

УДК 332.628

КЛАССИФИКАЦИЯ НАСЕЛЁННЫХ ПУНКТОВ ПО УРОВНЮ РАЗВИТОСТИ РЫНКА ЗЕМЕЛЬНЫХ УЧАСТКОВ ИНДИВИДУАЛЬНОГО ЖИЛОГО СТРОИТЕЛЬСТВА

Быкова Елена Николаевна¹,
vilena@spmi.ru

Балтыжакова Татьяна Игоревна¹,
tatiyana.baltyzhakova@gmail.com

Волкова Яна Александровна¹,
yana.docenko@inbox.lv

¹ Санкт-Петербургский горный университет,
Россия, 199106, г. Санкт-Петербург, Васильевский остров, 21 линия, 2.

Актуальность исследования обосновывается высокой степенью влияния недостаточного количества рыночных исходных данных на результаты кадастровой оценки, что является значимой проблемой при работе с населёнными пунктами с небольшим количеством сделок. В статье представлено обоснование идеи о дифференцированном подходе к кадастровой оценке в условиях разной развитости рынка в населённых пунктах России.

Цель исследования: обоснование необходимости дифференцированного подхода к кадастровой оценке в условиях разной развитости сектора рынка «земли для индивидуального жилого строительства» в населённых пунктах России. Для достижения этой цели в рамках исследования предлагается классифицировать населённые пункты России по уровню развитости рынка земель.

Объектом исследования являются населённые пункты РФ, в которых в период с 1 января 2011 по 31 декабря 2015 г. зарегистрированы сделки с землями индивидуального жилого строительства.

Результаты. Разработана классификация населённых пунктов по уровню развитости рынка, которая является основой для создания методики кадастровой оценки земель в условиях малоразвитого рынка. Населённые пункты были разделены на три группы: населённые пункты с развитым рынком земельных участков для индивидуального жилого строительства – более 129 сделок купли-продажи в год и в среднем – более 700 за указанный период; населённые пункты с условно развитым рынком земельных участков для индивидуального жилого строительства – те населённые пункты, где количество сделок распределено неравномерно: от 10 до 150 сделок в год, в среднем за период около 300; Населённые пункты с малоразвитым рынком земельных участков для индивидуального жилого строительства – до 10 сделок в год, в среднем за весь оцениваемый период не более 50. Из чего можно сделать вывод, что идея о дифференцированном подходе к кадастровой оценке является обоснованной.

Ключевые слова:

Кадастровая оценка, налоги, индивидуальное жилое строительство, классификация населённых пунктов, метод к-средних.

Введение

В современной России в основе понятия «кадастровая стоимость» лежит концепция рыночной стоимости объекта недвижимости, определённой методами массовой оценки. Реализация такого подхода требует большого объёма информации о важнейших характеристиках объектов и данных о сделках купли-продажи на оцениваемой территории. В последние годы исследователи всё чаще отмечают, что недостаточная точность результатов кадастровой оценки вызвана не только низким качеством исходных данных, но и, в первую очередь, их недостаточным количеством.

В соответствии с градостроительными регламентами, жилая зона является базовой в составе земель населённых пунктов, в связи с чем основное бремя налогообложения ложится на плечи владельцев жилой недвижимости. Тогда как земельные участки, входящие в состав имущества многоквартирного дома, полностью освобождены от налогообложения, собственникам земель индивидуального жилого строительства (ИЖС) приходится платить налог в полном объёме [1].

Эта проблема приобретает чрезвычайную актуальность в связи с ростом заинтересованности

граждан переездом из центров больших городов на окраины, о чём свидетельствует устойчивое увеличение количества вводимых в эксплуатацию объектов ИЖС [2]. Данный факт обосновывает выбор в качестве объекта исследования именно земельные участки с видом разрешенного использования ИЖС.

Сопоставление удельных показателей кадастровой стоимости (УПКС) (стоимости 1 м²) с удельными показателями рыночной стоимости (УПРС) земельных участков индивидуального жилого строительства в населённых пунктах показывает, что кадастровая стоимость земельных участков, в большинстве случаев, выше рыночной в среднем на 30 % (рис. 1), что определяет превышение земельного налога в этом соотношении.

Тема результатов кадастровой оценки всё чаще поднимается средствами массовой информации, которые внушают властям намерение нечестно пополнить казну. Рядовые граждане-собственники объектов недвижимости выражают своё несогласие при помощи обращений в суды (рис. 2) и в территориальные комиссии Росреестра (рис. 3) с целью пересмотра результатов кадастровой оценки [3].



Рис. 1. Сопоставление УПКС и УПРС земельных участков индивидуального жилого строительства в населённых пунктах Ненецкого АО, Республики Коми и Калининградской области

Fig. 1. Correlation between specific indicators of cadastral cost (SICC) and specific indicators of market cost (SIMC) for land properties in Nenets Autonomous District, Komi Republic and Kaliningrad region

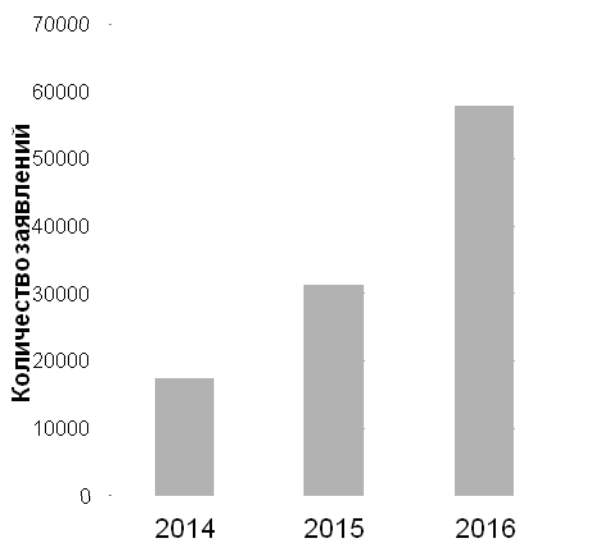


Рис. 2. Статистика обращения граждан в комиссии при территориальных органах Росреестра с целью пересмотра кадастровой стоимости недвижимого имущества за 2014–2016 гг.

Fig. 2. Statistics for citizens' petitions to commissions within Rosreestr local areas for reviewing real estate cadastral costs for 2014–2016

Совершенствованию методического обеспечения кадастровой оценки с целью уточнения её результатов посвящали свои работы А.А. Варламов, С.В. Грибовский, О.Ю. Лепихина, Л.А. Лейфер и многие другие отечественные учёные и практики. Анализ международного опыта методик кадастро-

вой оценки (или её аналогов) показал высокую актуальность направления и в других странах. Особенно стоит отметить работы Р. Борста (США), Д. Вельда (Нидерланды), Т. Кауко (Финляндия), Н. Синяка (Белоруссия).



Рис. 3. Статистика судебных споров в отношении результатов определения кадастровой стоимости объектов недвижимости за 2011–2016 гг.

Fig. 3. Statistics for legal cases on the results of real estate cadastral costs for 2011–2016

Исследователями всё чаще предлагаются технологически новые и нетривиальные методы определения стоимости методами массовой оценки, од-

нако по-прежнему остаются не разработанными направления оценки при дефиците исходных данных.

Уже на предварительных этапах исследования было выявлено, что проблема недостаточного количества исходных данных о ценах сделок купли-продажи земель ИЖС характерна для значительного количества населённых пунктов в России. В первую очередь, это населённые пункты, удалённые от центров субъектов РФ, и населённые пункты, не относящиеся ни к одной из агломераций, – там количество производимых сделок минимально. Кроме того, зачастую совершённая сделка в таких населённых пунктах не отражает реальную рыночную ситуацию и, как следствие, не должна использоваться для определения кадастровой стоимости.

Исходя из текущего положения, объективно можно сформулировать основную цель исследования – обоснование необходимости дифференцированного подхода к кадастровой оценке в условиях разной развитости сектора рынка «земли ИЖС» в населённых пунктах России. Для достижения этой цели в рамках исследования впервые предлагается классифицировать населённые пункты России по уровню развитости рынка земель.

Методы и материалы

Подготовка исходных данных

Как уже было сказано ранее, исходными данными для кадастровой оценки являются значения факторных признаков и сведения мониторинга сделок купли-продажи. Традиционные методики требуют большого объёма исходных данных для построения статистически значимой модели кадастровой стоимости, что вынуждает оценщиков

прибегать к использованию цен предложений, информации по другим видам разрешённого использования, а также рассчитывать единый для всего населённого пункта УПКС в случаях, когда это не обусловлено единообразием условий рынка [4–7]. Классификация по уровню развитости рынка недвижимости позволит определить положение населённого пункта в классификации и, в свою очередь, подобрать метод оценки, соответствующий конкретным рыночным условиям.

Для разработки классификации была собрана информация о сделках купли-продажи земель ИЖС по всем населённым пунктам Российской Федерации, сведения о которых имеются в автоматизированной информационной системе «Мониторинг рынка недвижимости».

В рамках исследования было изучено более 20000 населённых пунктов различного типа. Информация о зарегистрированных сделках купли-продажи земельных участков индивидуального жилого строительства была собрана для почти 7000 населённых пунктов за период с 1 января 2011 г. по 1 января 2016 г. Временной интервал обоснован периодом действия результатов предыдущей кадастровой оценки недвижимого имущества, которая производилась в 2011–2012 гг. (в зависимости от региона). На момент исследования не были опубликованы в полном объёме данные о сделках за 2016 г., поэтому информация за этот период не использовалась.

Поскольку развитость рынка в пределах населённых пунктов РФ может зависеть не только от количества сделок за определённый период, в качестве дополнительного параметра классификации было решено использовать плотность населения. В связи с этим была выдвинута гипотеза о

Таблица 1. Исходные данные для классификации населённых пунктов России по уровню развитости рынка земельных участков индивидуального жилого строительства (фрагмент)

Table 1. Basic data for classifying Russian inhabited localities by individual residential construction (IRC) land property market development (fragment)

Субъект Российской Федерации Subject of the Russian Federation	Населённый пункт Inhabited locality	Количество зарегистрированных сделок (за год) Number of registered deals (per year)					Плотность населения чел./км ² Population density (person per 1 km ²)	
		2011	2012	2013	2014	2015		Всего за пять лет Total for five years
Ивановская область/Ivanovo region	Лух/Lukh	0	2	0	0	0	2	952,34899
Ингушетия/Ingushetia	Магас/Magas	0	0	0	1	0	1	544,73476
Волгоградская область/Volgograd region	Манойлин/Manoylin	0	5	0	0	0	5	2,8557867
Владимирская область/Vladimir region	Меленки/Melenki	80	47	18	12	28	185	1290,0000
Кемеровская область/Kemerovo region	Мыски/Myski	98	70	45	126	39	570	385,81417
Хабаровский край Khabarovsk region	Некрасовка Nekrasovka	0	5	1	1	0	7	6738,48440
Ханты-Мансийский АО Khanty-Mansiysk Autonomous Region	Нижневартовск Nizhnevartovsk	5	8,0	2	8	0	23	998,25667
Свердловская область Sverdlovsk region	Нижний Тагил Nyzhniy Tagil	0	0	1	0	0	2	1199,62290
Томская область Tomsk region	Томск/Tomsk	515	574	170	339	67	1665	1944,13
Томская область Tomsk region	Стрежевой Strezhevoy	3	4	1	4	0	10	754,7984

том, что количество сделок связано с плотностью населения. Такое предположение обосновано тем, что:

Во-первых, конкуренция за земельные участки в населенных пунктах зависит от численности населения, которая тем выше, чем привлекательнее населенный пункт.

Во-вторых, увеличение численности населения и инвестиционной привлекательности населенного пункта влечет его развитие в пространственном отношении, то есть способствует увеличению площади в границах.

Исходя из вышесказанного в качестве исходных данных в работе, кроме данных мониторинга сделок купли-продажи, использовалась информация о площади и количестве жителей в населённых пунктах. На основании этой информации рассчитана плотность населения для 1007 населённых пунктов из исходной выборки (табл. 1).

Для проверки гипотезы связи между развитостью рынка земельных участков и плотностью населения была построена матрица корреляции (табл. 2).

Для проверки связи между количеством зарегистрированных сделок и плотностью населения был рассчитан линейный коэффициент корреляции Пирсона [4], описываемый формулой (1).

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}}, \quad (1)$$

где x_i и y_i – объекты выборок $x^m=(x_1, \dots, x_m)$, $y^m=(y_1, \dots, y_m)$; \bar{x} и \bar{y} – выборочные средние [8].

Выявлено, что связь между плотностью населения и количеством сделок купли-продажи земельных участков отсутствует, так как коэффициенты корреляции между этими признаками менее 0,1 (табл. 2).

В связи с этим использование плотности населения как параметра классификации является нецелесообразным, поэтому в качестве исходных данных выступает исключительно количество за-

регистрированных сделок купли-продажи (по годам). По итогам сбора и подготовки информации была сформирована выборка из 6926 объектов.

Предварительный анализ данных

Все расчеты и визуализация результатов выполнялись в системе статистических вычислений R. Это открытая программная среда, которая поддерживает широкий спектр методов обработки и визуализации данных [9].

Для аналитического осмысления исходной информации и оценки ее распределения на этапе предварительного анализа исходных данных были построены гистограммы распределения количества зарегистрированных сделок купли-продажи по годам (рис. 4–8). При этом для построения гистограммы определена ее ширина на основе размаха вариации ($R=664$) и ширина интервала ($h=2$). На оси X отложены интервалы числа сделок, а по оси Y – частота, т. е. количество населенных пунктов, попавших в данный интервал.

Прежде всего следует отметить, что согласно эмпирическим распределениям на построенных графиках, число зарегистрированных сделок купли-продажи по годам стабильно, т. е. незначительно изменяется. Кроме того, в большинстве рассматриваемых населенных пунктов количество сделок в течение года невелико. В значительном числе населённых пунктов совершается до 10 сделок в год. Меньше всего городов, в которых происходит более 100 сделок в год. Таким образом, можно предположить, что населенные пункты по уровню развитости рынка делятся на две группы: от 0 до 10 сделок в год и от 11 до 664.

Переход к обобщенным показателям развитости рынка

Рассматриваемые данные о количестве сделок обладают высокой корреляцией (табл. 2), что может привести к некорректным результатам при их анализе. Кроме того, такая тесная связь между признаками затрудняет наглядное отображение результатов. Поэтому было принято решение о переходе к обобщенным показателям развитости рынка с помощью метода анализа главных компонент.

Таблица 2. Матрица корреляции между уровнем развитости рынка и плотностью населения

Table 2. Correlation matrix between market development and population density

Количество зарегистрированных сделок (за год) Number of registered deals (per year)	Количество зарегистрированных сделок (за год) Number of registered deals (per year)					Плотность населения Population density	
	2011	2012	2013	2014	2015		Всего за 5 лет Total for 5 years
2011	1,0000	0,9484	0,8462	0,8151	0,7864	0,9456	-0,0068
2012		1,0000	0,8546	0,8128	0,7859	0,9407	-0,0069
2013			1,0000	0,9361	0,8943	0,9295	0,0016
2014				1,0000	0,8985	0,9110	0,0091
2015					1,0000	0,8882	-0,0014
Всего за 5 лет/Total for 5 years						1,0000	-0,0024
Плотность населения Population density							1,0000

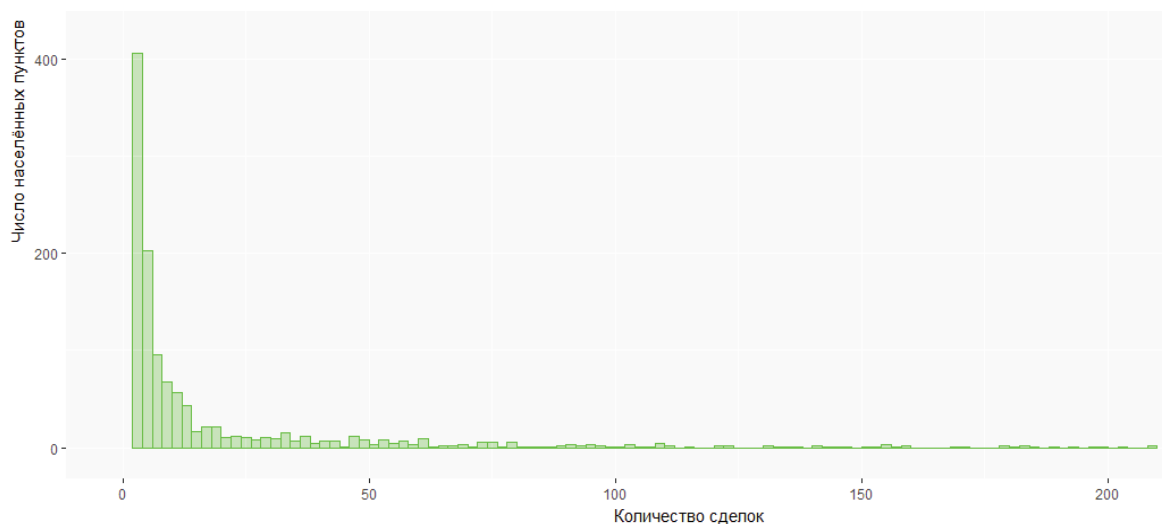


Рис. 4. Гистограмма распределения количества сделок за 2011 г.

Fig. 4. Histogram of allocation of a number of deals for 2011

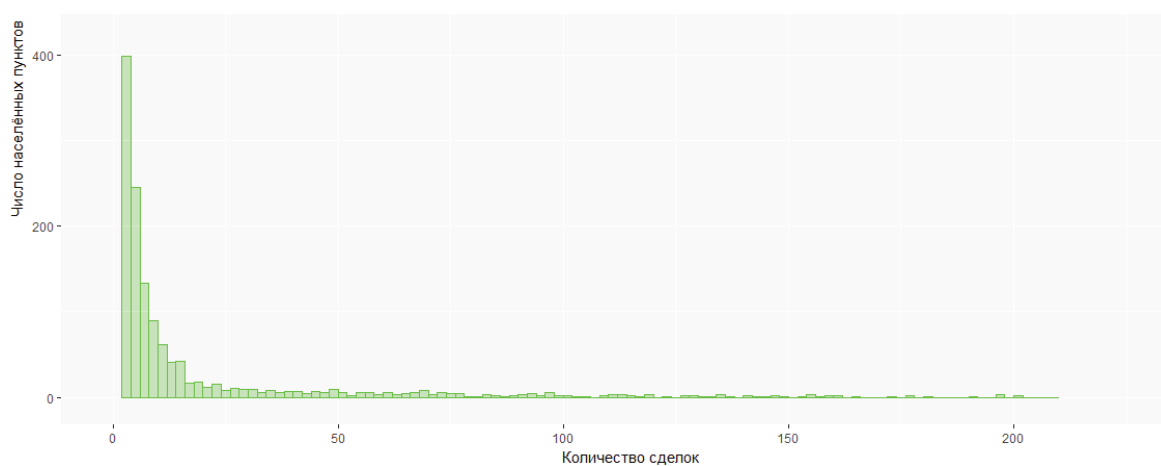


Рис. 5. Гистограммы распределения количества сделок за 2012 г.

Fig. 5. Histogram of allocation of a number of deals for 2012

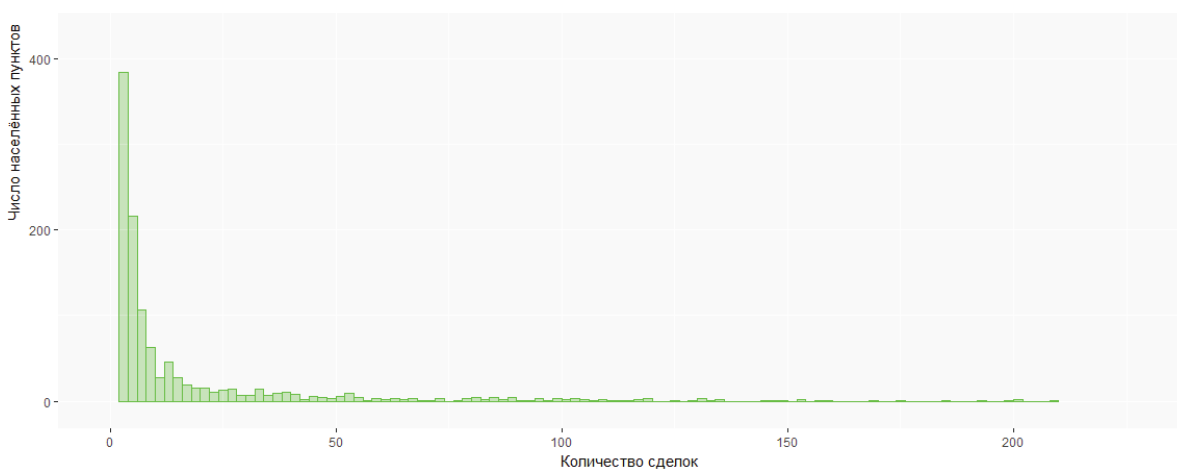


Рис. 6. Гистограмма распределения количества сделок за 2013 г.

Fig. 6. Histogram of allocation of a number of deals for 2013

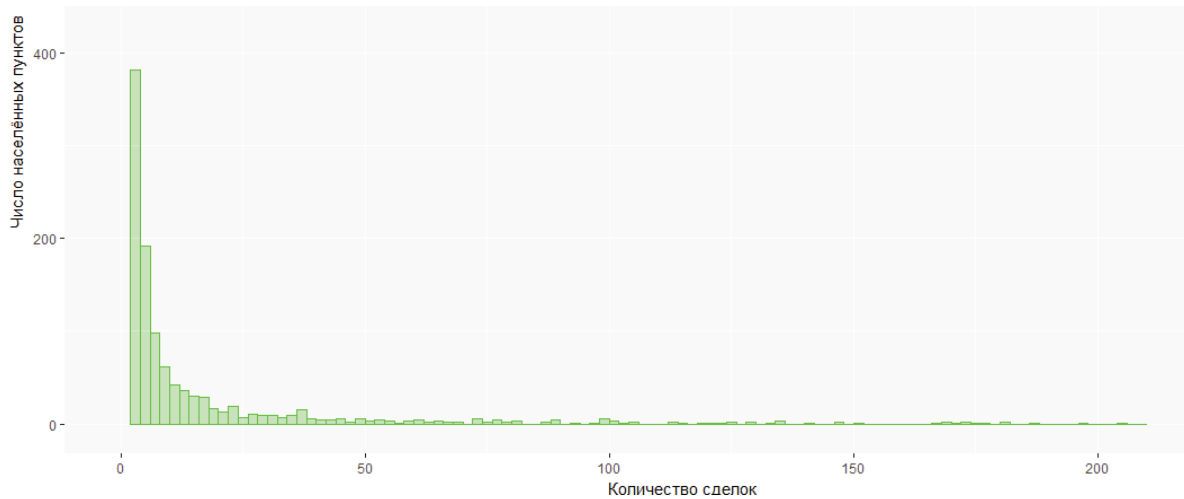


Рис. 7. Гистограмма распределения количества сделок за 2014 г.

Fig. 7. Histogram of allocation of a number of deals for 2014

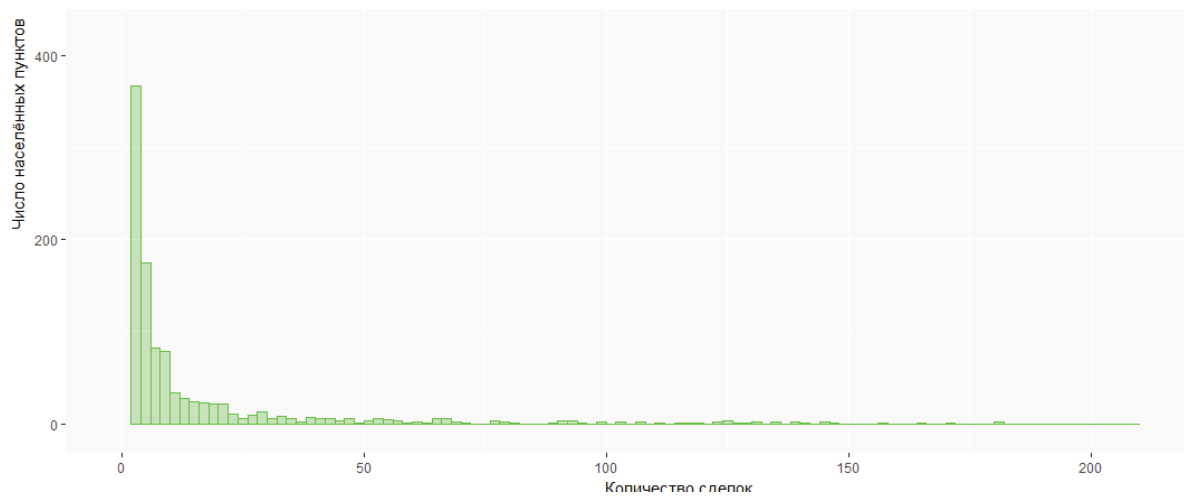


Рис. 8. Гистограмма распределения количества сделок за 2015 г.

Fig. 8. Histogram of allocation of a number of deals for 2015

Этот метод является одной из разновидностей методов классификации и факторного анализа и, как правило, применяется тогда, когда есть необходимость выявить независимые друг от друга факторы (компоненты), максимально точно описывающие данные без потери информации, и сократить размерность данных.

Метод анализа главных компонент построен на основе линейных преобразований систем координат. Исходными показателями в данном случае являются значения количества сделок по годам. На основании полученного облака данных выбирается направление, соответствующее максимальному разбросу данных, – эта ось становится первой главной компонентой, вторая главная компонента задается как перпендикулярная первой. Таким образом, происходит разворот осей и переход к новой системе независимых показателей, которые максимально точно описывают данные.

Формально метод описывается с помощью формулы (2) [10].

$$X = TP^t + E = \sum_{a=1}^A t_a P_a^t + E, \quad (2)$$

где X – матрица исходных данных; T – матрица счетов (scores), ее размерность $(I \times A)$; P – матрица нагрузок (loadings), ее размерность $(J \times A)$; E – матрица остатков, размерностью $(I \times J)$.

Новые переменные t_a называются главными компонентами (Principal Components).

Матрица счетов T дает проекции исходных образцов (J -мерных векторов x_1, \dots, x_j) на подпространство главных компонент (A -мерное). Строки t_1, \dots, t_j матрицы T – это координаты образцов в новой системе координат. Столбцы t_1, \dots, t_A матрицы T ортогональны и представляют проекции всех образцов на одну новую координатную ось.

Матрица нагрузок P – это матрица перехода из исходного пространства переменных x_1, \dots, x_j (J -мерного) в пространство главных компонент (A -мерное). Каждая строка матрицы P состоит из коэффициентов, связывающих переменные t и x .

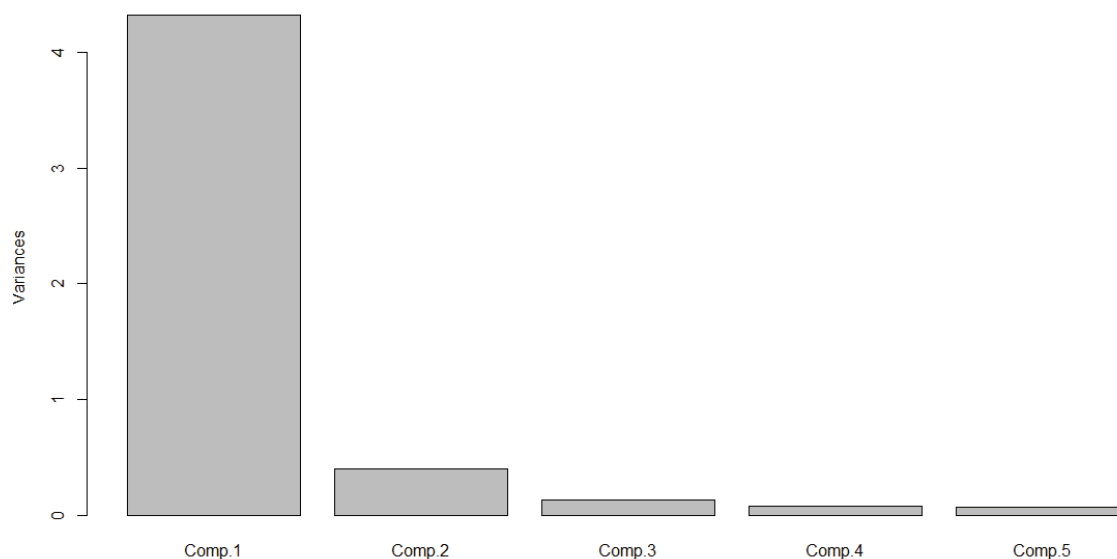


Рис. 9. Служебный график анализа главных компонент

Fig. 9. Support diagram showing the principal component analysis

Например, a -я строка – это проекция всех переменных x_1, \dots, x_j на a -ю ось главных компонент. Каждый столбец P – это проекция соответствующей переменной x_j на новую систему координат [9, 11].

Особенностью реализации этого метода в R является то, что количество полученных компонент равно количеству исходных признаков. Необходимое же количество компонент определяется на основе уровня объясненной вариации данных, то есть на основе того, какой процент разброса данных объясняет каждая компонента. Самой значимой всегда является компонента под номером один, так как она вводится по направлению максимального разброса данных.

Рассмотрим служебный график, показывающий вклады каждой компоненты (рис. 9).

На графике видно, что компонент, как и факторов, пять, однако основной вклад вносят первые две. Величины дисперсий и объясненной вариации каждой компоненты приведены в табл. 3.

Таблица 3. Характеристики главных компонент

Table 3. Characteristics of principal components

Характеристики Values	Номера компонент Numbers of the components				
	C_I	C_{II}	C_{III}	C_{IV}	C_V
Дисперсия компоненты Component dispersion	2,0400	0,6947	0,3861	0,3495	0,2908
Доля в вариации Variative fraction	0,8323	0,0965	0,0298	0,0244	0,01690
Кумулятивная доля Cumulative fraction	0,8323	0,9288	0,9587	0,9831	1,0000

Из рис. 9 и табл. 3 видно, что максимальное влияние на результат оказывают первые две компоненты, поэтому при необходимости можно ограничиться только ими. Однако, так как в данном

исследовании нет необходимости в сокращении размерности, для дальнейших расчетов были использованы все пять компонент.

Для перехода от исходных показателей к компонентам применяется матрица нагрузок. Полученная матрица нагрузок приведена в табл. 4.

Таблица 4. Матрица нагрузок

Table 4. Loading matrix

	Компонента/Component				
	1	2	3	4	5
2011	-0,443	0,537	-0,171	0,101	-0,691
2012	-0,451	0,454	-0,209	0,172	0,719
2013	-0,464		0,503	-0,727	
2014	-0,428	-0,593	-0,661	-0,165	
2015	-0,449	-0,392	0,488	0,636	

Из матрицы нагрузок видно, что почти все исходные показатели влияют на первые две компоненты, причем величина этого влияния невелика. Это обусловлено, прежде всего, высоким уровнем корреляции между исходными показателями.

Выбор метода классификации

Одной из основных задач исследования является определение количества групп населенных пунктов по уровню развитости рынка и интервалов значений каждой группы. Такие задачи решаются с применением методов классификации.

Метод классификации выбирается с учетом имеющихся исходных данных, а также решаемой в исследовании задачи. В том случае, если есть объекты с известной принадлежностью к группам, либо есть информация о распределении наблюдений внутри классов, применяются методы классификации с обучением. В прочих случаях используются методы классификации без обучения.

Основными методами классификации без обучения являются методы кластеризации и метод главных компонент. Для решения поставленной задачи оптимальными являются методы кластеризации, так как неизвестна принадлежность объектов к конкретным классам (группам) и необходимо определить параметры классов.

Методы кластеризации делятся на две основные группы: иерархические и последовательные методы. Первая группа методов применяется в том случае, если предполагается иерархия групп и при анализе небольшой выборки (до 150 объектов). Во всех остальных случаях более эффективными являются последовательные методы кластеризации.

Последовательные алгоритмы кластеризации являются итеративными и применяются для больших массивов данных, так как они позволяют производить расчеты для небольших групп объектов на каждой итерации.

Среди методов последовательной кластеризации стоит выделить метод k -средних, который можно отнести к наиболее традиционным и развиваемым алгоритмам. Его достоинствами являются: прозрачность алгоритма, высокая линейная скорость работы и эффективная обработка больших объёмов данных [12].

Алгоритм метода k -средних состоит из следующих шагов:

- 1) случайным образом выбирается k центров кластеров;
- 2) каждый объект приписывается к ближайшему центру;
- 3) производится пересчет центров кластеров как среднее значение параметров всех объектов, попавших в определенный кластер;
- 4) каждый объект приписывается к ближайшему центру кластера.

Шаги 3 и 4 повторяются до тех пор, пока центры кластеров не перестанут изменяться при пересчете или пока не будет достигнуто максимально возможное количество итераций. В результате должно быть получено k кластеров с минимальным значением целевой функции.

В качестве целевой функции, как правило, используется сумма квадратов расстояний внутри кластера, которая описывается формулой (3).

$$s(k) = \sum_{i=1}^n \sum_{j=0}^p (x_{ij} - \bar{x}_{kj})^2, \quad (3)$$

где k – кластер; x_{ij} – значение j -й переменной для i -го наблюдения (объекта); \bar{x}_{kj} – среднее значение j -й переменной в k -м кластере; p – количество переменных [9].

Исходными данными для проведения кластеризации с помощью метода k -средних разбиения являются: анализируемый массив данных, используемая метрика, количество кластеров и метод разбиения.

Определение количества кластеров производилось с использованием специализированного пакета NbClust [13]. Он разработан на основе исследо-

вания Г.В. Миллигана и М.С. Купер (1985), посвящённого анализу правил останковки – подходов, позволяющих определить оптимальное количество кластеров. Всего в пакете используется 30 правил останковки, таких как индекс СН (Calinski and Harabasz), индекс Дуда, Псевдо t^2 (Pseudo t^2), С-индекс, Гамма-индекс, индекс Биля (Beale), кубический критерий кластеризации (Cubic clustering criterion, CCC), коэффициент точечной двусторонней корреляции (Point-biserial correlation coefficient), Gplus-индекс и другие. В качестве критериев определения количества кластеров используются: матрица дисперсий внутри кластеров, матрица дисперсий между кластерами, общее количество данных в выборке, общее количество данных, входящих в кластер, сумма расстояний между кластерами, сумма расстояний в пределах кластеров.

Для примера, индекс Дуда описывается формулой (4):

$$Duda = \frac{Je(2)}{Je(1)}, \quad (4)$$

где $Je(2)$ – сумма квадратов внутрикластерных расстояний в случае, когда данные распределены по двум кластерам; $Je(1)$ – сумма квадратов внутрикластерных расстояний в случае одного кластера.

Гипотеза о существовании единого кластера элементов подтверждается, если значение критерия Дуда больше критического значения, определяемого по формуле (5).

$$\text{Критическое значение} = 1 - \frac{2}{p\rho} - z\sqrt{\frac{2(1 - \frac{8}{\pi^2\rho})}{n_m\rho}}, \quad (5)$$

где ρ – число признаков; n_m – квантиль стандартного нормального распределения уровня [14].

Критерий Дуда называют локальным правилом, так как он позволяет оценить корректность при $k=1$, то есть при отсутствии разбиения выборки на группы. При работе с данными, кластеризованными на две и более группы, необходимо использовать глобальные правила останковки. Например, индекс СН, который указывает на наиболее вероятное количество кластеров, задается с помощью формулы (6):

$$CH_{(q)} = \frac{\text{trace}(B_q) / (q-1)}{\text{trace}(W_q) / (n-q)}, \quad (6)$$

где B_q – матрица дисперсий между кластерами; W_q – матрица дисперсий внутри кластеров; n – количество наблюдений; q – количество кластеров [14]. В основе определения итогового количества кластеров на базе всех правил останковки лежит принцип большинства. Принцип большинства представляет собой правило, в соответствии с которым количество кластеров считается оптимальным, если оно получено по большинству критериев (табл. 5).

Графическое представление результатов приведено на рис. 10. На графике по оси X показано количество кластеров, а по оси Y отложено количе-

Таблица 5. Количество кластеров, определённое с использованием интегрированных в NbClust правил остановки

Table 5. Number of clusters defined by stopping rules integrated into NbClust

	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW
Number_clusters	2,000	2,000	10,000	2,000	5,000	5,000000e+00	3,0
Value_Index	6,2492	5901,491	4825,086	-34,968	3033,448	1,364096e+16	669453,2
	TraceW	Friedman	Rubin	Cindex	DB	Silhouette	Duda
Number_clusters	5,000	5,000	5,000	2,000	2,000	2,000	2,000
Value_Index	3410,754	5,6843	-0,6781	0,028	0,4902	0,9581	0,6122
	PseodoT2	Beale	Ratkowsky	Ball	PtBiserial	Frey	McClain
Number_clusters	2,000	4,000	2,000	3,000	7,000	NA	2e+00
Value_Index	19,6366	1,1478	0,493	2987,084	0,8102	NA	0,1677
	Dunn	Hubert	SDindex	Dindex	SSbw		
Number_clusters	2,000	0	9,000	0	9,000		
Value_Index	0,1677	0	1,5993	0	3,6456		

Примечание: Number_clusters – количество кластеров согласно одному из критериев; Value_Index – вес правила остановки.

Note: Number_clusters is the number of clusters according to one of the criteria; Value_Index is the value of the stopping rule.

ство критериев, согласно которым получено соответствующее число кластеров.

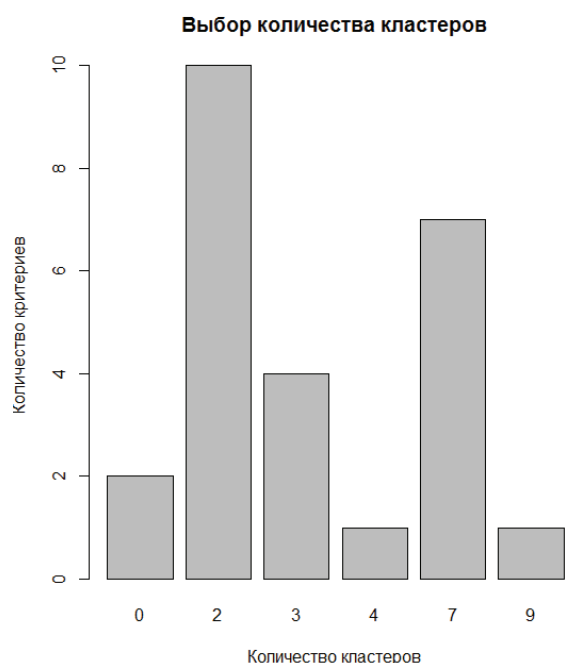


Рис. 10. Определение количества кластеров

Fig. 10. Definition of the number of clusters

Достаточная согласованность между результатами применения правил остановки позволяет сказать, что у информации, используемой для кластеризации, есть определённая структура. В результате проведенных расчетов, согласно табл. 4 и рис. 10, оптимальное количество кластеров для анализируемой совокупности составляет два.

Кроме определения количества групп, важной задачей при кластеризации является выбор метрики. По своей сути метрика – это мера сходства одного элемента выборки с другими. Для расчёта метрики используются: расстояние Чебышева, процент несогласия, а также степенное, манхэттенское и Евклидово расстояния [15]. Расстояние Чебышева применяется в том случае, если известно,

что объекты отличаются только по одной координате (по одному признаку). Степенное расстояние используется тогда, когда необходимо увеличить или уменьшить вес для той размерности, по которой объекты значительно отличаются. Для нечисловых данных применяется процент несогласия.

Для непрерывных данных наиболее распространёнными метриками при проведении кластеризации являются евклидово и манхэттенское расстояния.

Евклидово расстояние используется в том случае, если компоненты вектора наблюдений однородны по своему физическому смыслу и понятие близости объектов совпадает с понятием геометрической близости объектов в n -мерном пространстве [15] и описывается формулой (7).

$$d = \sqrt{((A_1 - B_1)^2 + (A_2 - B_2)^2 + \dots + (A_n - B_n)^2)}, \quad (7)$$

где A_1, \dots, A_n – положение объекта A в n -мерном пространстве; B_1, \dots, B_n – положение объекта B в n -мерном пространстве; d – расстояние между объектом A и B в n -мерном пространстве [16].

Манхэттенское расстояние (расстояние городских кварталов) по геометрическому смыслу является средним разностей по координатам. Как правило, результаты, полученные с использованием манхэттенского и евклидова расстояния, отличаются незначительно, однако следует отметить, что, если точки отличаются только по одной из координат, данная метрика сгладит эту разницу. Также следует учитывать то, что эта метрика наиболее применима в двумерном пространстве, то есть только при расчетах по двум признакам [17, 18].

Так как исследуемые данные являются числовыми, классификация производится по пяти параметрам и объекты отличаются друг от друга не по одному признаку, в качестве метрики было использовано евклидово расстояние.

Результаты кластеризации

В результате классификации по уровню развитости рынка земельных участков для ИЖС насе-

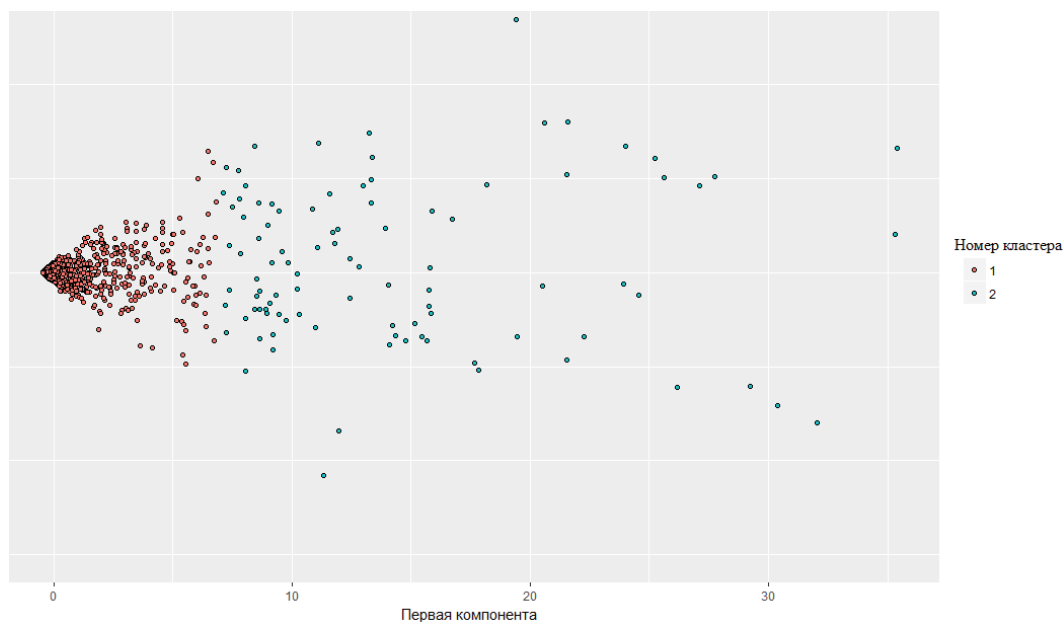


Рис. 11. Результат классификации населенных пунктов по уровню развитости рынка ИЖС

Fig. 11. Results of classification of inhabited localities by IRD property market

ленные пункты России были разделены на две группы. Результаты классификации в системе координат двух главных компонент представлены на рис. 11. На представленном графике каждый объект выборки – населенный пункт – представлен точкой [19].

Из рисунка видно, что полученные группы достаточно обособлены и практически не пересекаются, что является показателем хорошего качества проведенной кластеризации.

В первую группу попало 98 населённых пунктов. Среднее количество сделок в год составляет 129, за весь исследуемый период – более 700. Как правило, сюда входят населённые пункты, являющиеся столицами крупных и интенсивно развивающихся субъектов РФ – Красноярск, Уфа, Ульяновск, или городами федерального значения: Москва, Санкт-Петербург.

Вторая группа включает в себя 6828 населённых пунктов. Следует отметить, что населенные пункты в этой группе значительно отличаются друг от друга, поэтому было принято решение провести процедуру кластеризации еще раз только для населенных пунктов этой группы. В результате было выделено три подгруппы:

1. Населённые пункты с низким количеством зарегистрированных сделок купли-продажи земель ИЖС

К этой подгруппе относятся населённые пункты, о сделках в которых информация весьма ограничена. В среднем производится до 10 сделок в год на территории всего населённого пункта. В основном в эту группу входят сёла, посёлки, заимки и другие типы сельских населённых пунктов и малых городских населенных пунктов. Однако присутствуют и крупные города – например, Ростов-на-Дону или

Калуга. Такая ситуация может быть вызвана тем, что системе территориального планирования этих городов не характерно ИЖС, или рынок земель ИЖС на данный момент находится на стадии стагнации. Всего таких населённых пунктов 6517.

2. Населённые пункты со значительным количеством сделок в один (два) из годов периода изучения

Так называемые «скачки» количества сделок купли-продажи можно объяснить следующими обстоятельствами:

- изменение вида разрешённого использования земель в связи с перспективным развитием населённых пунктов для малоэтажной застройки (в т. ч. ИЖС). Например, в деревне Анкудиновка Нижегородской области активно осваиваются земли, предназначенные под индивидуальную жилую застройку. Общая площадь земель ИЖС составляет 22,36 га, деревня оборудована всеми необходимыми инженерными и транспортными коммуникациями [20];
- повышение инвестиционной привлекательности земель вследствие строительства новых объектов инфраструктуры и промышленности. За изучаемый период для города Бикине в Хабаровском крае было принято решение о реализации экономической программы создания локальной зоны предпринимательства;
- перспективное развитие населённых пунктов. Например, большое количество сделок в посёлке городского типа Анопино Владимирской области в 2011–2012 гг. связано с утверждением в 2010 г. нового генерального плана, который предполагает расширение площади населённого пункта и предусматривает развитие инфраструктуры.

Как правило, в таких населённых пунктах среднее количество сделок, как и для группы первого типа, находится в пределах 10 сделок в год, а в один–два года происходит резкий скачок, что является выбросом и не может учитываться при классификации.

Всего 99 населённых пунктов. Данные мониторинга сделок купли-продажи в таких населённых пунктах должны тщательно проверяться на возможность применения при построении модели кадастровой стоимости.

3. Населённые пункты, которые необходимо рассматривать отдельно

Населённым пунктам данной группы характерна достаточно высокая плотность сделок и стабильная динамика рынка. В среднем в таких населённых пунктах зарегистрировано около 30 сделок в год. При условии территориальной компактности размещения ИЖС в населённом пункте и высокой достоверности данных о сделках купли-продажи такие населённые пункты можно отнести к условно развитым.

Всего таких населённых пунктов 212.

Таким образом, классификация населённых пунктов России в зависимости от активности рынка земельных участков ИЖС представлена следующими группами:

1. Населённые пункты с развитым рынком земельных участков ИЖС – более 129 сделок купли-продажи в год, и, в среднем, более 700 за указанный период.
2. Населённые пункты с условно развитым рынком земельных участков ИЖС – те населённые пункты, где количество сделок распределено неравномерно: от 10 до 150 сделок в год, в среднем за период около 300.
3. Населённые пункты с малоразвитым рынком земельных участков ИЖС – до 10 сделок в год, в среднем за весь оцениваемый период не более 50.

Разработанная классификация предусматривает динамическую актуализацию данных, поэтому обоснованные интервалы и положение населённого пункта в классификации могут изменяться в зависимости от рыночных условий.

Заключение

Анализируя результаты классификации, можно заключить, что удельный вес населённых пунктов с развитым рынком земель индивидуального жилого строительства невелик (рис. 12). В то время как рыночные условия между группами различаются значительно, действующая в настоящее время методика кадастровой оценки земель насе-

лённых пунктов ориентирована исключительно на населённые пункты с большим количеством сделок купли-продажи.

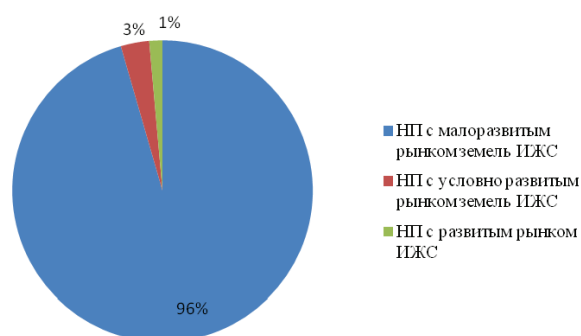


Рис. 12. Процентное соотношение групп населённых пунктов в зависимости от развитости рынка земель ИЖС

Fig. 12. Percentage of inhabited locality group by IRD land property market development

Проведенное исследование является основой для разработки предложений в области развития института кадастровой оценки недвижимости в России в условиях разного уровня развитости рынка:

1. В условиях разного уровня развитости рынка недвижимости населённых пунктов идея о дифференцированном подходе к кадастровой оценке является обоснованной, что подтверждается результатами выполненной работы.
2. С учётом того, что более 96 % населённых пунктов России имеют малоразвитый рынок земель индивидуального жилого строительства, а существующая методика предусматривает использование достаточной объективной исходной информации, следует осуществить разработку методики определения кадастровой стоимости, учитывающую недостаточность рыночной информации.
3. Представленная методика классификации населённых пунктов России должна явиться одним из этапов оценочных работ по кадастровой оценке, на основании которого должен определяться метод расчета кадастровой стоимости в различных условиях развитости рынка недвижимости.
4. Предложенная последовательность классификации населённых пунктов по уровню развитости рынка земель ИЖС может быть реализована и для других видов разрешенного использования земель, а также иного недвижимого имущества. Это обеспечит обоснование дифференцированного и объективного подхода к методическому аппарату расчета налогооблагаемой базы недвижимого имущества.

СПИСОК ЛИТЕРАТУРЫ

1. Налоговый кодекс Российской Федерации от 31 июля 1998 года № 146 (ред. 27.11.2017). – М.: Эксмо, 2017. – 1120 с.
2. Бюллетень о текущих тенденциях российской экономики // Аналитический центр при правительстве Российской Федерации URL: <http://ac.gov.ru/files/publication/a/10017.pdf> (дата обращения: 11.11.2017).
3. Федеральная служба государственной регистрации кадастра и картографии. URL: <https://rosreestr.ru/site/> (дата обращения: 11.11.2017).
4. Киселёв В.А., Лепихина О.Ю. Анализ достаточности и достоверности рыночной информации в малых и средних городах Северо-западного округа для оценки возможности применения статистического метода определения кадастровой стоимости земель // Записки Горного института. – 2011. – № 7. – С. 217–221.
5. Гладких Н.И., Кузнецова В.В. Определение необходимого количества аналогов при заданном числе ценообразующих факторов для целей оценки недвижимости методами корреляционно-регрессионного анализа // Имущественные отношения в РФ. – 2016. – № 6 (177). – С. 75–84.
6. Григорьев В.В. Совершенствование механизма управления государственной кадастровой оценкой в Российской Федерации // Государственное и муниципальное управление. – 2016. – № 1. – С. 83–89.
7. Бойко А.Ю. Кадастровая оценка. Проблемы и перспективы // Земля и бизнес. – 2016. – № 258. – С. 11–21.
8. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. – М.: Юнити, 2001. – 657 с.
9. Kabacoff R. R in action. – New York: Manning Publications Co., 2015. – 563 p.
10. Everitt B.S., Hothorn T. A Handbook of Statistical Analyses Using R // The Comprehensive R Archive Network. URL: https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_principal_components_analysis.pdf (дата обращения: 10.11.2017).
11. Venables W.N., Smith D.M., the R Core Team. An Introduction to R. Notes on R: a Programming Environment for Data Analysis and Graphics // The Comprehensive R Archive Network. URL: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf> (дата обращения: 18.11.2017).
12. Peng R. Exploratory Data Analysis with R // LeanPub: Publish Early, Publish Often. URL: <https://leanpub.com/exdata> (дата обращения: 11.11.2017).
13. NbClust: an R Package for Determining the Relevant Number of Clusters in a Data Set / M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs // Journal of Statistical Software. – 2014. – V. 61. URL: <https://www.jstatsoft.org/index.php/jss/article/view/v061i06/v61i06.pdf> (дата обращения: 11.11.2017).
14. Яцкив И., Гусарова Л. Методы определения количества кластеров при классификации без обучения // Transport and Telecommunication. – 2003. – № 1. – С. 23–28.
15. Leek J. The Elements of Data Analytic Style // LeanPub: Publish Early, Publish Often. URL: <https://leanpub.com/datastyle> (дата обращения: 11.11.2017).
16. Гитис Л.Х. Статистическая классификация и кластерный анализ. – М.: Изд-во московского горного университета, 2003. – 157 с.
17. Wickham H., Grolemund G. R for Data Science. – Sebastopol: O'Reilly Media, 2016. – 483 p.
18. Teetor P. R Cookbook. – Sebastopol: O'Reilly Media, 2011. – 397 p.
19. Burchell J., Vargas M. The Hitchhiker's Guide to Ggplot2 in R // LeanPub: Publish Early, Publish Often. URL: https://leanpub.com/hitchhikers_ggplot2 (дата обращения: 11.11.2017).
20. Генеральный план сельского поселения «Афонинский сельсовет» Кстовского района Нижегородской области // Кстовский район Нижегородской области. 2017. URL: http://www.kstovoadm.ru/about/rural_councils/afoninsky_selsove/town_planning.php (дата обращения: 20.11.2017).

Поступила 21.12.2017 г.

Информация об авторах

Быкова Е.Н., кандидат технических наук, доцент кафедры инженерной геодезии Санкт-Петербургского горного университета.

Балтыжакова Т.И., кандидат технических наук, ассистент кафедры инженерной геодезии Санкт-Петербургского горного университета.

Волкова Я.А., аспирант, ассистент кафедры инженерной геодезии Санкт-Петербургского горного университета.

UDC 332.628

CLASSIFICATION OF INHABITED LOCALITIES BY THE LEVEL OF DEVELOPMENT OF INDIVIDUAL RESIDENTIAL LAND MARKET

Elena N. Bykova¹,

vilena@spmi.ru

Tatiana I. Baltyzhakova¹,

tatiyana.baltyzhakova@gmail.com

Yana A. Volkova¹,

yana.docenko@inbox.lv

¹ St. Petersburg Mining University,
2, 21st Line, St.-Petersburg, 199106, Russia.

The relevance of the research is caused by high level of effect of insufficient amount of market data on the result of cadastral appraisal. This is the great problem for inhabited localities with low level of development of land registry. The paper introduces the motivation of the idea of case-specific approach to cadastral appraisal within the different market development in Russian inhabited localities.

The aim of the research is substantiation of the need in differentiated approach to cadastral appraisal within the different development of Land for Individual Residential Development land property market in Russian inhabited localities. The authors propose the classification of Russian inhabited localities by land property market development level.

Objects of research are all Russian inhabited localities where the deals with lands for individual residential construction development were registered from January 1st, 2011 to December 31st, 2015.

Methods. The inhabited localities were classified by real estate market development level using *k*-means, principal component analysis, histograms, stopping rules – approaches defining the best suitable number of the clusters.

Results. The authors have developed the classification of inhabited localities by real estate market development level. It is represented by the following groups: inhabited localities with developed individual residential construction land property market (more than 129 purchase/sale deals per year and on the average more than 700 deals for the studied period); inhabited localities with relatively developed individual residential construction land property market where the number of deals is spread irregularly, from 10 to 150 deals per year with the average of 300 deals for the studied period; inhabited localities with weakly developed individual residential construction land property market (up to 10 deals per year, up to 50 deals for the studied period). It allows making a conclusion, that the idea of different approach to cadastral appraisal is reasonable.

Key words:

Cadastral appraisal, taxes, individual residential construction, classification of inhabited localities, *k*-means.

REFERENCES

1. *Nalogovy kodeks Rossiyskoy Federatsii* [Code of Tax Laws of the Russian Federation]. Moscow, Eksmo, 2017. 1120 p.
2. Grigorev L., Golyashev A., Buryak E., Lobanova A., Kulpina V. *Byulleten o tekushchikh tendentsiyakh v rossiyskoy ekonomike* [Report on Current Trends in Russian Economy]. Available at: <http://ac.gov.ru/files/publication/a/10017.pdf> (accessed 11 November 2017).
3. *Federalnaya sluzhba gosudarstvennoy registratsii kadastra i kartografii* [Federal Service for State Registration, Cadaster and Cartography]. Available at: <https://rosreestr.ru/> (accessed 11 November 2017).
4. Kiselev V.A., Lepikhina O.Yu. Analysis of sufficiency and reliability of market information in small and average settlements in the northwest district for estimating the possibility of applying the statistical method for defining ground area cadastral cost. *Zapiski Gornogo Inzituta*, 2011, no. 7, pp. 217–221. In Rus.
5. Gladkih N.I., Kuznetsova V.V. Determination of the required number of analogues for a given number of pricing factors to estimate real estate using correlation-regression analysis. *Imushchestvennye otnosheniya v RF*, 2016, no. 6 (177), pp. 75–84. In Rus.
6. Grigorev V.V. Development of government mass appraisal in Russian Federation. *Gosudarstvennoe i munitsipalnoe upravlenie*, 2016, no. 1, pp. 83–89. In Rus.
7. Bojko A.Y. Mass appraisal. Problems and Prospects. *Zemlya i biznes*, 2016, no. 258, pp. 11–21. In Rus.
8. Ayvazyan S.A., Mkhitaryan V.S. *Teoriya veroyatnostey i prikladnaya statistika* [Theory of probability and applied statistics]. Moscow, Unity Publ., 2011. 657 p.
9. Kabacoff R. *R in action*. New York, Manning Publications Co., 2015. 563 p.
10. Everitt B.S., Hothorn T. A Handbook of Statistical Analyses Using R. *The Comprehensive R Archive Network*. Available at: https://cran.r-project.org/web/packages/HSAUR/vignettes/Ch_principal_components_analysis.pdf (accessed 10 November 2017).
11. Venables W.N., Smith D.M., the R Core Team. An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. *The Comprehensive R Archive Network*. Available at: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf> (accessed 18 November 2017).
12. Peng R. Exploratory Data Analysis with R. *LeanPub: Publish Early, Publish Often*. Available at: <https://leanpub.com/exdata> (accessed 11 November 2017).
13. Charrad M., Ghazzali N., Boiteau V., Niknafs A. NbClust: an R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 2014. Available at: <http://www.jstatsoft.org/> (accessed 11 November 2017).
14. Yatzkiv I., Gusarova L. Methods of Clusters Quantity Determination without Training. *Transport and Telecommunication*, 2003, no. 1, pp. 23–28. In Rus.
15. Leek J. The Elements of Data Analytic Style. *LeanPub: Publish Early, Publish Often*. Available at: <https://leanpub.com/datasty-le> (accessed 11 November 2017).

16. Gitis L.H. *Statisticheskaya klassifikatsiya i klasterny analiz* [Statistical classification and cluster analysis]. Moscow, Moscow Mining University Press, 2003. 157 p.
17. Wickham H., Golemund G. *R for Data Science*. Sebastopol, O'Reilly Media, 2016. 483 p.
18. Teetor P. *R Cookbook*. Sebastopol, O'Reilly Media, 2011. 397 p.
19. Burchell J., Vargas M. The Hitchhiker's Guide to Ggplot2 in R. *LeanPub: Publish Early, Publish Often*. Available at: https://leanpub.com/hitchhikers_ggplot2 (accessed 11 November 2017).
20. *Generalny plan Afoninskogo selsoveta v Kstovskom rayone Nizhegorodskoy oblasti* [Site Plan of Afoninsky Rural Council in Kstovsky District in Nizniy Novgorod Region]. Available at: http://www.kstovo-adm.ru/about/rural_councils/afoninsky_selsove/town_planning.php (accessed 20 November 2017).

Received: 21 December 2017.

Information about the authors

Elena N. Bykova, Cand. Sc., associate professor, St. Petersburg Mining University.

Tatiana I. Baltyzhakova, Cand. Sc., assistant professor, St. Petersburg Mining University.

Yana A. Volkova, postgraduate student, assistant professor, St. Petersburg Mining University.