

УДК 519.237.5

## Методика сравнения эффективности методов компьютерной массовой оценки

А. В. Беляева

Федеральное государственное бюджетное учреждение науки  
«Институт проблем управления им. В. А. Трапезникова Российской академии наук»,  
Россия, 117997, г. Москва, ул. Профсоюзная, д. 65  
Телефон: +7 495 334-89-10, факс: +7 495 334-93-40, +7 499 234-64-26  
E-mail: belyaevaav@gmail.com

Получено 26 ноября 2014 г.,  
после доработки 15 декабря 2014 г.

Одним из направлений построения компьютеризированной массовой оценки (КМО) объектов недвижимости является учет местоположения объекта. При учете местоположения объекта с использованием пространственных авторегрессионных моделей структура модели (тип пространственной автокорреляции, выбор числа «ближайших» соседей) не всегда может быть определена однозначно до ее построения. Кроме того, на практике имеются ситуации, когда более эффективными являются методы, учитывающие другой тип зависимости цены объекта от его местоположения. В связи с этим в рамках исследования эффективности применения методов оценки, использующих пространственные методы, важными вопросами являются:

- области эффективности применения методов различной структуры;
- чувствительность методов, использующих пространственные модели, относительно выбора типа пространственной модели и числа ближайших соседей.

В статье проведен сравнительный анализ оценки эффективности методов КМО объектов недвижимости, использующих информацию о местоположении объекта, приведены результаты апробации, выделены области эффективного применения рассматриваемых методов.

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

## Comparing the effectiveness of computer mass appraisal methods

A. V. Belyaeva

*Institute of Control Sciences V. A. Trapeznikov Academy of Sciences,  
65 Profsoyuznaya st., Moscow, 117997, Russia. Tel.: +7 495 334-89-10, fax: +7 495 334-93-40, +7 499 234-64-26*

**Abstract.** — Location-based models — one of areas of CAMA (computer-assisted mass appraisal) building. When taking into account the location of the object using spatial autoregressive models structure of models (type of spatial autocorrelation, choice of “nearest neighbors”) cannot always be determined before its construction. Moreover, in practice there are situations where more efficient methods are taking into account different rates depending on the type of the object from its location. In this regard there are important issues in spatial methods area:

- fields of methods efficacy;
- sensitivity of the methods on the choice of the type of spatial model and on the selected number of nearest neighbors.

This article presents a methodology for assessing the effectiveness of computer evaluation of real estate objects. There are results of approbation on methods based on location information of the objects.

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

Citation: *Computer Research and Modeling*, 2015, vol. 7, no. 1, pp. 185–196 (Russian).

## Введение

Массовая оценка объектов недвижимости — это оценка стоимости объектов недвижимости, которая проводится систематически на определенную дату, при этом используются стандартные процедуры статистического анализа [Эккерт, 1997]. Главное отличие массовой оценки от экспертной — работа не с отдельными объектами, а со статистическими данными, описывающими совокупность объектов, что, в свою очередь, обеспечивает объективность полученной оценки в сравнении с экспертной оценкой. Это дает возможность пояснить налогоплательщикам причины размера налогооблагаемой базы, что является одной из основных характеристик эффективной системы налоговой оценки недвижимости согласно практическому исследованию, проведенному Международной ассоциацией налоговых оценщиков (IAAO).

Еще одним преимуществом массовой оценки перед экспертной является то, что она позволяет провести оценку совокупности объектов недвижимости, расположенных на обширной территории, с минимальными трудозатратами в сравнении с экспертной оценкой.

Направление исследования зависимости стоимости объекта от его пространственных характеристик развивалось в последние три десятилетия в работах зарубежных ученых, но в России среди оценщиков этот подход имеет ограниченное применение, несмотря на то что обеспечивает существенное повышение точности результатов оценки.

Все вышеперечисленное в сочетании с планами внедрения в России единого налога на недвижимость, который должен заменить сразу два ныне действующих налога — на землю и на имущество физических лиц, говорит о том, что проблема разработки эффективных методов построения компьютерной массовой оценки (КМО) объектов недвижимости с учетом их местоположения и методологии их применения является весьма актуальной в настоящее время. В контексте вышесказанного в данной статье приводятся результаты исследования эффективности текущих и нового [Беляева, Гребенюк, 2014] методов КМО с целью их дальнейшего применения на практике.

Среди методов, использующих пространственные характеристики объектов, можно выделить следующие 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; к таким объектам, например, можно отнести парк, который увеличит стоимость близлежащих объектов, или помойку, которая стоимость снизит) [Eckert, O'Connor, Chamberlain, 1993];
- методы, учитывающие наличие пространственной автокорреляции между характеристиками недвижимости [Anselin, 1988; Cliff, Ord, 1981], которая выражается во влиянии цен близкорасположенных объектов друг на друга и/или в одинаковом характере влияния независимых переменных на цену; эти методы подразумевают вычисление в явном виде зависимости каждого наблюдения от всех остальных посредством матрицы учета пространственных взаимодействий;
- методы, основанные на разделении области оценивания на однородные зоны и анализе поведения цен в каждой из зон [Стерник, 2005].

Анализ существующих методов КМО дал предпосылки для попытки объединить первые два метода и оценить качество полученного результата. В работе [Беляева, Гребенюк, 2014] был предложен метод компьютерной массовой оценки, который одновременно учитывает влияние двух пространственных факторов на цену недвижимости: расстояние до центров влияния и пространственную автокорреляцию. Была проведена апробация метода на большом объеме экспериментальных данных (в том числе данных московского рынка недвижимости), давшая положительный результат.

При построении компьютеризированной оценки методов, включающих пространственные модели [Anselin, 1988; Cliff, Ord, 1981; Беляева, Гребенюк, 2014], выбор типа пространственной корреляции с использованием теста множителей Лагранжа [Anselin, 1988b] не всегда позволяет

однозначно определить характер пространственной корреляции (в лагах или в ошибках) и отсутствуют эффективные методы определения числа ближайших соседей, которое используется при построении матрицы учета пространственных взаимодействий. В связи с этим в рамках исследования методов оценки стоимости объектов недвижимости, учитывающих местоположение объектов, важными вопросами являются:

- определение областей эффективности применения методов различной структуры;
- анализ чувствительности методов, использующих пространственные модели, относительно выбора типа пространственной модели и числа ближайших соседей.

В настоящей статье проводится анализ и сравнение методов КМО с использованием имитационного моделирования, представлены результаты сравнения методов, а также анализ чувствительности методов, использующих пространственные модели относительно выбора типа пространственной модели и числа ближайших соседей.

## Методика оценки эффективности КМО

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

1. Выбор методов для сравнения их эффективности. Для сравнения были выбраны следующие методы.

- Метод (М1) линейной регрессии, модель оценивания:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon, \tag{1}$$

где  $Y$  — цена объекта недвижимости,  $(X_1, \dots, X_n)$  — непространственные характеристики объекта недвижимости,  $\beta_i$  — численные коэффициенты при непространственных характеристиках объекта, показывающие вклад соответствующих характеристик в стоимость,  $\varepsilon$  — случайная составляющая.

- 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:

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

где  $Y$  —  $n \times 1$ -вектор зависимых переменных;  $X$  —  $n \times k$ -матрица непространственных характеристик объекта недвижимости;  $\rho$  — коэффициент пространственной зависимости;  $\beta$  —  $k \times 1$ -вектор коэффициентов при непространственных характеристиках объекта;  $\varepsilon$  — случайная составляющая;  $W$  —  $n \times n$ -матрица учета пространственных взаимодействий, элементы  $w_{ij}, i, j = 1, \dots, n$  которой формируются по правилу

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

б) модели с пространственной корреляцией в ошибках:

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

где  $u$  —  $n \times 1$ -вектор возмущений,  $\lambda$  — 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 [Беляева, Гребенюк, 2014], 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.

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.

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:

- критерии 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 ( $K_2$ )

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

где  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ ,  $\bar{Y}_m = \frac{1}{n} \sum_{i=1}^n Y_i^m$  — 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 ( $K_4$ ):

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

- 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  $\alpha$  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

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

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.

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

$$\begin{aligned} \hat{Y}_k &= \rho W_{pc} Y_l + X_k \beta \text{ для модели (2),} \\ \hat{Y}_k &= X_k \beta + \lambda W_{pc} (Y_l - X_l \beta) \text{ для модели (3),} \end{aligned}$$

где  $X_k$  — значения регрессоров контрольной выборки,  $Y_l = (y_1, \dots, y_n)$  и  $X_l$  — значения цен и регрессоров обучающей выборки, коэффициенты  $\beta, \lambda, \rho$  оценены по обучающей выборке.

По результатам анализа значений критериев для каждой модели формирования цен определяется метод, являющийся наиболее эффективным для данной модели. Остальные проверяемые методы ранжируются в порядке убывания их эффективности.

4. В результате анализа поведения критериев для различных моделей формирования цен формируются группы моделей (которые 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.

## 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 [Беляева, Гребенюк, 2014].

Построение модели включает следующие шаги.

1. Генерация объектов. На плоскости  $[0,100] \times [0,100]$  генерируются координаты 150 объектов (из них 100 — объекты обучающей 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 объекта до центра влияния; для учета воздействия центра влияния на цену выбирается функция вида

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

где параметр  $\alpha$  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  $i$ -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  $100 \times 100$  для объектов обучающей выборки, где  $i = 2, 4, 8, 12$ . Для объектов контрольной выборки строятся матрицы  $W_{ic}$  размера  $50 \times 100$ , 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  $\rho$  и  $\lambda$  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  $\alpha$  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).

## 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, не учтенными регрессионной моделью;

Ситуация 4. Слабая пространственная автокорреляция между факторами, не учтенными 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, различающиеся 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 рисунках 1 и 2 представлены графики значений критерия, оценивающего качество построенной модели — среднеквадратичного отклонения для различных Ситуаций: сплошной линией — для случая одного центра влияния, пунктирной линией — для случая трех центров влияния. По оси ординат указано значение среднеквадратичного отклонения, 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).

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.

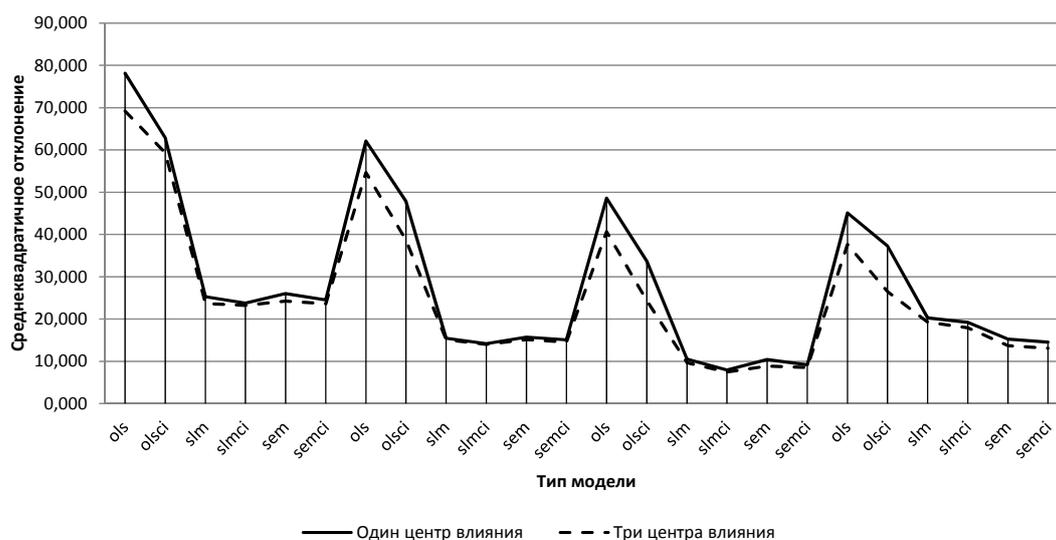


Рис. 1. Среднеквадратичное отклонение для Ситуации 1

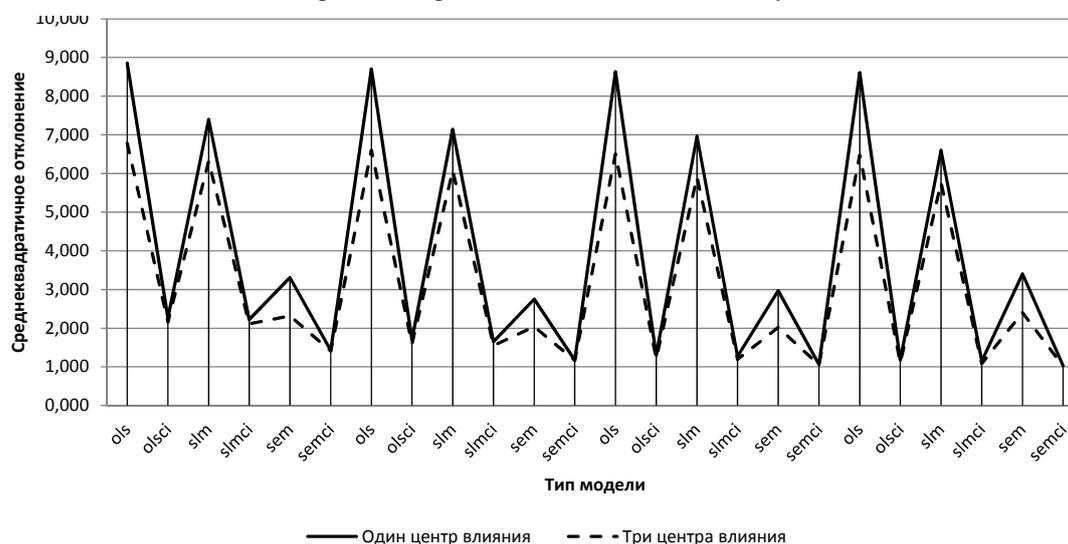


Рис. 2. Среднеквадратичное отклонение для Ситуации 3

Из графиков видно, что в рамках одной ситуации поведение значения критерия  $K_4$  для одного и трех центров влияния идентично, но отличается для разных ситуаций.

Это подтверждается соответствующими расчетами.

В Ситуации 1:

- методы М4 и М3 с учетом пространственной автокорреляции во всех ситуациях являются более эффективными по критерию среднеквадратичного отклонения, чем методы М1 и М2;
- при использовании метода М4 значения критерия «среднеквадратичное отклонение» ниже на 3–24 % его значений, полученных при использовании метода М3;
- пространственные модели уменьшают значение критерия метода М2 в 2–3 раза.

В Ситуации 3:

- методы с учетом центров влияния М2 и М4 во всех экспериментах являются более эффективными, чем методы М1 и М3: среднеквадратичная ошибка уменьшается на 68–88 %;
- метод М4 и метод М2 в данной ситуации силы пространственной зависимости по значению критерия «среднеквадратичное отклонение» не являются равнозначными. Так, метод М4 позволяет уменьшить среднеквадратичную ошибку на 8–37 %.

Таблица 1. Применение методов

Ситуация	Качество прогноза	Диагностика модели			ИТОГ
		Критерии Акаике (КА)	Корреляция модельных и реальных цен (КК)	Среднеквадратичное отклонение остатков (СКО)	
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 КА. 2) М3 по этому критерию значительно эффективнее М1 и М2.	1) М3 и М4 имеют КК выше, чем М1 и М2. 2) М4 и М3 по значению КК не различаются.	1) М3 и М4 имеют СКО остатков меньше, чем М1 и М2. 2) М4 имеет значение СКО остатков на 3% - 24% ниже, чем М3.	Наилучший 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. 4) М4 с SEM моделью-ошибка прогноза меньше, чем по М2 на 13%-43%.	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 КА М1 на 28% - 55%, М3 – на 5% - 35%. 2) В свою очередь, М4 уменьшает значение КА на 24% - 45% по сравнению с М2.	1) М2 и М4 имеют КК выше, чем М1 и М3. 2) М4 и М2 по значению КК не различаются.	1) М2 уменьшает значение СКО М1 на 49% - 73%, М3 – для усредненного по двум методам $\text{slm}$ и $\text{sem}$ на 12% - 44%. Но в экспериментах 5, 6, 21, 22, 23 модель $\text{sem}$ обладает качеством, лучше, чем М2 на 3%-15%. 2) М4 уменьшает значение СКО М2 на 30% - 49%. При этом для большого числа соседей (12 соседей, эксперименты 8 и 24) лучшей с точки зрения этого критерия является модель $\text{sem}$ .	Наилучший метод – М4 с SEM моделью для числа соседей больше двух, М4 с SLM моделью – для числа соседей, равных двум
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 КА на 38%-90% по сравнению с М1 и М3. 2) Причем для такого варианта пространственной зависимости лучшим является метод $\text{sem}$ . М4 позволяет уменьшить значение КА М2 на 28% - 60%.	1) М2 и М4 имеют КК выше, чем М1 и М3: на 2% - 7%. 2) М4 и М2 по значению КК не различаются.	1) М2 и М4 имеют СКО остатков на 68% - 88% меньше, чем М1 и М3. 2) М4 уменьшает значение СКО остатков М2 на 8% - 37%.	М4 с SEM моделью лучше М2 и значительно лучше М3
Ситуация 4	1) М2 и М4 – ошибка прогноза меньше, чем по М1 и М3. 2) М4 и М2 по критерию “ошибка прогноза” одинаково эффективны. 3) М3– ошибка прогноза увеличивается в 16-29 раз по сравнению с ошибкой прогноза по М2.	1) Методы с учетом центров влияния М2 и М4 во всех экспериментах являются более эффективными, чем М1 и М3: значение КА уменьшается на 92% - 97%. 2) М2 в случае слабой пространственной зависимости в остатках модели дает более низкие значения КА по сравнению с М4 на 5%-20%.	1) М2 и М4 имеют КК выше, чем М1 и М3: на 2% - 7%. 2) М4 и М2 по значению КК не различаются.	1) М2 и М4 уменьшают СКО остатков М1 и М3 на 80% - 89%. 2) М4 и М2 по значению СКО не различаются.	М2 наилучший метод

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 качество моделей и прогнозные характеристики 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.

## 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 и числа ближайших соседей. 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 тип 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

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

Рассмотрим критерий  $K_5$ . Положительное значение  $\alpha$  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. На рисунке 3 представлены значения  $\alpha$  для метода М4 (сплошная линия), результаты метода М3 (пунктир). По оси ординат отложены значения  $\alpha$ , по оси абсцисс — номера экспериментов.

По критерию Стьюдента для парных выборок значения  $\alpha$  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 использова-

нии метода М4. В остальных случаях метод М4 не является чувствительным к выбору типа модели.

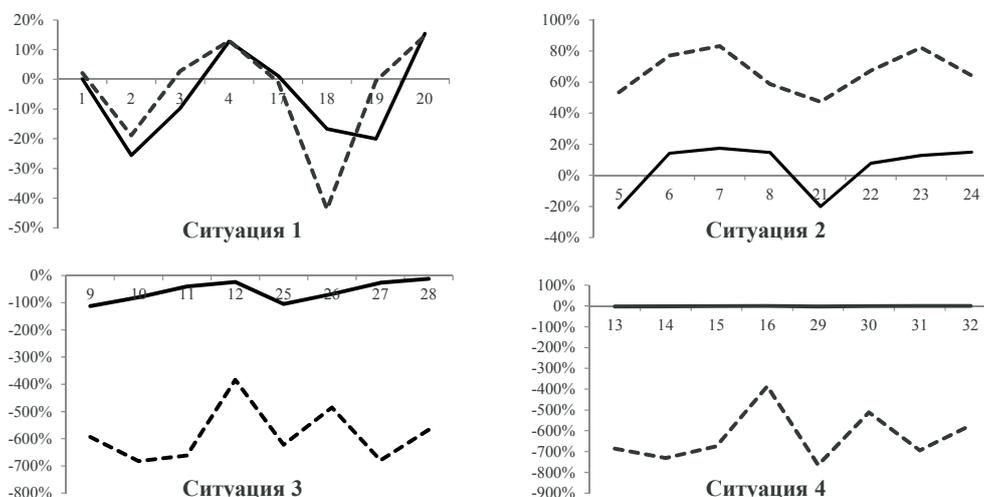


Рис. 3. Чувствительность методов

Как видно из графиков, М4 в Ситуациях 2–4 обладает меньшей чувствительностью к выбору типа модели, 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.

## Заключение

Проведена апробация методики сравнения на определенном наборе методов КМО. 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.

## Список литературы

- Беляева А. В., Гребенюк Е. А. Построение моделей массовой оценки объектов недвижимости с учетом пространственной зависимости // Пробл. управл. — 2014. — № 1. — С. 45–52.
- Стерник Г. М. Математические основы методологии построения дискретных пространственно-параметрических моделей рынка жилья // Рынок недвижимости России. — 2005.

- Эккерт Дж.* Организация оценки и налогообложения недвижимости: в 2 т. / Дж. Эккерт, Р. Дж. Глаудеманс, Р. Р. Олми. — М.: Рос. общество оценщиков, 1997. — Т. 1–2.
- Anselin L.* Spatial Econometrics: Methods and Models. — Dordrecht: Kluwer Academic, 1988a. — P. 284.
- Anselin L.* Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity // Geographical Analysis. — 1988b. — Vol. 20. — P. 1–17.
- Cliff A., Ord J. K.* Spatial Processes: Model and Application. — London: Pion, 1981. — P. 327.
- 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.