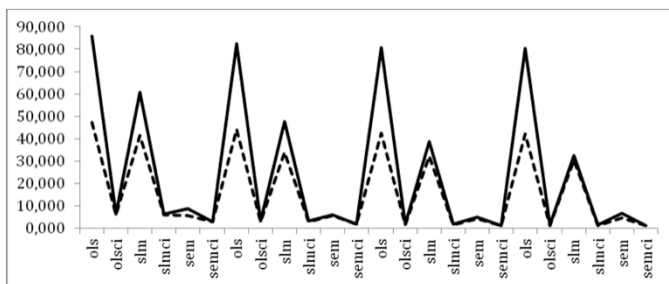


Рис. 2. Среднеквадратичная ошибка прогноза для Ситуации 3

Используются обозначения:

- ols – линейная регрессия;
- olsci – линейная регрессия с центром влияния;
- slm – модель с пространственной корреляцией в лагах;
- slmci – модель с пространственной корреляцией в лагах с учетом центра влияния;
- sem – модель с пространственной корреляцией в остатках;
- semci – модель с пространственной корреляцией в остатках с учетом центра влияния.

Из графиков видно, что в рамках одной ситуации поведение значения критерия K_5 для одного и трех центров влияния идентично, но отличается для разных ситуаций.

Это подтверждается соответствующими расчетами.

В Ситуации 1:

– метод М4 во всех экспериментах имеет наименьшую ошибку прогноза и является более эффективным, чем метод М3: применение метода М4 уменьшает ошибку прогноза на 14–25 % по сравнению с методом М3;

– метод М1 во всех экспериментах дает ошибку прогноза в 2–28 раз больше, чем метод М3, метод М2 во всех экспериментах дает ошибку прогноза в 2–8 раз больше, чем метод М3. Поэтому использование методов М1 и М2 не эффективно в рассматриваемой Ситуации.

В Ситуации 3:

– метод М4 с SEM моделью во всех экспериментах имеет наименьшую ошибку прогноза и является более эффективным, чем метод М2: уменьшает ошибку прогноза метода М2 на 11–52 %;

– метод М1 во всех экспериментах дает ошибку прогноза в 2–4 раза больше, чем метод М3 с SLM моделью для одного центра влияния и на 25% больше для трех центров влияния. Метод М3 с SEM моделью более эффективен, однако во всех экспериментах он дает ошибку прогноза в 2–13 раз большую, чем метод М2. Поэтому использование этих методов М1 и М3 не эффективно в рассматриваемой ситуации.

В таблице 1 подведены итоги анализа эффективности применения методов. На пересечении строки, соответствующей Ситуации, и столбца, соответствующего критерию, указаны методы, наиболее эффективные по этому критерию. В крайнем правом столбце указана лучшая модель для ситуации в целом.

Значение критерия K_3 отлично от нуля только для случаев, где модель данных строится по большому числу соседей – а именно, 12 соседей, поэтому в таблицу он не включен.

В результате проведенного анализа для каждой ситуации были определены методы, использование которых предпочтительнее в данной Ситуации. Результаты анализа показали, что:

– учет пространственных факторов повышает эффективность оценок прогноза в несколько раз, по сравнению с линейной регрессионной моделью М1;

– метод М4 значительно эффективнее методов М1–М3 в Ситуациях 1–3, а в Ситуации 4 сравним по эффективности с методом линейной регрессии с учетом расстояний до центров влияния М2, но последний предпочтительнее (по критерию Акаике), так как является более простым в реализации.

– в Ситуации 1 метод М4 эффективнее метода М3 в среднем на 15–25 % (в зависимости от используемой модели цен), а метод М2 хуже М4 в несколько раз; в Ситуациях 2 и 3 метод М4 эффективнее методов М2 и М3, в Ситуации 4 метод М4 незначительно уступает по эффективности методу М2, а применение метода М3 ухудшает значения критериев в несколько раз.

Таблица 1. Свод по применению методов

Ситуация	Качество прогноза (K_5)	Диагностика модели			ИТОГ
		Критерий Акаике (K_1)	Корреляция модельных и реальных цен (K_2)	Среднеквадратичное отклонение остатков (K_4)	
Ситуация 1	1) М4 с SLM моделью дает минимальные ошибки прогноза, уменьшает ошибку прогноза М3 на 14–25 % . 2) М1 – ошибка прогноза в 2–28 раза больше, чем по М3, М2 ошибка прогноза в 2–8 раз больше, чем по М3.	1) М4 во всех экспериментах дает минимальное значение K_1 . 2) М3 по этому критерию значительно эффективнее М1 и М2.	1) М3 и М4 имеют K_2 выше, чем М1 и М2. 2) М4 и М3 по значению K_2 не различаются.	1) М3 и М4 имеют K_4 остатков меньше, чем М1 и М2. 2) М4 имеет значение K_4 остатков на 3–24 % ниже, чем М3.	Эффективный метод – М4 с SLM–моделью
Ситуация 2	1) М2 и М4 – ошибка прогноза меньше, чем по М1 и М3, соответственно. 2) М4 – наименьшая ошибка прогноза. 3) М3 с SLM моделью – ошибка прогноза в 2–6 раз больше, чем по М2.	1) Методы с учетом центров влияния М2 и М4 во всех экспериментах позволяют построить модель лучшего качества, чем М1 и М3. Так, М2 уменьшает значение K_1 М1 на 28–55%, М3 – на 5–35%. 2) В свою очередь, М4	1) М2 и М4 имеют K_2 выше, чем М1 и М3. 2) М4 и М2 по значению K_2 не различаются.	1) М2 уменьшает значение K_4 М1 на 49–73%, М3 – для усредненного по двум методам slm и sem на 12–44%. Но в экспериментах 5, 6, 21, 22, 23 модель sem обладает качеством, лучшим, чем М2 на	Эффективный метод – М4 с SEM моделью для числа соседей больше двух, М4 с SLM моделью – для числа соседей, равных двум

Ситуация	Качество прогноза (K_5)	Диагностика модели			ИТОГ
		Критерий Акаике (K_1)	Корреляция модельных и реальных цен (K_2)	Среднеквадратичное отклонение остатков (K_4)	
	4) М4 с SEM моделью-ошибка прогноза меньше, чем по М2 на 13–43%.	уменьшает значение K_1 на 24–45% по сравнению с М2.		3–15%. 2) М4 уменьшает значение K_4 М2 на 30–49%. При этом для большого числа соседей (12 соседей, эксперименты 8 и 24) наилучшей с точки зрения этого критерия является модель semci.	
Ситуация 3	1) М4 с SEM моделью-минимальная ошибка прогноза. 2) М1 – ошибка прогноза в 2–4 раза больше, чем по М3. 3) М3 – ошибка прогноза в 2–13 раз больше, чем по М2. 4) М4 – ошибка прогноза на 11–52% меньше, чем по М2.	1) Методы М2 и М4 во всех экспериментах являются более эффективными, чем М1 и М3. Они позволяют уменьшить значение K_1 на 38–90% по сравнению с М1 и М3. 2) Причем для такого варианта пространственной зависимости лучшим является метод semci. М4 позволяет уменьшить	1) М2 и М4 имеют K_2 выше, чем М1 и М3: на 2–7%. 2) М4 и М2 по значению K_2 не различаются.	1) М2 и М4 имеют K_4 остатков на 68– 88% меньше, чем М1 и М3. 2) М4 уменьшает значение K_4 остатков М2 на 8–37%.	М4 с SEM моделью эффективнее М2 и значительно эффективнее М3

Ситуация	Качество прогноза (K_5)	Диагностика модели			ИТОГ
		Критерий Акаике (K_1)	Корреляция модельных и реальных цен (K_2)	Среднеквадратичное отклонение остатков (K_4)	
		значение K_1 M2 на 28–60%.			
Ситуация 4	<p>1) M2 и M4 –ошибка прогноза меньше, чем по M1 и M3.</p> <p>2) M4 и M2 по критерию «ошибка прогноза» одинаково эффективны.</p> <p>3) M3 – ошибка прогноза увеличивается в 16–29 раз по сравнению с ошибкой прогноза по M2.</p>	<p>1) Методы с учетом центров влияния M2 и M4 во всех экспериментах являются более эффективными, чем M1 и M3: значение K_1 уменьшается на 92– 97%</p> <p>2) M2 в случае слабой пространственной зависимости в остатках модели дает более низкие значения K_1 по сравнению с M4 5–20%.</p>	<p>1) M2 и M4 имеют K_2 выше, чем M1 и M3: на 2–7%.</p> <p>2) M4 и M2 по значению K_2 не различаются.</p>	<p>1) M2 и M4 уменьшают K_4 остатков M1 и M3 на 80–89%.</p> <p>2) M4 и M2 по значению K_4 не различаются.</p>	Эффективный метод – M2

5. Анализ чувствительности методов

С использованием построенной имитационной модели было проведено исследование чувствительности методов, использующих пространственные модели. При построении пространственных моделей выбор типа модели (SLM, SEM) и числа ближайших соседей обычно определяется пользователем, так как тесты множителей Лагранжа зачастую не дают однозначного ответа о типе пространственной корреляции и не существует эффективных методов определения числа ближайших соседей на основе анализа выборки. В связи с этим возникает вопрос о чувствительности методов М3 и М4 относительно выбора типа модели. Метод, обладающий меньшей чувствительностью к выбору типа модели и заданию числа ближайших соседей, является более робастным.

Чувствительность относительно выбора типа пространственной модели измеряется величиной относительного изменения значения критерия при построении модели, тип которой совпадает с типом модели цен, до значения критерия, где тип модели цен и оцениваемой модели не совпадают. Пусть x_1 – значение какого-либо критерия K_i при построении SLM модели по данным цен генерируемым моделью SLM, x_2 – значение критерия K_i при построении SEM модели по тем же данным. Относительное изменение значения критерия при замене типа модели рассчитывается по формуле:

$$(5) \quad \alpha = \frac{x_1 - x_2}{x_1} \cdot 100\% .$$

Рассмотрим критерий K_5 . Положительное значение α указывает на то, что ошибка прогноза модели SLM выше, чем модели SEM для анализируемых данных, и в данном эксперименте предпочтительнее SEM модель.

Сравним чувствительность методов М3 и М4 в Ситуации 1. На рис. 3 представлены значения α для метода М4 – сплошная линия, результаты метода М3 – пунктир. По оси ординат отложены значения α , по оси абсцисс – номера экспериментов.

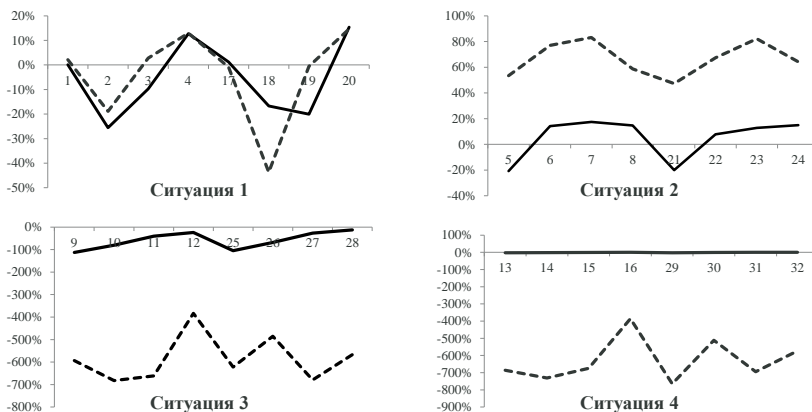


Рис. 3. Чувствительность методов

По критерию Стьюдента для парных выборок значения α для алгоритмов М4 и М3 значимо не различаются, таким образом, в Ситуации 1 методы М3 и М4 имеют одинаковую чувствительность.

В Ситуации 2 чувствительность метода М3 к выбору модели в несколько раз превосходит чувствительность метода М4.

В Ситуациях 3 и 4 для методов М4 и М3 использование SLM модели увеличивает ошибку прогноза в несколько раз в методе М3 и на 70–100 % при малом числе соседей при использовании метода М4. В остальных случаях метод М4 не является чувствительным к выбору типа модели.

Как видно из графиков, М4 в Ситуациях 2–4 обладает меньшей чувствительностью к выбору типа модели и заданию числа ближайших соседей, чем метод М3.

Выбор числа ближайших соседей может значительно изменить качество пространственной модели. Сравним чувствительность методов М4 и М3 в различных Ситуациях. Для вычисления чувствительности в каждом эксперименте вычислим относительное изменение ошибки прогноза по формуле (5). Вычислим среднюю величину ошибки прогноза по всем экспериментам рассматриваемой Ситуации для методов М3 и М4. Результаты представлены в таблице 2.

Таблица 2. Результат анализа чувствительности методов

Метод	Ситуация 1	Ситуация 2	Ситуация 3	Ситуация 4
Метод 3	48%	11%	10%	3%
Метод 4	46%	8%	19%	0%

Методы М3 и М4 имеют одинаковую чувствительность к точности задания числа ближайших соседей.

6. Заключение

Сравнительный анализ эффективности рассмотренных методов при различных вариантах влияния пространственного фактора на цену (Ситуациях) показал, что метод М4, учитывающий оба пространственных фактора, является наиболее эффективным в трех рассматриваемых Ситуациях и сравним по эффективности с методом линейной регрессии с учетом расстояний до центров влияния в четвертой Ситуации. Метод М4 является менее чувствительным к заданию типа пространственной модели, чем метод М3.

Литература

1. БЕЛЯЕВА А.В. , ГРЕБЕНЮК Е.А. , *Построение моделей массовой оценки объектов недвижимости с учетом пространственной зависимости* // Проблемы управления. – 2014. – №1. – С. 45–52.
2. СТЕРНИК Г.М. *Математические основы методологии построения дискретных пространственно-параметрических моделей рынка жилья*. – Рынок недвижимости России, 2002. [Электронный ресурс]. – URL: <http://realtymarket.ru/metodi-eskie-materiali/Sternik-G.M.-MATEMATI-ESKIE-OSNOVI-METODOLOGII-POSTROENIYA-DISKRETNIIH-PROSTRANSTVENNO-PARAMETRI-ESKIH-MODELEI-RINKA-JILYA.html> (дата обращения 02.01.2015)
3. ANSELIN L. *Spatial Econometrics: Methods and Models* – Dordrecht: Kluwer Academic, 1988. – 284 p.

4. BIDANSET P., LOMBARD J. *Evaluating Spatial Model Accuracy in Mass Real Estate Appraisal: a Comparison of Geographically Weighted Regression and the Spatial Lag Model* // Cityscape: A Journal of Policy Development and Research. – 2014. – No 3. – P. 169–182.
5. CLIFF A., ORD J.K. *Spatial Processes: Model, and Application*. – London: Pion, 1981. – 327 p.
6. CONWAY D., LI C. Q., WOLCH J. AND ETC. *A spatial autocorrelation approach for examining the effects of urban green space on residential property values* // The Journal of Real Estate Finance and Economics. – 2010. – No. 41(2). – P. 150–169.
7. 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.
8. FINGLETON B. *A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices* // Empirical Economics. – 2008. – No. 34. – P. 35–57.
9. MÖLLER J. *Regional variations in the price of building land: a spatial econometrics approach for West Germany* // The Annals of Regional Science – 2009. – No. 43(1). – P. 113–132.

LOCATION FACTORS IN REAL ESTATE MASS APPRAISAL: COMPARING APPROACHES

Anna Belyaeva, Institute of Control Sciences of RAS, Moscow, (belyaevaav@gmail.com).

Abstract: In practice, the structure of a location-based model for real estate appraisal (type of the spatial autocorrelation, neighbor count) is unobvious. We develop the simulation-based methodology to compare accuracy of computer-assisted mass appraisal (CAMA) routines. Then we compare routines differing in the form of location information accounting and build the best routine selection rule basing on spatial autocorrelation type in source data. We also provide routine robustness comparison.

Keywords: computer-assisted real estate mass appraisal, spatial autoregressive model, center of influence, simulation model, robustness.

*Статья представлена к публикации
членом редакционной коллегии С.А. Баркаловым*

*Поступила в редакцию 22.11.2014.
Опубликована 31.01.2015.*