УДК 339.13.017

© М. Б. Ласкин, канд. физ.-мат. наук, ст. науч. сотр.
(Санкт-Петербургский институт информатики
и автоматизации Российской академии наук
(СПИИРАН), Санкт-Петербург, Россия)
E-mail: laskinmb@yahoo.com

© Л. Б. Дампилов, студентка
(Санкт-Петербургский государственный
университет, Санкт-Петербург, Россия)
E-mail: dampilonlyudmila@gmail.com

DOI 10.23968/1999-5571-2020-17-3-208-220

© М. Б. Laskin, PhD in Sci. Phys.-Math.,
senior researcher
(St. Petersburg Institute for Informatics
and Automation of the Russian Academy
of Sciences (SPIIRAS), St. Petersburg, Russia)
E-mail: laskinmb@yahoo.com

© L. B. Dampilon, student
(Saint Petersburg State University,
St. Petersburg, Russia)
E-mail: dampilonlyudmila@gmail.com

КОРРЕКТИРОВКИ ОБЪЕКТОВ СРАВНЕНИЯ ОБРАБОТКОЙ БОЛЬШИХ МАССИВОВ ДАННЫХ КАДАСТРОВОГО УЧЕТА

ADJUSTMENT OF COMPARISON OBJECTS BY PROCESSING OF LARGE AMOUNTS OF CADASTRAL DATA

Предложен метод корректировок объектов сравнения в сравнительном подходе при оценке недвижимости, основанный на сравнении кластеров различных групп объектов недвижимости, сформированных по ценообразующим факторам и темпам изменений за период между датами кадастровой оценки. Ценообразующие факторы разделяются на качественные и непрерывные. Разделение объектов на кластеры проводится по темпам роста в период между кадастровой оценкой и по качественным факторам. Затем в каждом кластере рассматриваются двумерные распределения кадастровых стоимостей и вытекающие из них зависимости изменившейся кадастровой стоимости от непрерывных ценообразующих факторов. Предложенный метод позволяет проводить корректировки объектов сравнения рыночных данных (в том числе и для малых выборок), опираясь на данные кадастрового учета двух периодов. Приведенные примеры опираются на сравнение результатов кадастровой оценки жилой недвижимости Санкт-Петербурга в 2015 и 2018 годах.

Ключевые слова: рыночная стоимость недвижимости, стохастическая модель ценообразования, мода логарифмически нормального закона распределения, корректировка цен объектов сравнения.

The article proposes an adjustment method of comparison objects in the comparative approach to real estate valuation, based on the comparison of clusters of different groups of real estate objects formed by price-forming factors and the rate of change between the dates of cadastral valuation. Price-forming factors are divided into qualitative and continuous ones. The division of objects into clusters is carried out according to the growth rate in the period between cadastral valuation and to the qualitative factors. Then, in each cluster, two-dimensional distributions of cadastral values and the resulting dependencies of the changed cadastral value on continuous price-forming factors are considered. The proposed method makes it possible to adjust the objects of comparison of market data (including for small samples), based on the data of cadastral accounting of two periods. These examples are based on a comparison of cadastral valuation results of residential real estate in St. Petersburg in 2015 and 2018.

Keywords: real estate market value, stochastic model of pricing, mode of the logarithmically normal law of distribution, the price adjustment of the comparison objects.

В теории и практике оценки недвижимого имущества при применении сравнительного подхода существует ряд проблем.

Первая проблема — выбор объектов сравнения. Оценщики стараются подобрать данные по 5-6 аналогам (бывают и большие, но все

равно малые выборки) как можно более близких по своим характеристикам к объекту оценки. В силу относительной закрытости информации о реальных сделках подбор аналогов часто связан с определенными трудностями. Малый объем выборки в большом мегаполисе часто приво-

дит к существенно разным результатам оценки для одного и того же объекта у разных оценщиков — все зависит от объектов сравнения.

Вторая проблема — корректировки по методу парных сравнений. Практически невозможно подобрать объекты сравнения так, чтобы обеспечить пары объектов, различающихся только по одному ценообразующему фактору. К тому же совершенно необязательно, чтобы разница в цене объяснялась одним этим фактором.

Третья проблема — взвешивание результатов, которое в практике оценки часто не имеет никакого обоснования, кроме мнения оценщика. В настоящей статье рассматривается подход, призванный разрешить указанные трудности с помощью статистического анализа (можно сказать и статистического обучения — в смысле ставшего популярным в последнее время *machine learning*) больших данных¹, которые на рынке недвижимости, несомненно, присутствуют. Существенно важным инструментом для такого анализа могут оказаться данные кадастрового учета, в которых в 2015 году (и уже второй раз в 2018 году) появились сведения о кадастровой стоимости всех введенных в эксплуатацию объектов недвижимости. Становится возможным изучение двумерных распределений данных кадастрового учета в разных периодах, многомерных распределений данных кадастрового учета² и рыночных данных в разных периодах и получении важных характеристик для различных кластеров объектов. Кадастровый номер при этом оказывается уникальным идентификатором, позволяющим объединять разные наборы данных. Несмотря на то что кадастровая стоимость введена прежде всего для целей налогообложения, она может быть использована и для аналитических исследований. В 2015 году³ када-

стровая стоимость определялась как рыночная на дату оценки. В 2018 году с принятием Федерального закона № 237-ФЗ «О государственной кадастровой оценке» от 03.07.2016 из стандарта исчезло указание на связь кадастровой и рыночной стоимости, но оно осталось в методических указаниях Минэкономразвития⁴. В частности, в используемых нами данных отчетов о кадастровой оценке 2015 и 2018 года, опубликованные ГУП ГУИОН Санкт-Петербурга в 2015 г.⁵ и Санкт-Петербургским ГУ «Кадастровая оценка» в 2018 г.⁶, кадастровая стоимость определялась как рыночная на дату оценки.

В качестве примера в статье показано, как можно было бы оценить рыночную стоимость объекта жилого фонда, опираясь на данные кадастрового учета 2015 и 2018 года, выделение кластеров объекта оценки и объектов сравнения, расчета корректировок кластеров целиком и применение полученных корректировок к рыночным данным. Теоретические обоснования для такого подхода были опубликованы ранее в статьях [3, 4, 5]. Мы рассматриваем результаты отчетов о кадастровой оценке⁷ как сборник рыночных цен объектов недвижимости на 01.01.2015 и 01.01.2018. Подробное обоснование расчета кадастровых стоимостей как рыночных дано в отчете об оценке⁶ (том 1 «Отчета», с. 107, 138, 111, п. 4.9.5).

Министерства экономического развития РФ от 22 октября 2010 г. № 508). <https://base.garant.ru/14152457/>

⁴ Приказ Минэкономразвития России от 12.05.2017 № 226 (ред. от 09.08.2018) «Об утверждении методических указаний о государственной кадастровой оценке» (Зарегистрировано в Минюсте России 29.05.2017 № 46860). <http://base.garant.ru/71686152/>

⁵ Приказ Комитета имущественных отношений Санкт-Петербурга от 27.08.2015 № 59-п «Об утверждении результатов определения кадастровой стоимости помещений площадью менее 3000 кв. м на территории Санкт-Петербурга». URL: <http://gov.spb.ru/gov/otrasl/kio/documents/inve-dokumenty/4434/> (дата обращения: 20.09.2019).

⁶ Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга № 1/2018, Санкт-Петербургское государственное бюджетное учреждение «Городское управление кадастровой оценки», Санкт-Петербург, 2018. URL: <http://www.ko.spb.ru/interim-reports/> (дата обращения: 20.09.2019).

⁷ В отчетах, безусловно, есть оценки, с которыми владельцы обоснованно не согласны, обычно такие оценки оспариваются на соответствующей комиссии или в суде, но, в целом, составляют не более 3–4 % от всего массива кадастровых данных. Это довольно внушительное количество, тем не менее мало влияет на общую рыночную картину, которую данные кадастрового учета (по Санкт-Петербургу) в целом отражают вполне адекватно.

¹ Неустоявшийся термин «большие данные» мы понимаем в узком, прикладном смысле: как данные, размер которых не позволяет применять обычные методы анализа, проверки гипотез и т. д.

² Федеральный Закон № 237-ФЗ «О государственной кадастровой оценке» предусматривает проведение государственной кадастровой оценки не чаще, чем один раз в три года (для городов федерального значения один раз в два года), но не реже чем один раз в пять лет, а кроме того, предусматривает возможность внеочередной кадастровой оценки в случаях изменения индекса недвижимости более чем на 30 % (Федеральный закон № 237-ФЗ «О государственной кадастровой оценке» от 03.07.2016 г. http://www.consultant.ru/document/cons_doc_LAW_200504/).

³ Федеральный стандарт оценки «Определение кадастровой стоимости объектов недвижимости (ФСО № 4)» (утв. Приказом

Исходные допущения

1. Рынок недвижимости — это совокупность объектов и система взаимоотношений между участниками рынка и регулятором.

2. Объекты недвижимости могут быть разделены на различные сегменты и кластеры, развитие которых во времени происходит взаимосвязано и обусловлено общими макроэкономическими условиями (доходы населения, доступность кредитов, градостроительная практика и т. д.) в пределах определенного региона.

3. Рынок недвижимости по сравнению с финансовыми рынками отличается замедленной реакцией на внешние изменения условий.

4. Разделение рынка на кластеры с различными темпами роста и ценообразующими факторами позволяет исследовать развитие каждого кластера в отдельности и во взаимосвязи с другими кластерами.

5. Полученные на определенную дату характеристики кластеров можно распространить на некоторый промежуток времени для краткосрочного прогнозирования.

6. Кадастровые базы являются обобщенной оценкой состояния рынка на дату.

Кластеризация

Как известно, рыночная стоимость является функцией ценообразующих факторов:

$$MV = f(x_1, \dots, x_n, y_1, \dots, y_m),$$

где MV — рыночная стоимость (*marketvalue*); $x_1, \dots, x_n$ — качественные факторы (тип дома, этаж и т. п.); $y_1, \dots, y_m$ — количественные факторы (площадь, расстояние до центра и т. п.).

Все объекты могут быть разбиты на кластеры с помощью качественных факторов. Затем в каждом кластере изучаются зависимости рыночной стоимости от количественных факторов.

Таким образом, общая совокупность объектов недвижимости на рынке может быть использована для оценки рыночной стоимости любого объекта недвижимости и прогноза его стоимости в краткосрочном периоде. Для этого сначала исследуются положение объекта оценки и объектов сравнения в общем множестве объектов, его принадлежность к определенному кластеру, темпы роста от базового периода к дате оценки. Проводятся три вида корректировок:

- корректировка на время;

- первая группа корректировок — по качественным факторам, т. е. корректировки кластеров объектов сравнения на кластер объекта оценки;

- вторая группа корректировок — по непрерывным ценообразующим факторам на параметры объекта оценки.

Темп роста в кластере

Пусть имеются данные по кадастровой оценке в двух периодах (например, на 2015 и 2018 г.). Пусть каждый кластер имеет линейный тренд в логарифмах вида

$$\ln(V_{2018}) = b + a \ln(V_{2015}),$$

где V_{2015} , V_{2018} — кадастровые стоимости в 2015 и 2018 г.

Тогда $V_{2018} = \exp(b) V_{2015}^a$, а с другой стороны, по формуле сложных процентов $V_{2018} = (1+i)^3 V_{2015}$, где i — темп роста рынка в год.

Получаем формулу для расчета темпа роста в каждом кластере:

$$1+i = \sqrt[3]{\exp(b) V_{2015}^{a-1}}. \quad (1)$$

Корректировка объектов сравнения на заданный момент времени

Прогноз роста за произвольный период δ (в годах от даты последней кадастровой съемки) рассчитывается по формуле

$$(1+i)^\delta = \sqrt[3]{\exp(b) V_{2015}^{a-1}^\delta}. \quad (2)$$

Для прогноза на период времени δ от даты последней кадастровой съемки определяем тренды в каждом кластере и по нему рассчитываем темп роста. Затем увеличиваем/уменьшаем значения кадастровой стоимости последней кадастровой съемки на мультипликатор $M = (1+i)^\delta$. Получаем прогноз рыночной стоимости на дату оценки по всем объектам кластера (обозначим его V).

Первая группа корректировок

Пусть кластеры, содержащие объект оценки и объект сравнения, имеют линейные тренды в логарифмах (верхние индексы: oo — объект оценки, oc — объект сравнения):

$$\ln(V^{oo}) = b_1 + a_1 \ln(V_{2015}) \quad \ln(V^{oc}) = b_2 + a_2 \ln(V_{2015}).$$

$$\text{Тогда } V^{oo} = \exp(b_1) V_{2015}^{a_1} \text{ и } V^{oc} = \exp(b_2) V_{2015}^{a_2}.$$

Так как можно записать, что

$$V^{oo} = \frac{\exp(b_1)}{\exp(b_2)} (V_{2015})^{a_1 - a_2} V^{oc}, \text{ то коэффициент}$$

корректировки равен

$$K_1 = \frac{\exp(b_1)}{\exp(b_2)} (V_{2015})^{a_1 - a_2}. \quad (3)$$

Вторая группа корректировок

Корректировки по непрерывным факторам проводятся в кластере, сформированном после первой группы корректировок. Предположим, что рыночная стоимость имеет линейный тренд в логарифмах по всем непрерывным факторам. Тогда $\ln(V) = b + a_1 \ln(y_1) + \dots + a_m \ln(y_m)$;

$$V = \exp(b) y_1^{a_1} \cdot \dots \cdot y_m^{a_m}.$$

В этом случае объекты оценки и сравнения отличаются набором ценообразующих факторов $y_1^{oo}, \dots, y_m^{oo}$ и $y_1^{oc}, \dots, y_m^{oc}$, формула для корректировки

$$V^{oo} = K_2 \cdot V^{oc},$$

$$\text{где } K_2 = \left(\frac{y_1^{oo}}{y_1^{oc}} \right)^{a_1} \cdot \dots \cdot \left(\frac{y_m^{oo}}{y_m^{oc}} \right)^{a_m}. \quad (4)$$

Коэффициенты K_1, K_2 являются мультипликаторами, которые применяются к результату корректировки на время данных последнего кадастрового учета по всем кластерам объектов сравнения, кроме того, мультипликатор K_2 применяется к кадастровой стоимости базового периода в целях приведения всех объектов сравнения (теперь уже кластеров целиком) к одинаковым значениям ценообразующих факторов, равных значениям параметров объекта оценки.

Применяем полученные корректировки к рыночным данным по 5–6 объектам сравнения на дату оценки. Получаем оценку рыночной стоимости объекта оценки. Затем строим прогноз и сравниваем результаты.

Прогноз рыночной стоимости на дату оценки

После объединения скорректированных кластеров объекта оценки и объектов сравнения получаем кластер «псевдоодинаковых» объектов, имеющих одинаковые значения учтенных ценообразующих факторов, образующих множество точек на плоскости (V_{2015}, V) .

При подтверждении гипотезы о совместной нормальности логарифмов $\ln(V_{2015}), \ln(V)$ оценка рыночной стоимости проводится по условным распределениям случайной величины V , по известному значению V_{2015} .

Плотность двумерного нормального распределения с нулевым вектором средних при $x, y \in R$ равна

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$$

$$\times \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_x^2} - 2\rho \frac{xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} \right) \right],$$

где (σ_x^2, σ_y^2) — дисперсии случайных величин X, Y ; — коэффициент корреляции компонент X и Y .

Так как условная плотность вектора, имеющего совместную плотность $p(x, y)$, равна

$$\rho_{X|Y=y}(x) = \rho(x|y) = \frac{\rho(x,y)}{\int_{-\infty}^{+\infty} \rho(x,y) dx} = \frac{\rho(x,y)}{\rho(y)},$$

то условная плотность двумерного нормального вектора $f(X|y)$

$$f(X|y) = \frac{1}{\sqrt{2\pi} \cdot \sigma_x \sqrt{1-\rho^2}} \times \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{x}{\sigma_x} - \rho \frac{y}{\sigma_y} \right)^2 \right], \quad (5)$$

что означает: при каждом y случайная величина X нормальна с параметрами

$$E(X|Y=y) = \rho \frac{\sigma_x}{\sigma_y} y, D(X|Y=y) = \sigma_x^2 (1-\rho^2).$$

Рассмотрим двумерную случайную величину (V_{2015}, V) с совместно нормальным распределением логарифмов с параметрами $\mu_{2015}, \sigma_{2015}, \mu, \sigma$, коэффициентом корреляции ρ .

Пусть $X = \ln V - \mu, Y = \ln V_{2015} - \mu_{2015}$, тогда подставляя X и Y в (5), получим условную плотность $\ln V$ при условии $\ln V_{2015} = \ln v_{2015}$:

$$f(\ln V | \ln V_{2015} = \ln v_{2015}) = \frac{1}{\sqrt{2\pi} \cdot \sigma \sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \times \left(\frac{\ln V - \mu}{\sigma} - \rho \frac{\ln v_{2015} - \mu_{2015}}{\sigma_{2015}} \right)^2 \right],$$

т. е. $\ln V$ (при условии $\ln V_{2015} = \ln v_{2015}$) нормален с параметрами

$$E(\ln V | \ln V_{2015} = \ln v_{2015}) = \mu + \rho \frac{\sigma}{\sigma_{2015}} (\ln v_{2015} - \mu_{2015}),$$

$$D(\ln V | \ln V_{2015} = \ln v_{2015}) = \sigma^2 (1-\rho^2)$$

и V распределено логарифмически нормально с теми же параметрами⁸.

⁸ Следует обратить внимание, что условная дисперсия зависимой переменной не зависит от дисперсии регрессора и при значениях ρ , близких к единице, будет достаточно малой.

Таблица 1

Пример: данные по цене продаж шести объектов сравнения

Объекты	Адрес, СПб.	Цена продажи
Объект оценки	Улица Рентгена, дом XX	
Объект сравнения	Улица Профессора Попова, дом XX	7 400 000 Р
Объект сравнения	Улица Куйбышева, дом XX	15 500 000 Р
Объект сравнения	Большая Монетная улица, дом XX	7 900 000 Р
Объект сравнения	Улица Всеволода Вишневского, дом XX	5 100 000 Р
Объект сравнения	Съезжинская улица, дом XX	4 500 000 Р
Объект сравнения	Малый проспект П.С., дом XX	13 100 000 Р

Условные медиана, мода, математическое ожидание величины V имеют вид степенных функций:

$$\text{Median}(V | V_{2015} = v) = A_{\text{median}} v^{\frac{\rho \sigma}{\sigma_{2015}}}, \quad (6)$$

$$\text{где } A_{\text{median}} = \exp\left(\mu - \rho \frac{\sigma}{\sigma_{2015}} \mu_{2015}\right),$$

$$\text{Mode}(V | V_{2015} = v) = A_{\text{mode}} v^{\frac{\rho \sigma}{\sigma_{2015}}}, \quad (7)$$

$$\text{где } A_{\text{mode}} = \exp\left(\mu - \rho \frac{\sigma}{\sigma_{2015}} \mu_{2015} - \sigma^2 \left(1 - \rho^2\right)\right),$$

$$E(V_{2018} | V_{2015} = v) = A_E v^{\frac{\rho \sigma}{\sigma_{2015}}}, \quad (8)$$

$$\text{где } A_E = \exp\left(\mu - \rho \frac{\sigma}{\sigma_{2015}} \mu_{2015} + \frac{1}{2} \sigma^2 \left(1 - \rho^2\right)\right).$$

Уравнение главной оси эллипсов рассеяния

$$\ln(V) = \rho \frac{\sigma}{\sigma_{2015}} \ln(v_{2015}) + \mu - \rho \frac{\sigma}{\sigma_{2015}} \mu_{2015} \quad (9)$$

при потенцировании дает уравнение зависимости медианных значений случайной величины V при заданном значении $V_{2015} = v_{2015}$.

Пример

Предположим, требуется оценить рыночную стоимость объекта оценки на основании данных о продажах шести объектов жилой недвижимости в период сентябрь-октябрь 2019 г. Данные сведены в табл. 1.

Для определения корректировок по качественным факторам по кадастровым базам 2015 и 2018 г. находим необходимые качественные и количественные факторы объекта оценки и объектов сравнения (поиск легко выполняется по кадастровому номеру) (рис. 2).

Качественными переменными в этом примере являются тип дома, количество комнат, этаж;

Таблица 2

Значения ценообразующих факторов, кадастровые стоимости для объекта оценки и объектов сравнения

Кадастровый номер	КС2015 кв. м	КС2018 кв. м	Адрес, СПб.	Площадь, кв. м	Centre	Rooms	Type_dom	Etag_kv
78:07:XXXXXXX:XXXX	101 368,48		Улица Рентгена, дом XX	80,90	3 310,20	3	КИРП_5эт	2
78:07:XXXXXXX:XXXX	105 796,58	122 420,47	Улица Профессора Попова, дом XX	66,80	3 864,20	3	СФ_КР	2
78:07:XXXXXXX:XXXX	94 737,31	99 411,43	Улица Куйбышева, дом XX	134,90	2 306,80	6	СФ	2
78:07:XXXXXXX:XXXX	123 803,56	125 551,80	Большая Монетная улица, дом XX	78,60	2 972,10	1	СТАЛ	2
78:07:XXXXXXX:XXXX	142 411,81	142 585,04	Улица Всеволода Вишневского, дом XX	46,80	3 590,20	1	КИРП_совр.	3
78:07:XXXXXXX:XXXX	132 072,04	133 687,54	Съезжинская улица, дом XX	36,70	2 049,00	1	ИНД	2
78:07:XXXXXXX:XXXX	117 003,01	129 814,86	Малый проспект П.С., дом XX	115,90	2 370,80	4	КИРП_совр.	2

количественными — площадь, расстояние до центра города⁹.

Приведенное ниже исследование построено следующим образом. Сначала в качестве рыночных стоимостей объектов жилой недвижимости через 3 года от даты кадастровой оценки выбираются данные кадастровой оценки 2018 года. Кадастровая стоимость объекта оценки в 2018 году не рассматривается, так как основной задачей этой части исследования является расчет корректировок первой и второй группы. Значения кадастровых стоимостей в кластерах объекта оценки и объектов сравнения корректируются на мультипликатор (см. формулу (2)) для $\delta = 22/24$ (корректировка на время оценки). При этом будет получена оценка рыночной стоимости объекта оценки (независимо от того, какое значение кадастровой стоимости было присвоено объекту оценки во время последней кадастровой съемки). Смысл такого результата — получить оценку как результат общей направленности рынка за трехлетний кадастровый период. Мы исходим из предположения, что общее развитие рынка, наблюдаемое по всем объектам одновременно, «диктует» объекту оценки его рыночную стоимость. Полученные корректировки применяются к заданным заранее ше-

сти объектам сравнения. Результат сравнивается с прогнозным, полученным по условным распределениям, выбирается интервальная оценка на основе доверительных интервалов.

Соотношение кадастровых стоимостей 2015 и 2018 года для всей жилой недвижимости Санкт-Петербурга представлено на рис. 1. Объем выборки 2 025 757 объектов (объекты, имеющие совпадения кадастрового номера в 2015 и 2018 годах).

Очевидно, что все множество объектов жилой недвижимости в Санкт-Петербурге делится на 5–7 кластеров, демонстрирующих разную динамику в указанном трехлетнем периоде. Объект оценки и объекты сравнения относятся к самому большому кластеру, выделенному серым цветом. Этот кластер также может быть разделен на кластеры, но уже не по «геометрическому» принципу, а по ценообразующим факторам. Пример такого разбиения показан на рис. 2.

Поскольку для объекта оценки и объектов сравнения кластеры определяются набором значений ценообразующих качественных факторов, из табл. 1 легко заметить, что все семь объектов находятся в разных кластерах, каждый из которых находится внутри кластера, выделенного на рис. 1 серым цветом, и полностью определяемых по набору значений качественных факторов: тип дома–количество комнат–этаж. Введением фиктивных переменных, принимающих бинарные значения 0,1 по каждому значению качественных

⁹ В кадастровых базах имеются и другие факторы. Для приведенного в статье примера мы ограничились только указанными пятью как бесспорными факторами, влияющими на рыночную стоимость.

Рис. 1. Двумерное распределение кадастровых стоимостей 1 кв. м объектов жилой недвижимости (отдельные квартиры) в Санкт-Петербурге в 2015 и 2018 годах. По горизонтали — кадастровая стоимость 1 кв. м в руб. в 2015 году, по вертикали — в 2018 году

Рис. 2. Пример разбиения кластера на «подкластеры» по ценообразующим факторам.
«Серый» кластер, двумерное распределение кадастровых стоимостей 1 кв. м
в Санкт-Петербурге в 2015 и 2018 годах. По горизонтали — кадастровая стоимость
1 кв. м в руб. в 2015 году, по вертикали — в 2018 году

Рис. 3. Кластер объекта оценки (выделен черным цветом) и кластеры объектов сравнения (цвета).
Двумерное распределение кадастровых стоимостей 1 кв. м в Санкт-Петербурге в 2015 и 2018
годах. Слева: по горизонтали — кадастровая стоимость 1 кв. м в руб. в 2015 году по вертикали —
в 2018 году; справа в логарифмах от соответствующих кадастровых стоимостей

факторов и применением стандартных процедур иерархического кластерного анализа получаем семь кластеров, необходимых для дальнейшего исследования. Результат показан на рис. 3.

Кластер объекта оценки показан черным цветом. Корректируем все кластеры на время, применяя мультипликатор $M = (1+i)^{\delta}$ (значение мультипликатора для каждого объекта зависит от его кадастровой стоимости в базовом периоде 2015 года (см. формулу (2)). Параметры линейных трендов логарифмов для кластеров объекта сравнения и объектов оценки приведены в табл. 3.

Таблица 3

**Параметры линейных трендов логарифмов
для расчета корректировки на время оценки**

Кластеры	Коэффициенты тренда	
	b	a
Кластер объекта оценки	- 0,4838	1,1066
Кластер объекта сравнения 1	- 0,4083	1,1039
Кластер объекта сравнения 2	0,1919	0,9779
Кластер объекта сравнения 3	- 0,0294	1,0057
Кластер объекта сравнения 4	- 0,3453	1,0788
Кластер объекта сравнения 5	- 0,3883	1,0811
Кластер объекта сравнения 6	1,3910	0,7150

На рис. 4 показаны откорректированные на время кластеры объекта оценки и объектов сравнения.

По первой группе корректировок кластер объекта оценки (черный) не корректируется. Ко всем остальным кластерам применены корректировки по формуле (3). Значения соответствующих параметров, и оценки качества линейной регрессии логарифмов¹⁰ показаны в табл. 4.

После первой группы корректировок кластеры расположились компактнее, результат показан на рис. 5.

¹⁰ *** Стандартная подсказка R означает p-value t-теста < 0,001 при проверке гипотезы о незначимом отличии от нуля параметра (аналогично **p-value t-теста < 0,01 *p-value t-теста < 0,05, т. е. уровни значимости, на которых отвергается нулевая гипотеза и принимается альтернативная: коэффициенты значимо отличаются от нуля).

Применяем вторую группу корректировок по формуле (4). Значения параметров линейных трендов и критерии качества множественной линейной регрессии приведены в табл. 5 и 6.

После второй группы корректировок разброс в объединенном множестве скорректированных кластеров уменьшился. Результат показан на рис. 6.

Исследуем полученную совокупность откорректированных кластеров на совместную нормальность. На рис. 7 показана совокупность откорректированных кластеров (в логарифмах) и эллипсы рассеяния, соответствующие различным доверительным уровням в предположении о совместном нормальном распределении логарифмов. Эллиптические облака рассеяния явля-

Рис. 4. Кластер объекта оценки и кластеры объектов сравнения после корректировки на дату оценки. Слева: по горизонтали — кадастровая стоимость 1 кв. м в руб. в 2015 году, по вертикали — прогноз на дату оценки; справа — в логарифмах от соответствующих величин

Таблица 4

Расчетные параметры для определения корректировок, корректировки и параметры оценки качества линейной регрессии по первой группе факторов (качественные факторы)

Параметры	Кластер ОО	Кластер ОС № 1	Кластер ОС № 2	Кластер ОС № 3	Кластер ОС № 4	Кластер ОС № 5	Кластер ОС № 6
a	1,1718	1,1673	0,9644	1,0091	1,1270	1,1307	0,5408
b	- 0,7795	- 0,6578	0,3092	- 0,0474	- 0,5564	- 0,6256	2,2411
Значимость коэффициентов *	***	***	***	***	***	***	***
RSE	0,047	0,045	0,048	0,047	0,068	0,038	0,058
Степени свободы	22 251	18 260	5 498	2 298	2 610	18 719	12 481
R^2	0,91	0,90	0,77	0,88	0,77	0,93	0,37
p-value F-теста	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16
Корректировка K1 коэффициента A	0,00000	0,00444	0,20734	0,16266	0,04480	0,04106	0,63094
Корректировка K1 коэффициента B	1	0,88540	0,33665	0,48089	0,80001	0,85732	0,04877

Рис. 5. Результат корректировки по первой группе факторов. Кластер объекта оценки (выделен черным цветом) и кластеры объектов сравнения (цвета). Двумерное распределение кадастровых стоимостей 1 кв. м в Санкт-Петербурге в 2015 году и прогнозное значение на октябрь 2019 года. Слева: по горизонтали — кадастровая стоимость 1 кв. м в руб. в 2015 году, по вертикали — прогноз на октябрь 2019 года; справа — в логарифмах от соответствующих величин

Таблица 5

Параметры линейных трендов в кластерах объекта оценки и объектов сравнения для кадастровой стоимости в базовом периоде (2015 г.)

Параметры	Кластер ОО	Кластер ОС № 1	Кластер ОС № 2	Кластер ОС № 3	Кластер ОС № 4	Кластер ОС № 5	Кластер ОС № 6
b	5,5558	5,2700	4,7849	5,3323	6,0655	6,2323	5,0703
Значимость коэффициента	***	***	***	***	***	***	***
a_1	- 0,0601	- 0,0353	0,0181	0,0324	- 0,0389	- 0,1040	0,0505
Значимость коэффициента	***	***	***	***	***	***	***
a_2	0,0964	- 0,0749	- 0,0282	- 0,0822	- 0,1337	- 0,1297	- 0,0830
Значимость коэффициента	***	***	***	***	***	***	***
RSE	0,1065	0,0908	0,0870	0,1102	0,0875	0,0877	0,0730
Степени свободы	22 250	18 259	5 497	2 297	2 609	18 718	12 480
R^2	0,285	0,412	0,072	0,250	0,356	0,475	0,208
p-value F-теста	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16
Показатель степени к корректировке на площадь	- 0,0601	- 0,0353	- 0,0181	0,0324	- 0,0389	- 0,1040	0,0505
Показатели степени к корректировке на расстояние	- 0,0964	- 0,0749	- 0,0282	- 0,0822	- 0,1337	- 0,1297	- 0,0830

ются признаком, указывающим на возможную совместную нормальность.

Особенностью полученной двумерной выборки является ее размер — 82 131 элемент. Обычные статистические тесты при такой размерности выборки не применимы из-за «проклятия размерности» [6]. Тестирование проводилось следующим способом: из выборки случайным образом выбирались «подвыборки» по 100 элементов, к полученной подвыборке применялся тест Колмогорова–Смирнова, процедура случайной генерации повторялась 100 раз.

Результат проверки маргинальных распределений на нормальность. На рис. 8 показаны распределения p-value для величин $\ln(V_{2015})$ и $\ln(V_{2018})$ на промежутке $[0, 1]$. При этом для $\ln(V_{2015})$ $\min(p\text{-value}) = 0,2455479$, $\max(p\text{-value}) = 0,9981777$, для $\ln(V_{2018})$ $\min(p\text{-value}) = 0,08266929$, $\max(p\text{-value}) = 0,995018$.

Проверка на совместную нормальность. В прикладных статистических пакетах до недавнего времени отсутствовали тесты на нормальность многомерных распределений. Возможно, первой работой в этом направлении является

Таблица 6

Параметры линейных трендов в кластерах объекта оценки и объектов сравнения для прогноза цен на дату оценки

Параметры	Кластер ОО	Кластер ОС № 1	Кластер ОС № 2	Кластер ОС № 3	Кластер ОС № 4	Кластер ОС № 5	Кластер ОС № 6
b	5,5951	5,3702	4,6798	5,6489	6,7550	6,6047	5,5329
Значимость коэффициента	***	***	***	***	***	***	***
a_1	- 0,0204	- 0,0124	0,0265	0,0451	- 0,0557	- 0,1108	0,0787
Значимость коэффициента	*	**	***	***	***	***	***
a_2	- 0,1201	- 0,1017	- 0,0475	- 0,1201	- 0,1990	- 0,1650	- 0,1482
Значимость коэффициента	***	***	***	***	***	***	***
RSE	0,1287	0,1022	0,1086	0,1258	0,1073	0,1014	0,0910
Степени свободы	22 250	18 259	5 497	2 297	2 609	18 718	12 480
R^2	0,309	0,509	0,129	0,353	0,449	0,523	0,345
p-value F-теста	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16	2,20E-16
Показатель степени к корректировке на площадь	- 0,0204	- 0,0124	0,0265	0,0451	- 0,0557	- 0,1108	0,0787
Показатели степени к корректировке на расстояние	- 0,1201	- 0,1017	- 0,0475	- 0,1201	- 0,1990	- 0,1650	- 0,1482

Рис. 6. Результат корректировки по второй группе факторов. Объединение скорректированных кластеров. Двумерное распределение кадастровой стоимости 1 кв. м в Санкт-Петербурге в 2015 и прогноз на октябрь 2019 года. Слева: по горизонтали — кадастровая стоимость 1 кв. м в руб. в 2015 году, по вертикали — прогноз на октябрь 2019 г.; справа — в логарифмах от соответствующих величин. Зеленый цвет — общая совокупность кластеров объекта оценки и объектов сравнения до корректировок, красный цвет — общая совокупность кластеров после корректировки на время и первой корректировки, желтый — общая совокупность кластеров после трех корректировок

разработка такого теста, описанная в работе [7], и включение его в библиотеку статистического пакета R [8]. Мы применили более простой способ, который, однако, можно применить только в двумерном случае. Применялся «круговой» тест Колмогорова–Смирнова [9,10] для подвы-

борок размером 100. Результат 100-кратного повторения показан на рис. 9.

Пересечение линии p-value с линией критического уровня наблюдается в области углов поворота, близких к углу наклона главной оси эллипса рассеяния и в большей степени относится к особенностям работы статистического теста

Рис. 7. Совокупность откорректированных кластеров, эллипсы рассеяния, соответствующие разным доверительным уровням по мере уменьшения 0,9; 0,8; 0,7; 0,6)

при малой дисперсии. В целом в результате тестирования мы сохраняем как рабочую гипотезу о совместной нормальности логарифмов.

Оценка рыночной стоимости объекта оценки. Применение полученных корректировок к ценам продаж

Применим полученные корректировки к ценам продаж объектов сравнения. Результаты по-

следовательных корректировок цен продаж показаны на рис. 10.

При расчете корректировок кластеров объектов сравнения по 1-й и по 2-й группам объекты сравнения находились внутри своих кластеров. Объекты сравнения, таким образом, представляют собой малую выборку, извлеченную из логарифмически нормальной совокупности. Точечной оценкой рыночной стоимости объекта оценки является наиболее вероятное значение $Mode(V) = 103,299$ руб. за 1 кв. м. Стандартное отклонение $\sigma = 0,043$. Интервальная оценка с надежностью 68,2 % (интервал $\pm \sigma$ в логарифмах) $[99,137; 107,838]$ (в тыс. руб./кв. м). Интервальная оценка с надежностью 95,4 % (интервал $\pm 2\sigma$ в логарифмах) $[94,965; 112,787]$ (в тыс. руб./кв. м).

Прогноз рыночной стоимости объекта оценки по корректировкам кластеров объекта оценки и объектов сравнения, учитывающих данные кадастрового учета 2015 и 2018 г.

Для скорректированной выборки получены следующие значения средних, стандартных отклонений и корреляции: $\mu_{2015} = 4,5907$,

Рис. 8. Распределения p-value для величин $\ln(V_{2015})$ (слева) и прогноза на октябрь 2019 г. (справа) на промежутке $[0, 1]$ после 100-кратного повторения теста Колмогорова–Смирнова

Рис. 9. 100-кратное повторение «кругового» теста Колмогорова–Смирнова для двумерной случайной выборки. По горизонтали — угол поворота системы координат, по вертикали — значение p-value для величины $\ln(V_{2018})$. Пунктир — критическое значение p-value = 0,05

	Цена продажи (руб.)	S (Площадь)	Расстояние до центра	КС2015/кв.м	Цена продажи 1 кв.м. (тыс.руб.)	Коэффициент "а" 1 группы корректировок	Коэффициент "b" 1 группы корректировок	Результат корректировок по 1 группе	Коэффициент "a1" 2 группы корректировок	Коэффициент "a2" 2 группы корректировок	Результат корректировок по 2 группе	Логарифмы скорректи- рованных цен
Объект оценки		80,9	3 310	101,368								
Объект сравнения 1	7 400 000	66,8	3 864	105,797	110,778	0,004	0,885	100,134	- 0,012	- 0,102	101,480	4,6199
Объект сравнения 2	15 500 000	134,9	2 307	94,737	114,900	0,207	0,337	99,383	0,027	- 0,048	96,375	4,5682
Объект сравнения 3	7 900 000	78,6	2 972	123,804	100,509	0,163	0,481	105,841	0,045	- 0,120	104,617	4,6503
Объект сравнения 4	5 100 000	46,8	3 590	142,412	108,974	0,045	0,800	108,868	- 0,056	- 0,199	107,321	4,6758
Объект сравнения 5	4 500 000	36,7	2 049	132,072	122,616	0,041	0,857	128,461	- 0,111	- 0,165	108,732	4,6889
Объект сравнения 6	13 100 000	115,9	2 371	117,003	113,028	0,631	0,049	111,237	0,079	- 0,148	102,915	4,6339

Среднее логарифмов	4,640
Стандартное отклонение логарифмов	0,043
Мода скорректированных цен	103,299
Медиана скорректированных цен	103,493
Среднее скорректированных цен	103,590

Рис. 10. Применение корректировок кластеров к ценам продаж объектов сравнения

$\mu_{2018} = 4,6116$, $\sigma_{2015} = 0,1406$, $\sigma_{2018} = 0,1690$, $\rho = 0,9497$. Подставляя полученные значения в формулы (6), (7), (8) получаем оценки условных математических ожиданий и дисперсии логарифмов рыночной стоимости объекта оценки на дату оценки (октябрь 2019):

$$\begin{aligned}
 E(\ln V) &= 4,6116 + \\
 &+ 0,9497 \frac{0,1690}{0,1406} (\ln v_{2015} - 4,5907) = \\
 &= -0,6448 + 1,1415 \cdot \ln v_{2015}, \\
 D(\ln V) &= 0,1690^2 (1 - 0,9497^2) = 0,0028, \\
 \sigma &= \sqrt{0,0027} = 0,053. \quad (10)
 \end{aligned}$$

В предположении, что V распределено логарифмически нормально с теми же параметрами¹¹, условные медиана, мода, математическое ожидание величины V имеют вид степенных функций:

$$\begin{aligned}
 Median(V) &= 0,5332v^{1,1415}, \quad Mode(V) = 0,5317v^{1,1415}, \\
 E(V) &= 0,5340v^{1,1415}.
 \end{aligned}$$

Нетрудно заметить, что в силу малого стандартного отклонения оценка моды отличается от оценки математического ожидания на мультипликатор 0,996, т. е. примерно на 0,4 %. С учетом обычной оценочной практики любая из приведенных числовых характеристик может быть выбрана как оценка рыночной стоимости. Оценка рыночной стоимости объекта на октябрь 2019 г. равна $Mode(V) = 0,5317v^{1,1415}$, где v — кадастровая стоимость объекта оценки в 2015 году (см. табл. 2, $v = 101\,368,48$ руб./кв. м $\sim 101,368$ тыс. руб./кв. м). При проведении вто-

рой корректировки кадастровая стоимость 2015 года у объекта оценки не изменилась, так как корректировка проводилась на ценообразующие факторы объекта оценки. Точечная оценка: $Mode(V) = 0,5317 \cdot 101,368^{1,1415} = 103,610$ тыс. руб./кв. м. Стандартное отклонение условных распределений не зависит от дисперсии регрессора (т. е. не зависит от кадастровой стоимости 2015 года) и равна (см. (10)) $\sigma = 0,053$. Интервальная оценка с надежностью 68,2 % (интервал $\pm \sigma$) [98,539; 109,558] (в тыс. руб./кв. м). Интервальная оценка с надежностью 95,4 % (интервал $\pm 2\sigma$) [93,452; 115,521] (в тыс. руб./кв. м).

Таким образом, полученные применением корректировок, определенных по взаимному расположению кластеров объектов оценки, точечная и интервальные оценки рыночной стоимости не значительно отличаются от прогнозных значений, полученных по кадастровым данным двух периодов.

Выводы

1. Кадастровые базы могут быть использованы для статистического обучения и построения прогноза по рыночной стоимости объекта недвижимости в краткосрочном (до следующей кадастровой оценки) периоде.

2. На основании данных кадастрового учета можно рассчитывать корректировки по различным ценовым факторам, основанные на обработке больших данных. Такие корректировки могут применяться к текущим рыночным данным.

3. Предложенный подход имеет более сложную технику расчетов, чем это принято в оценочной практике, но он разрешает типичные для сравнительного подхода проблемы, а именно:

¹¹ Следует обратить внимание, что условная дисперсия зависимой переменной не зависит от дисперсии регрессора и при значениях ρ , близких к единице, будет достаточно малой.

- объекты сравнения могут быть любыми, не обязательно «похожими» на объект сравнения;

- корректировки проводятся по кластерам целиком; чтобы их поменять, надо изменить кластеры; расхождения из-за слишком дешевых или слишком дорогих объектов сравнения устраняются;

- взвешивание результатов корректировок не требуется — происходит автоматически при объединении откорректированных кластеров разного размера.

4. Рынок оценочных услуг нуждается в создании обширных рыночных баз данных. Включение кадастрового номера в базы данных может придать новый импульс к разработке новых методов оценки.

Библиографический список

1. Ласкин М. Б., Гадасина Л. В. Как определить кадастровую стоимость // Имущественные отношения в Российской Федерации. 2018. № 3 (198).
2. Русаков О. В., Ласкин М. Б., Джаксумбаева О. И. Оценка показателей рынка недвижимости по статистическим данным на основе многомерного логарифмически нормального закона // Экономический журнал Высшей школы экономики. 2016. Т. 20, №. 2. С. 268–284.
3. Rusakov O. V., Laskin M. B., Jaksumbaeva O. I. Pricing in the real estate market as a stochastic limitLog Normal approximation//International Journal of mathematical models and methods in applied sciences. 2016. Vol. 10. Pp. 229–236.
4. Лемешко Б. Ю., Лемешко С. Б., Семенова, М. А. К вопросу статистического анализа больших данных // Вестник Томского университета. 2018. № 44.
5. Korkmaz S., Goksuluk D., Zararsiz G. MVN: An R package for assessing multivariate normality // The R Journal. 2014. Vol. 6, №. 2. Pp. 151–162.
6. Korkmaz S., Goksuluk D., Zararsiz G. MVN: An R package for assessing multivariate normality // Trakya University, Faculty of Medicine, Department of Biostatistics, Edirne, TURKEY 1 selcukorkmaz@gmail.com MVNversion 5.7 (Last revision 2019-03-18).
7. Русаков О. В., Ласкин М. Б., Джаксумбаева О. И., Стабровская К. Ю. Определение скидки на торг по статистическим данным // Вестник гражданских инженеров. 2016. № 2. С. 268–284.
8. Ласкин М. Б. Корректировка рыночной стоимости по ценообразующему фактору «площадь объекта» // Имущественные отношения в Российской Федерации. 2017. № 8 (191). С. 86–99.

References

1. Laskin M. B., Gadasina L. V. *Kak opredelit' kadas-trovuyu stoimost'* [How to determine the cadastral value]. *Imushchestvennyye otnosheniya v Rossiyskoy Federatsii – Property relations in the Russian Federation*, 2018, no. 3 (198).
2. Rusakov O. V., Laskin M. B., Dzhaksumbaeva O. I. *Otsenka pokazateley rynka nedvizhimosti po statisticheskim dannym na osnove mnogomernogo logarifmicheski normal'nogo zakona* [Evaluation of real estate market indicators based on statistical data based on a multidimensional logarithmically normal law]. *Ekonomicheskiy zhurnal Vysshey shkoly ekonomiki – Economic journal of the Higher school of Economics*, 2016, vol. 20, no. 2, pp. 268–284.
3. Rusakov O. V., Laskin M. B., Jaksumbaeva O. I. Pricing in the real estate market as a stochastic limitLog Normal approximation. *International Journal of mathematical models and methods in applied sciences*, 2016, vol. 10, pp. 229–236.
4. Lemesheko B. Yu., Lemesheko S. B., Semenova M. A. *K voprosu statisticheskogo analiza bol'shikh dannykh* [On the issue of statistical analysis of big data]. *Vestnik Tomskogo Universiteta – Bulletin of Tomsk University*, 2018, no. 44.
5. Korkmaz S., Goksuluk D., Zararsiz G. MVN: An R package for assessing multivariate normality. *The R Journal*, 2014, vol. 6, no. 2, pp. 151–162.
6. Korkmaz S., Goksuluk D., Zararsiz G. MVN: An R package for assessing multivariate normality. Trakya University, Faculty of Medicine, Department of Biostatistics, Edirne, TURKEY 1 selcukorkmaz@gmail.com MVNversion 5.7 (Last revision 2019-03-18).
7. Rusakov O. V., Laskin M. B., Dzhaksumbaeva O. I., Stabrovskaya K. Yu. *Opredelenie skidki na torг po statisticheskim dannym* [Determination of discounts for the trading according to the statistical data]. *Vestnik grazhdanskikh inzhenerov – Bulletin of Civil Engineers*, 2016, no. 2, pp. 268–284.
8. Laskin M. B. *Korrektirovka rynochnoy stoimosti po tsenobrazuyushchemu faktoru «ploshchad' ob'ekta»* [Market value adjustment according to the “object area” price-forming factor]. *Imushchestvennyye otnosheniya v Rossiyskoy Federatsii – Property relations in the Russian Federation*, 2017, no. 8 (191), pp. 86–99.