

Научная статья

Original article

УДК 332.642

doi: 10.55186/2413046X_2025_10_3_69

**РЕГРЕССИОННЫЕ МОДЕЛИ ОЦЕНКИ РЫНОЧНОЙ СТОИМОСТИ
НЕДВИЖИМОСТИ В САНКТ-ПЕТЕРБУРГЕ**
**REGRESSION MODELS FOR ESTIMATING THE MARKET VALUE OF
REAL ESTATE IN ST. PETERSBURG**



Чесноков Евгений Александрович, к.ф.-м.н., доцент кафедры математики, Санкт-Петербургский государственный архитектурно-строительный университет, Санкт-Петербург, E-mail: eachesn@yandex.ru

Chesnokov Evgenii Alexandrovich, PhD in Physical and Mathematical Sciences, Associate Professor of the Department of Mathematics, St. Petersburg State University of Architecture and Civil Engineering, St. Petersburg, E-mail: eachesn@yandex.ru

Аннотация. Работа посвящена актуальной проблеме получения объективной массовой оценки рыночной стоимости жилой недвижимости. Цель исследования состоит в том, чтобы на примере динамично развивающегося рынка однокомнатных квартир Санкт-Петербурга получить оптимальную, с точки зрения точности, регрессионную модель оценки стоимости недвижимости. Методами регрессионного анализа проанализировано влияние различных ценообразующих факторов, включая характеристики местоположения объектов недвижимости и уровня комфортности проживания. Получены четыре оптимизированных модели оценки стоимости недвижимости. Модели различаются по факторам, характеризующим уровень комфортности проживания, а также по их спецификации:

аддитивные и мультипликативные. Тестирование полученных моделей по контрольной выборке позволило отобрать модель с максимальной точностью оценки ($MedAPE=5,2\%$) и рекомендовать ее для практического применения. В качестве примера практического использования модели рассмотрено выявление недооцененных и переоцененных квартир на основе полученных интервальных оценок среднего значения стоимости объекта недвижимости.

Abstract. The work is devoted to the urgent problem of obtaining an objective mass assessment of the market value of residential real estate. The purpose of the study is to use the example of the dynamically developing market of one-room flats in St. Petersburg to obtain an optimal, in terms of accuracy, regression model of real estate assessment. Regression analysis methods had been used to analyze the influence of various price-forming factors, including the characteristics of the location of real estate and the level of comfort of living. Four optimized models of real estate assessment have been obtained. The models differ in the factors characterizing the level of comfort of living, as well as in their structure: additive and multiplicative. Testing of the obtained models based on a control representative sample made it possible to select a model with maximum estimation accuracy ($MedAPE=5,2\%$) and recommend it for practical use. As an example of the practical use of the model, the identification of undervalued and overvalued flats based on the obtained interval estimates of the average value of the real estate have been considered.

Ключевые слова: экономико-математическое моделирование, массовая оценка недвижимости, регрессионный анализ, ценообразующие факторы

Keywords: economic and mathematical modeling, mass assessment of real estate, regression analysis, price-forming factors

Введение. На протяжении многих лет вопрос получения объективных оценок недвижимости продолжает оставаться актуальным. Причин для этого достаточно много. Оценка недвижимости необходима для определения

кадастровой стоимости, залоговой стоимости объекта недвижимости, в целях ипотечного кредитования, налогообложения, для оценок страховых компаний и т.д.

Начиная с 90-х годов, с началом активного развития рыночных отношений, в частности, с активизацией сделок по объектам жилой недвижимости, большую актуальность приобрела оценка рыночной стоимости жилья в РФ. Примерно в это же время наметилось бурное развитие вычислительной техники. Потребность в получении объективных оценок объектов жилого фонда, в совокупности со вновь открывшимися вычислительными возможностями, привело к появлению ряда научных работ, посвященных моделям массовой оценки объектов недвижимости. Массовая оценка, основанная, как правило, на использовании статистических методов анализа, имеет ряд преимуществ по сравнению с индивидуальной, осуществляемой с привлечением экспертов. Во-первых, она является существенно более дешевой. Во-вторых, в ней сводится к минимуму количество субъективных факторов, способных сильно исказить оценку. Важность получения объективных оценок объектов недвижимости неоднократно отмечалась в научной литературе [1,2]. В качестве одной из первых отмечу работу по регрессионной оценке рыночной стоимости вторичного жилого фонда Москвы, результаты которой приведены, в качестве примера, в [3].

Несмотря на то, что за последние 20 лет появилось множество работ, посвященных оценке стоимости недвижимости, интерес к этой теме продолжает оставаться высоким. С одной стороны – каждая отдельно взятая область, город или, даже, район, наряду с хорошо известными общими факторами, влияющими на формирование рыночной цены на недвижимость, имеют, как правило, свои собственные особенности, учет которых необходим для построения эффективной модели оценки стоимости жилья. В связи с этим, в ряде работ, уже известные модели адаптируются на объекты

недвижимости, относящиеся к определенному городу или району. Например, работы [4,5,6] посвящены различным аспектам построения моделей массовой оценки жилой недвижимости в Москве, в работах [7,8] рассматриваются вопросы оценки стоимости жилой недвижимости в Самаре, работа [9] посвящена моделированию цен на недвижимость в Елабуге, [10] – в Нижнем Новгороде, [11] – в Ростове-на-Дону. Вопросам оценки недвижимости в Санкт-Петербурге посвящены работы [1,12]. С другой стороны – растет и разнообразие используемых моделей. Все чаще стали привлекаться к рассмотрению модели нелинейных регрессий [2,9], модели порядковой логистической регрессии [6]. Наряду с классическими моделями множественной регрессии по разным ценообразующим факторам, появились модели пространственной оценки, использующие в качестве объясняющих ценообразование переменных как факторы удаленности объекта до некоторых «мест притяжения» [4,13] (например, центра города, метро, объектов инфраструктуры, культурных объектов и т.д.), так и, непосредственно, географические координаты объектов [1,9]. Появились модели, учитывающие комфортность проживания в том или ином объекте жилой недвижимости [6,13], включающие такие факторы как экология, близость к индустриально – промышленным объектам, уровень шума и др. Существенно расширился перечень методов, используемых при оценке недвижимости. Появились так называемые «аналитические модели» оценки недвижимости, построенные на основе таких методов, как искусственные нейронные сети, методы пространственного анализа, экспертные системы. Достаточно широкое распространение получили модели машинного обучения [10,11].

Каждый класс моделей имеет свои достоинства и недостатки. Тем не менее, наиболее широко применимым классом моделей оценки недвижимости по сей день остаются классические регрессионные модели [2,5,14]. Этому есть, как минимум, три объяснения. Модели достаточно

простые, не требующие привлечения специального программного обеспечения. Модели, дающие неплохую, значимую по факторам, оценку стоимости недвижимости даже при не слишком большом объеме выборки, в отличие, скажем, от моделей машинного обучения. Модели позволяют дать ясную интерпретацию влияния своих факторов и предоставляют возможность не только оценки стоимости объекта недвижимости, но и возможность качественной и количественной оценки степени влияния факторов на ценообразование.

Рынок жилой недвижимости в Санкт-Петербурге, как и во многих городах РФ, динамично развивается. Это второй по масштабу рынок в РФ, после Москвы. По данным Росстата, ввод жилья в эксплуатацию начал резко расти, начиная с 2003 года. В 2004 г. объем жилья, введенного в эксплуатацию, превысил 2 млн. м², с 2007 г. – превысил 2,5 млн. м², начиная с 2014 г. – практически не опускался ниже отметки в 3 млн. м². Снижение на величину порядка 30% произошло только в 2024 г. Разнообразие выставляемых на рынке вторичного фонда объектов недвижимости в Санкт-Петербурге, как в старом фонде, так и в фонде недавно введенного в эксплуатацию жилья, создает дополнительные сложности, как для формирования репрезентативной выборки, так и для построения качественной модели оценки стоимости недвижимости.

Цель исследования – получить оптимальную, с точки зрения точности, регрессионную модель оценки стоимости недвижимости вторичного жилого фонда в Санкт-Петербурге. К целям работы относится также выявление на основе полученной модели, путем построения интервальных оценок среднего значения стоимости, классов недооцененных и переоцененных квартир на представленном рынке. Отмечу, что использование интервальных оценок, несмотря на явную его необходимость для объективного выявления недооцененных и переоцененных квартир, является среди перечня работ по данной тематике большой редкостью.

Материал и методы исследования. В качестве объекта исследования был выбран рынок однокомнатных квартир Приморского района Санкт-Петербурга. Обращение к Приморскому району не случайно. Это один из наиболее интенсивно застраиваемых районов города. Спектр жилья, представленного на вторичном рынке, как по качеству, так и по цене, весьма широкий. На сегодняшний день, число предложений на вторичном рынке жилья в Приморском районе – максимальное в городе. Цены на жилье, с учетом факторов года введения в эксплуатацию, комфортности проживания и престижности различаются более, чем в два раза за м².

Работа основана на использовании методов регрессионного, дисперсионного и корреляционного анализов. Сравниваются общие для столь разнообразного по ценообразующим факторам рынка модели. Предполагается, что модели должны быть достаточно полными, т.е. включать в себя, наряду со стандартными (указанными в техническом паспорте квартиры) ценообразующими факторами, факторы пространственного расположения объекта недвижимости и факторы, учитывающие комфортность проживания.

Результаты и их обсуждение. Для построения и последующего тестирования моделей по данным сайта bn.ru была сформирована выборка из выставленных на продажу квартир на вторичном рынке Приморского района Санкт-Петербурга. После удаления 3% самых дешевых и 3% самых дорогих квартир, в ней осталось 152 объекта недвижимости, которые, с применением генератора случайных чисел в каждой из ценовых подгрупп, были поделены на две части. В основную выборку, используемую для оценки параметров моделей и их статистических характеристик, вошли 132 квартиры, 20 других попали в тестовую выборку, предназначенную для тестирования точности моделей. Квартиры оказались в ценовом диапазоне от 6 до 15,2 млн. руб., с еще большей вариацией по цене за м²: от 155 до 432 тыс. руб. В качестве ценообразующих факторов первоначально были взяты практически все

факторы, которые были указаны на сайте в качестве общих характеристик квартиры. В их число вошли 5 непрерывных и 9 фиктивных (бинарных) переменных, все они приведены в таблице 1. Параметры оценены по методу наименьших квадратов в рамках линейной спецификации модели.

Таблица 1 – Перечень факторов, способных влиять на ценообразование, с указанием МНК оценок коэффициентов и их р-значений по t-статистике Стьюдента.

N	Обозначение	Описание фактора	Коэффициент	р-значение
	1	Свободный член в уравнении	-135153	0,000181401
1	So, м ²	Общая площадь	117,4341	0,001412587
2	Sk, м ²	Площадь кухни	77,53569	0,209657597
3	hp, м	Высота потолков	1415,9	0,310716024
4	lm, км	Расстояние до метро	-699,112	1,04714E-11
5	год	Год ввода в эксплуатацию	67,79473	0,000203628
6	тд	Тип дома (0 – панельные, 1 – кирпичные и монолитные)	909,5216	0,041355045
7	эт	Этаж (0 – первый или последний, 1 – остальные)	52,84131	0,909913575
8	су	Санузел (0 – совмещенный, 1 – отдельный)	-375,983	0,28900378
9	бл	Наличие балкона или лоджии (0 – нет, 1 – есть)	-218,962	0,459849651
10	рем	Наличие обычного ремонта (0 – нет, 1 – есть)	-26,2258	0,937269349
11	евро	Наличие евроремонта (0 – нет, 1 – есть)	-339,074	0,400003529
12	диз	Наличие ремонта по дизайнерскому проекту (0 – нет, 1 – есть)	2443,351	0,002116613
13	видул	Вид из окна (0 – только во двор, 1 – есть вид на улицу)	-589,634	0,180167801
14	прод	Вид продажи (0 – встречная, 1 – прямая)	311,6013	0,439009508

Как видно из таблицы, значимыми на 5-процентном уровне оказались только пять факторов: общая площадь, расстояние до метро, год постройки,

тип дома и наличие дизайнерского ремонта. Модель значима по F-статистике на уровне $P_F = 1,3 \cdot 10^{-19}$, однако коэффициент детерминации у нее относительно низкий $R^2 = 0,64$. Можно заметить, что некоторые незначимые параметры имеют отрицательное значение, противоречащее какой-либо логике. Например, отрицательное значение у параметра 9 означает, что наличие балкона или лоджии удешевляет стоимость квартиры в среднем на 219 тыс. руб., отрицательные значения параметров 10 и 11 говорят о том, что средняя стоимость квартир с ремонтом на 26 тыс. руб., а с евроремонтом – даже на 339 тыс. руб. ниже средней стоимости квартир без ремонта. Другие факторы, такие как площадь кухни, высота потолков и вид продажи имеют разумные с точки зрения знака значения, но, тем не менее, не являются значимыми.

Значительные искажения в уровень значимости и даже знаки коэффициентов может внести явление скрытой мультиколлинеарности факторов [3]. Для проверки параметров на возможную мультиколлинеарность были рассчитаны коэффициенты корреляции между всеми факторами. Результаты приведены в таблице 2.

Таблица 2 – Значения коэффициента корреляции между ценообразующими факторами

Корреляция	So, м2	Sk, м2	hп, м	lм, км	год	тд	эт	су	бл	рем	евро	диз	видул	прод
So, м2	1													
Sk, м2	0,5165	1												
hп, м	0,1669	0,3297	1											
lм, км	-0,1608	0,1365	0,0876	1										
год	0,0297	0,4567	0,3958	0,3761311	1									
тд	-0,004	0,3215	0,3418	0,21167875	0,6518	1								
эт	0,0647	-0,1353	-0,0726	0,06470317	0,0791	0,0083	1							
су	0,1485	-0,2248	-0,2529	-0,1681221	-0,5671	-0,5466	0,092	1						
бл	-0,0112	-0,0291	-0,067	0,08384157	-0,0382	-0,0238	0,2929	0,0795	1					
рем	-0,0367	-0,2005	-0,2776	0,13541214	-0,3748	-0,2952	-0,1081	0,235	-0,0825	1				
евро	0,0579	0,151	0,1651	0,00389073	0,0177	0,0152	0,0826	0,0804	0,1637	-0,3676	1			
диз	-0,0951	-0,0833	-0,0255	-0,1406407	-0,0381	-0,0199	0,0609	-0,0339	-0,069	-0,1494	-0,0769	1		
видул	0,0729	0,046	-0,1043	-0,1289985	-0,095	0,0527	-0,0412	0,0637	0,1297	-0,0416	0,1192	0,0826	1	
прод	-0,1669	0,0367	0,1669	0,17123876	0,2714	0,1816	-0,0652	-0,2265	0,0129	-0,1567	-0,0686	0,0702	-0,1499	1

Как видно из таблицы, имеется достаточно заметная (но не критичная с точки зрения мультиколлинеарности факторов) положительная корреляция между общей площадью и площадью кухни, годом постройки и площадью кухни, годом постройки и типом дома, а также отрицательная корреляция между раздельностью санузла и годом постройки, а также, типом дома. Положительная корреляция между общей площадью и площадью кухни логически понятна. Возможно, именно она нивелирует зависимость цены квартиры от площади кухни и делает данный фактор незначимым. Положительная корреляция между годом постройки и площадью кухни тоже понятна: в домах 70-х – 80-х годов преобладали маленькие кухни. Положительная корреляция между годом постройки и типом дома свидетельствует о том, что современные дома в районе делают преимущественно кирпичными, кирпично-монолитными и монолитными, в то время как среди относительно старых домов имеется достаточно большой процент панельных. Отрицательная корреляция раздельности санузла с типом дома и годом постройки говорит о том, что в более современных домах, среди которых преобладают кирпичные и монолитные, санузел, в целях увеличения его площади, чаще всего делают совместным. Несколько удивляет незначимость высоты потолков, которая особенно ни с какими другими факторами не коррелирует. Как показало детальное рассмотрение, возможное объяснение состоит в следующем. Все квартиры с максимальной высотой потолков в 3 метра находятся, видимо, в не очень престижных, возможно, не очень удачно расположенных домах и имеют среднюю стоимость заметно меньшую, чем средняя стоимость квартир с высотой потолков в 2,7-2,8 метра.

В ряде работ [2,12] в качестве объясняемой переменной предпочтение отдавалось удельной стоимости недвижимости за m^2 . Было отмечено, что при анализе удельной стоимости, в отличие от абсолютной, в некоторых случаях удается выявить дополнительные значимые факторы. Регрессия удельной

стоимости квартиры по всем включенным в рассмотрение факторам была выполнена и в настоящей работе, однако набор значимых факторов остался тем же, примерно с теми же уровнями значимости, а коэффициент детерминации лишь немного уменьшился до значения $R^2 = 0,59$. Модель с удельной стоимостью в качестве объясняемой переменной не показала каких-либо преимуществ. Тем не менее, она подтвердила хорошо известный факт уменьшения стоимости одного квадратного метра с ростом общей площади жилья. Параметр при переменной S_0 равен $-3,18$ при p -значении по t -статистике $p=0,0015$, т.е. цена за m^2 в среднем снижается на 3,18 тыс. руб. при увеличении общей площади на один метр.

В работах [1,2] было отмечено, что в ряде случаев объективно учесть влияние некоторых факторов возможно только в рамках нелинейных спецификаций. Для выявления возможных нелинейных зависимостей цены квартиры от факторов, принимающих непрерывный ряд значений, были рассмотрены точечные диаграммы корреляции между ценой квартиры и соответствующим фактором. Зависимость цены квартиры от года постройки приведена на рисунке 1.

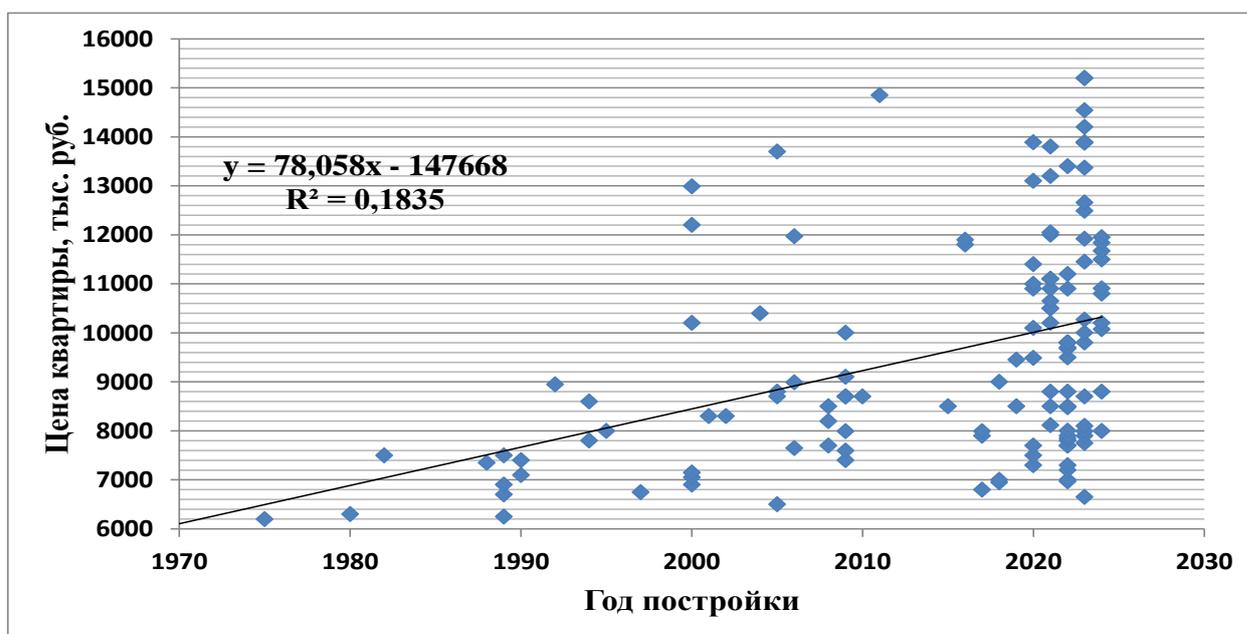


Рисунок 1 – Диаграмма зависимости цены квартиры от года введения дома в эксплуатацию

Из диаграммы видно, что большинство выставляемых на продажу квартир относятся к домам, введенным в эксплуатацию начиная с 2020 года. С одной стороны, введения жилья в эксплуатацию за эти годы шло с максимальной интенсивностью, с другой – явно наметилась тенденция вложений денежных средств в недвижимость именно с целью получения дохода. Причина этого достаточно очевидна: средний ежегодный прирост стоимости жилой недвижимости в Санкт-Петербурге за период 2000-2003 гг. составил величину порядка 24% (заметное снижение произошло только в 2024 г.), при средней ставке ипотечного кредитования на уровне 7,5-8,5%. Как показывает визуальный анализ диаграммы, вряд ли можно отдать предпочтение какой-либо нелинейной спецификации зависимости цены квартиры от года постройки. Рассмотрение парных нелинейных регрессий простейшего вида показало, что квадратичная зависимость ничуть не улучшает ситуации, использование показательной или степенной зависимости приводит лишь к незначительному увеличению коэффициента детерминации (с 0,18 он возрастает до 0,22), при этом линия регрессии визуально почти не меняется. Отметим также, что коэффициент при факторе года постройки в случае парной регрессии (78) не сильно отличается от соответствующего коэффициента в модели множественной регрессии по всем факторам (68).

На рисунке 2 приводится зависимость цены квартиры от общей площади.

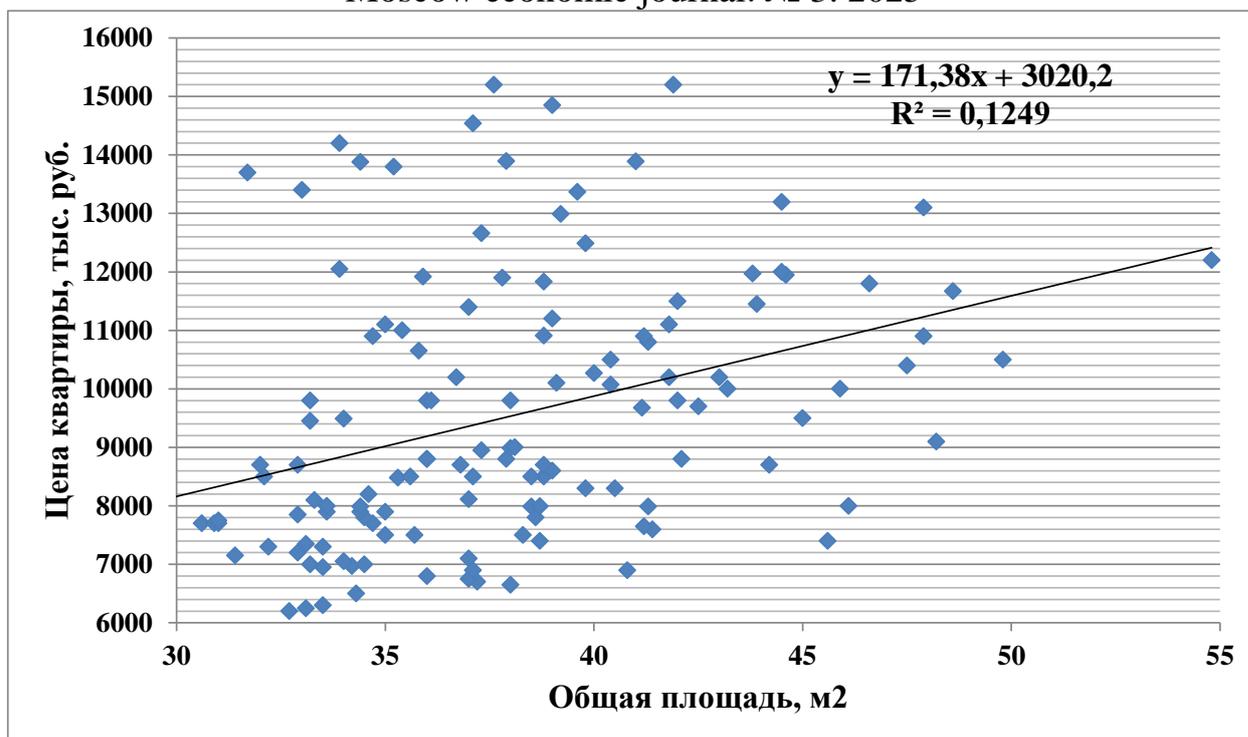


Рисунок 2 – Диаграмма зависимости цены квартиры от общей площади

Из диаграммы видно, что цена квартиры очень слабо коррелирует с общей площадью, причем максимальный разброс цен приходится на квартиры с площадью порядка 35-40 м². Использование какой-либо нелинейной спецификации здесь тоже представляется необоснованным. Как показывает рассмотрение простейших нелинейных регрессий, только показательная и степенная зависимости приводят к небольшому увеличению коэффициента детерминации (0,15 вместо 0,12 для линейной). Сама линия регрессии визуальнo при этом опять практически не меняется.

Рисунок 3 иллюстрирует влияние общей площади на стоимость одного квадратного метра жилья.

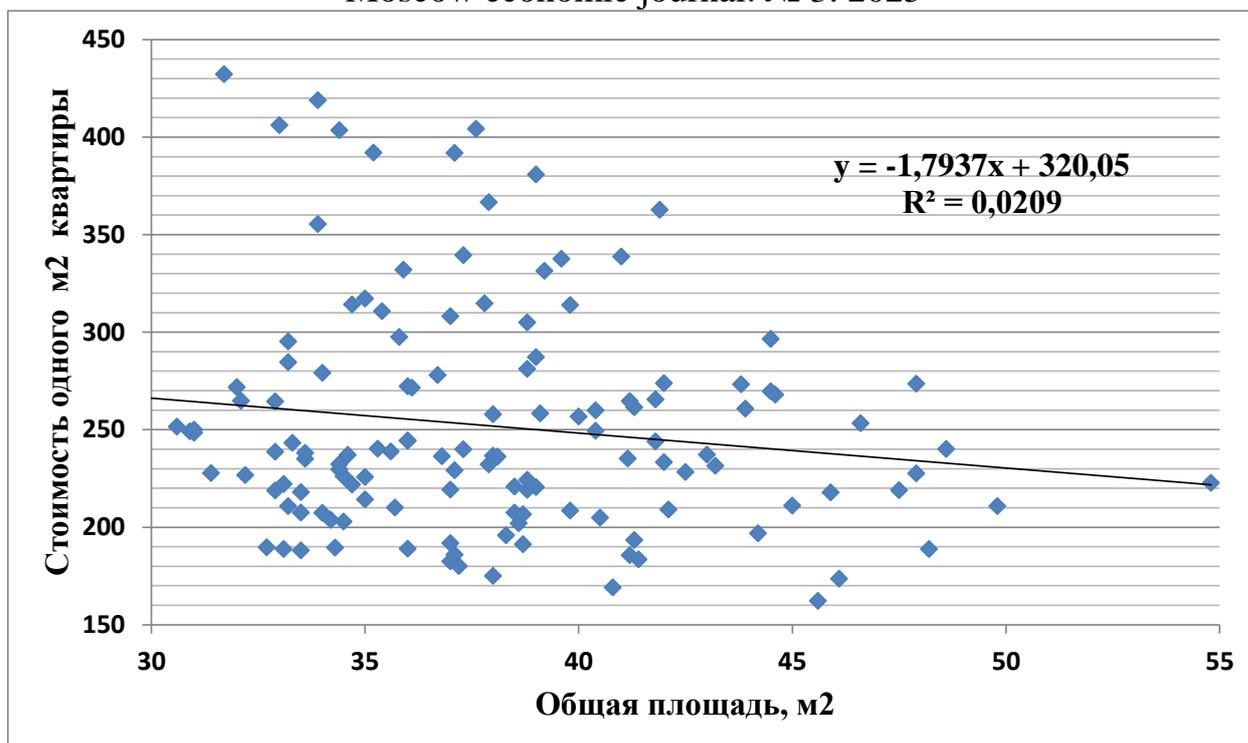


Рисунок 3 – Диаграмма зависимости стоимости одного квадратного метра жилья от общей площади

Как уже отмечалось, величина стоимости одного м² жилья очень сильно варьируется. Коэффициент детерминации парной регрессии стоимости одного м² жилья от общей площади составляет всего лишь 2%, значения коэффициентов при факторе общей площади в моделях парной и множественной регрессий тоже сильно различаются (–1,8 для парной и –3,2 для множественной). Тем не менее, в обоих случаях зависимость оказывается значимой и подтверждает уменьшение средней стоимости одного квадратного метра жилья с ростом общей площади.

На рисунке 4 приведена диаграмма зависимости цены квартиры от расстояния до метро.

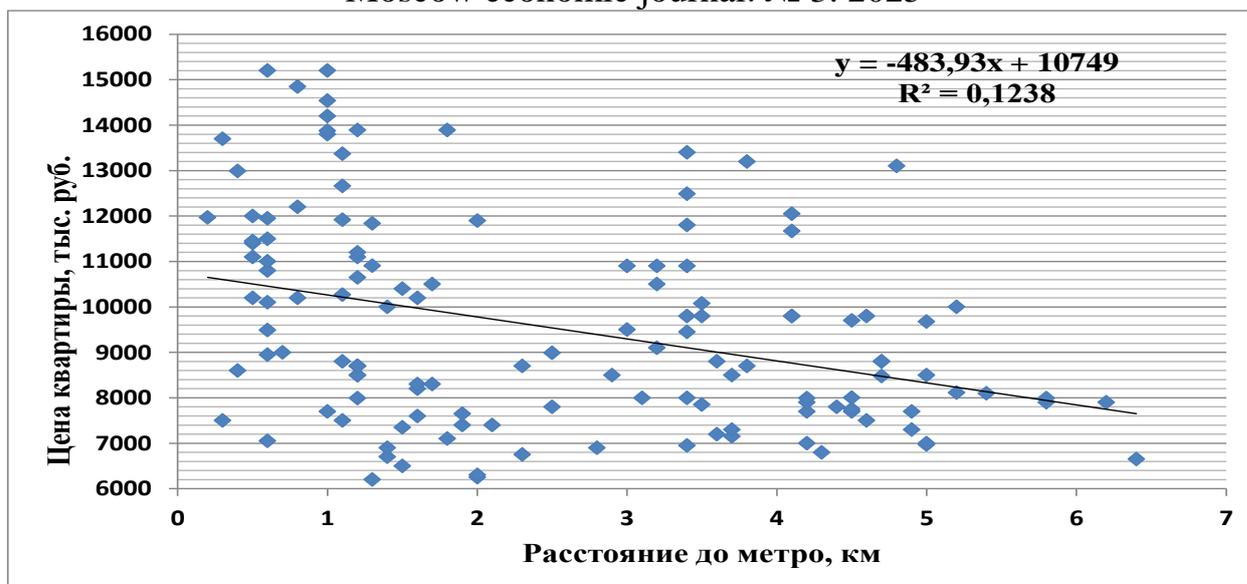


Рисунок 4 – Диаграмма зависимости цены квартиры от расстояния до метро

Диаграмма, как и следовало ожидать, демонстрирует снижение средней стоимости квартиры по мере удаления ее расположения от метро. Никаких оснований отдать здесь предпочтение какой-либо нелинейной спецификации опять-таки не прослеживается. При переборе всех простейших функций в качестве спецификации для парной регрессионной зависимости, коэффициент детерминации практически не меняется. Отметим, что коэффициент при факторе l_m , для парной и множественной регрессии различается весьма существенно (-484 для парной и -699 для множественной).

В итоге, по результатам проведенного анализа, предпочтение было отдано линейной регрессионной модели цены квартиры по ряду ценообразующих факторов. Далее, реализуя пошаговую процедуру исключения наиболее (по t -статистике) незначимого фактора, с пошаговым контролем матрицы коэффициентов корреляции на случай возможного возникновения мультиколлинеарности, была, окончательно получена модель, в которой все оставшиеся факторы оказались значимы на 5% уровне. Значимыми остались все те же пять факторов. Значение коэффициентов приведено в таблице 3.

Таблица 3 – Факторы, значения параметров при них и р-значения для аддитивной линейной регрессионной модели, после удаления всех незначимых на 5% уровне факторов

N	Обозначение	Описание фактора	Коэффициент	р-значение
	1	Свободный член в уравнении	-177144	6,55E-09
1	So, м ²	Общая площадь	133,8973	4,5E-06
2	lm, км	Расстояние до метро	-696,95	3,79E-13
3	год	Год ввода в эксплуатацию	90,58556	3,94E-09
4	тд	Тип дома (0 – панельные, 1 – кирпичные и монолитные)	1041,177	0,013534
5	диз	Наличие ремонта по дизайнерскому проекту (0 – нет, 1 – есть)	2527,941	0,001051

Окончательное уравнение регрессии в рамках линейной спецификации цены квартиры по пяти значимым факторам имеет вид

$$P = -177144 + 133,9 So - 697 lm + 90,586 год + 1041 тд + 2528 диз. \quad (1)$$

Коэффициент детерминации для него равен 0,60.

Матрица коэффициентов корреляции между пятью оставшимися факторами приведена в таблице 4.

Таблица 4 – матрица коэффициентов корреляции для факторов из уравнения (1)

Корреляция	So	lm	год	тд	диз
So	1				
lm	-0,1608	1			
год	0,0297	0,3761	1		
тд	-0,004	0,2117	0,6518	1	
диз	-0,0951	-0,1406	-0,0381	-0,01992	1

Как видно из таблицы, существенная (но не критичная) по величине положительная корреляция имеет место только между переменными *год* и *тд*.

Другой популярной спецификацией для оценки стоимости недвижимости (смотри, например, [3]) является мультипликативная,

$$P = C \cdot S_0^{\alpha_1} \cdot l_M^{\alpha_2} \cdot \text{год}^{\alpha_3} \cdot e^{\alpha_4 \text{тд}} \cdot e^{\alpha_5 \text{диз}}, \quad (2)$$

или, что то же самое, логарифмическая модель

$$\ln P = \alpha_0 + \alpha_1 \ln S_0 + \alpha_2 \ln l_M + \alpha_3 \ln \text{год} + \alpha_4 \text{тд} + \alpha_5 \text{диз}, \quad (3)$$

в которой коэффициенты при логарифмах непрерывных факторов представляют собой эластичности цены по соответствующему фактору (т.е. показывают, на сколько процентов изменится цена при увеличении соответствующего фактора на один процент), а коэффициенты при бинарных факторах показывают, на сколько процентов изменится цена при наличии соответствующего признака.

Оцененное уравнение логарифмической регрессии, также по пяти значимым факторам, имеет вид

$$\ln P = -130,3 + 0,622 \ln S_0 - 0,132 \ln l_M + 18,0 \ln \text{год} + 0,106 \text{тд} + 0,215 \text{диз}. \quad (4)$$

Коэффициент детерминации для него несколько выше (0,64).

Поскольку коэффициенты детерминации моделей оказались относительно низкими, была предпринята попытка, основываясь теперь на описании каждой конкретной квартиры, учесть дополнительные ценообразующие факторы, характеризующие расположение дома и уровень комфорта проживания. Кроме того, в целях создания более репрезентативной

выборки, из нее была удалена часть квартир, расположенных по одному адресу или в близко расположенных домах и имеющих во многом схожие значения основных факторов. В итоге, была сформирована основная выборка объемом в 81 объект недвижимости. Ценообразующие факторы, учитываемые в новой модели, значения параметров при них в рамках линейной спецификации и их р-значения приведены в таблице 5.

Таблица 5 – Перечень вновь отобранных факторов, способных влиять на ценообразование, с указанием МНК оценок коэффициентов линейной регрессии и их р-значений по t-статистике Стьюдента.

N	Обозначение	Описание фактора	Коэффициент	р-значение
	1	Свободный член в уравнении	-99496,9	0,019347772
1	So, м ²	Общая площадь	97,0145	0,026562691
2	Sk, м ²	Площадь кухни	-0,32951	0,995869455
3	lm, км	Расстояние до метро	-627,736	2,08204E-05
4	lp, км	Расстояние до ближайшего парка	155,3732	0,636268185
5	год	Год сдачи дома в эксплуатацию	52,11985	0,014738587
6	тд	Тип дома (0 – панельные, 1 – кирпичные и монолитные)	664,0902	0,123086191
7	комф	Дома повышенной комфортности (0 – обычный, 1 – повышенной)	1265,047	0,012944705
8	биз	Дома бизнес класса (0 – обычный, 1 – бизнес класса)	3025,757	4,66756E-05
9	парк	Наличие собственной подземной парковки (0 – нет, 1 – есть)	-487,627	0,287177622
10	рем	Наличие обычного ремонта (0 – нет, 1 – есть)	94,03058	0,843039842
11	евро	Наличие евроремонта (0 – нет, 1 – есть)	510,136	0,310219033
12	диз	Наличие ремонта по дизайнерскому проекту (0 – нет, 1 – есть)	1997,858	0,001061248
13	вид	Видовая квартира (1 – из окна открывается красивый вид, 0 – нет)	5,234958	0,98928914

Как видно из таблицы, значимыми на 5-процентном уровне опять оказываются коэффициенты при практически том же наборе факторов. Правда, чисто строительный фактор – тип дома заменяется теперь на два значимых фактора, 7 и 8, более детально характеризующие дом с точки зрения престижности и комфортности в нем проживания, при этом коэффициент детерминации возрастает до значения $R^2 = 0,69$. Как ни странно, фактор 4, характеризующий близость дома к парковой зоне, оказывается незначимым. По-видимому, большую роль в ценообразовании играет достаточно субъективный и трудно поддающийся учету фактор престижности микрорайона. Матрица коэффициентов корреляции между факторами содержит значения, не превосходящие 0,6, что говорит об отсутствии существенной мультиколлинеарности между ними.

Как и ранее, реализуя пошаговую процедуру исключения наиболее незначимых факторов, приходим к значимой по всем оставшимся факторам модели. Все значимые факторы, коэффициенты при них и р-значения приведены в таблице 6.

Таблица 6 – Перечень значимых ценообразующих факторов с указанием МНК оценок коэффициентов и их р-значений по t-статистике Стьюдента.

N	Обозначение	Описание фактора	Коэффициент	р-значение
	1	Свободный член в уравнении	-115090	0,001491426
1	$S_0, м^2$	Общая площадь	94,18695	0,005691496
2	$l_m, км$	Расстояние до метро	-673,145	9,28155E-08
3	год	Год сдачи дома в эксплуатацию	60,31823	0,000966642
4	комф	Дома повышенной комфортности (0 – обычный, 1 – повышенной)	1148,368	0,011828548
5	биз	Дома бизнес класса (0 – обычный, 1 – бизнес класса)	2781,324	6,81722E-06
6	диз	Наличие ремонта по дизайнерскому проекту (0 – нет, 1 – есть)	1836,401	0,000488644

Оцененная линейная регрессионная модель по новым шести факторам имеет вид

$$P = -115090 + 94,2 S_o - 673 l_m + 60,318 \text{ год} + 1148 \text{ комф} + 2781 \text{ биз} + 1836 \text{ диз}, \quad (5)$$

коэффициент детерминации ее равен 0,67 при р-значении по F-статистике $P_F = 9 \cdot 10^{-16}$. Тот же набор значимых факторов, примерно на том же уровне значимости, получается и для модели с удельной стоимостью (за м²) в качестве объясняемой переменной. Наибольший коэффициент корреляции в данной модели (0,52) имеет место между факторами года постройки и повышенной комфортности дома, что вполне соответствует истине: в последние годы процент жилья повышенной комфортности заметно возрос, что говорит об ориентации застройщиков на класс более состоятельных клиентов. Потребности менее обеспеченных клиентов отчасти удовлетворяются при этом квартирами – студиями.

Оцененная логарифмическая модель, построенная по факторам из таблицы 6, имеет вид

$$\ln P = -92,0 + 0,46 \ln S_o - 0,15 \ln l_m + 13,07 \ln \text{год} + 0,13 \text{ комф} + 0,28 \text{ биз} + 0,17 \text{ диз}, \quad (6)$$

Коэффициент детерминации ее равен 0,74, $P_F = 6,5 \cdot 10^{-20}$.

Шестифакторная логарифмическая модель имеет наилучшие статистические характеристики, однако это еще не означает, что она обладает наивысшей точностью оценки. Для объективной проверки качества оценивания все четыре оптимизированные по факторам модели были протестированы на основе дополнительной выборки, состоящей, как отмечалось ранее, из 20 случайным образом отобранных квартир. Результаты приведены в таблице 7.

Таблица 7 – Основные статистические характеристики и показатели точности оценивания для четырех, значимых по своим факторам, моделей

Название модели	Уравнение регрессии	R^2	R^2_{adj}	P_F	MAPE %	MedAPE %	MaxAPE %
5-факторная линейная	(1)	0,60	0,59	$9 \cdot 10^{-24}$	9,0	6,6	30
5-факторная логарифмическая	(4)	0,64	0,62	$5 \cdot 10^{-26}$	8,5	7,9	22
6-факторная линейная	(5)	0,67	0,64	$9 \cdot 10^{-16}$	7,5	6,2	24
6-факторная логарифмическая	(6)	0,74	0,72	$7 \cdot 10^{-20}$	7,5	5,2	19

В таблице, наряду с тремя статистическими характеристиками, полученными в ходе оценивания моделей по основной выборке (коэффициент детерминации R^2 , скорректированный коэффициент детерминации R^2_{adj} и р-значение по F-статистике Фишера P_F), приведены три показателя точности оценивания, полученные по тестовой выборке (MAPE - средняя абсолютная процентная ошибка, MedAPE – медианная абсолютная процентная ошибка и MaxAPE – максимальная абсолютная процентная ошибка). По всем показателям наилучшей оказывается 6-факторная логарифмическая модель, которая и будет выбрана для дальнейшего использования.

В качестве примера использования отобранной модели рассмотрим актуальную для рынка недвижимости задачу выявления недооцененных и переоцененных объектов. Данный вопрос рассматривался, например, в работах [5,8]. Правда, в работе [5] проводилось сравнение выставочной стоимости квартиры с точечной оценкой, полученной по уравнению регрессии. При этом все квартиры оказывались либо недооцененные, либо переоцененные и те, которые представляют наибольший интерес для покупателя с точки зрения стоимости, отбирались (среди класса недооцененных квартир) по максимальным значениям абсолютной

процентной ошибки. В работе [8], на основе использования порядковой логистической регрессии, весь ценовой диапазон был разбит на 6 интервалов, средние из которых имели ширины порядка 1; 1; 1,5; 4,5 млн. руб. В этой работе, однако, непопадание квартиры в предсказанный интервал интерпретировалось не как выявление недооцененной или переоцененной квартиры, а как ошибка прогноза.

В настоящей работе процедура выявления недооцененных и переоцененных квартир осуществлялась на основе сравнения выставочной стоимости квартиры с интервальной оценкой среднего значения стоимости, полученной на 5-процентном уровне значимости. К переоцененным относились квартиры с выставочной стоимостью выше верхней границы доверительного интервала, к недооцененным – с выставочной стоимостью ниже нижней границы интервала, при этом отмечался процент, на который выставочная стоимость превысила верхнюю границу или, наоборот, оказалась ниже нижней границы. Результаты приведены в таблице 8.

Таблица 8 – Выявление недооцененных и переоцененных квартир на основе интервальной оценки среднего значения стоимости по модели (6). Уровень значимости $\alpha = 5\%$. Все стоимости приведены в тыс. руб.

N	Rвыст	Rпрогн	Ширина интервала	Rmin	Rmax	Оценка стоимости	%
1	6200	6266,283	1164,83	5710,876	6875,705	объект	
2	6300	6553,379	800,9777	6165,116	6966,094	объект	
3	6750	7030,688	656,5345	6710,08	7366,615	объект	
4	7000	7367,304	1040,619	6865,344	7905,963	объект	
5	7000	8075,8	926,9168	7625,629	8552,546	заниж	-7,74696
6	7300	7129,956	1095,799	6603,077	7698,876	объект	
7	7500	7577,65	839,088	7169,711	8008,799	объект	
8	7900	7698,018	975,2482	7225,822	8201,07	объект	
9	8600	9200,614	1204,263	8618,165	9822,428	заниж	-0,19743
10	8700	8036,866	909,4243	7595,007	8504,431	завыш	2,433398
11	9200	9916,765	2214,369	8871,197	11085,57	объект	
12	9500	9854,53	1361,364	9197,328	10558,69	объект	

13	9800	10008,94	1713,291	9188,887	10902,18	объект	
14	10200	11920,95	1877,952	11018,9	12896,85	заниж	-6,86942
15	10740	10192,18	1115,988	9649,446	10765,43	объект	
16	11100	10022,13	1105,246	9484,734	10589,98	завыш	5,088934
17	11920	12196,43	2192,265	11149,45	13341,72	объект	
18	12200	10070,97	1958,835	9139,062	11097,9	завыш	10,94337
19	13700	11103,94	2547,408	9903,047	12450,45	завыш	11,25317
20	14850	12411	2374,5	11280,4	13654,9	завыш	9,629343

Как видно из таблицы, 60% квартир имеют объективную стоимость, 15% - заниженную относительно допустимого интервала и 25% - завышенную. Максимальное превышение цены составляет 11,3% от прогнозной, максимальное занижение цены – 7,7% от прогнозной.

Выводы. В работе, на основе собранных статистических данных о выставленных на продажу однокомнатных квартирах Приморского района Санкт-Петербурга, в рамках классических регрессионных моделей, было проанализировано влияние различных ценообразующих факторов на стоимость недвижимости. Совокупность рассмотренных факторов включает в себя как стандартные характеристики квартиры, так и переменные, значение которых может быть определено только при изучении описания квартиры. Из числа рассмотренных, значимое влияние на формирование цены квартиры оказывают следующие факторы: общая площадь, удаленность от метро, год ввода дома в эксплуатацию, факторы, характеризующие качество дома и комфорт проживания в нем (тип дома или, при более детальном описании, факторы, выделяющие дома повышенной комфортности и дома бизнес-класса) и фактор наличия дизайнерского ремонта в квартире. Влияние всех перечисленных факторов имеет ясную экономическую интерпретацию. Стоит отметить и то, что влияние некоторых факторов, отмеченных в ранее опубликованных работах (таких как площадь кухни, наличие евроремонта, вид продажи, вид из окна, близость к парковой

зоне) на стоимость квартиры, не нашло своего статистического подтверждения в настоящей работе.

Были проанализированы регрессионные модели в рамках различных, линейных и нелинейных, спецификаций. Наилучшей, как по своим статистическим характеристикам, так и по оценкам ее прогностического качества на основе специально сформированной тестовой выборки, признана шестифакторная логарифмическая модель (6). Модель демонстрирует весьма высокую точность прогноза ($MAPE = 7,5\%$, $MedAPE = 5,2\%$, $MaxAPE = 19\%$), она на 74% объясняет полную наблюдаемую вариацию цен на недвижимость и может быть рекомендована для практического использования. Для сравнения, ряд моделей из ранее опубликованных работ, различающихся как по методам их построения, так и по составу статистических данных, собранных на базе статистик для различных городов РФ, демонстрируют следующие показатели точности прогноза: $MaxAPE = 33\%$ [5], $MAPE = 14\%$ [7], $MedAPE = 6\%$ [10], $MedAPE = 13\%$ [11]. В работе [1] было отмечено, что при построении экономико-математических моделей оценки стоимости недвижимости приемлемыми могут быть признаны модели с $MAPE$, не превосходящим 15%. Из числа работ, известных автору, только в работе [12] была указана большая, чем в настоящей работе, прогностическая точность модели ($MAPE=0,5\%$). Столь высокий показатель точности, вызывает, однако, некоторые сомнения в репрезентативности тестовой выборки, поскольку в данной работе в качестве ценообразующих факторов учитывались только фактор пространственного расположения недвижимости (правда на весьма высоком уровне детализации) и общая площадь. Никакие факторы, характеризующие конкретную квартиру в данном доме, такие, например, как однозначно (по результатам настоящего исследования) влияющий на ценообразование фактор наличия дизайнерского ремонта, в работе учтены не были.

Построенная модель была применена для выявления недооцененных и переоцененных квартир. Модель выявила 15% недооцененных и 25% переоцененных квартир на 5-процентном уровне значимости. В отличие от работы [5], выявление недооцененных и переоцененных квартир производилось на основе сравнения выставочной стоимости с объективно полученной интервальной оценкой для среднего значения цены квартиры, а не на основе относительно субъективного мнения о большом или малом процентном завышении (занижении) выставочной стоимости относительно точечного прогнозного значения. Стоит отметить, что ширина доверительных интервалов варьируется в диапазоне от 0,66 до 2,55 млн. руб. Данные значения неопределенности оценки не представляются слишком большими. Для сравнения, в работе [8], в рамках более сложной (как с точки зрения сбора статистической информации, так и с точки зрения ее вычислительной обработки) модели, прогнозировалось попадание рыночной стоимости квартиры в ценовые интервалы шириной от 1 до 4,5 млн. руб. при, приблизительно, том же диапазоне цен, что и в настоящей работе.

В заключении хотелось бы отметить, что ситуация на рынке недвижимости достаточно быстро меняется. Значения коэффициентов в полученной модели имеют смысл скорее как приблизительные количественные оценки степени влияния того или иного фактора на стоимость квартиры. Уже через пару месяцев значения их могут существенно измениться. Да и сам набор значимых ценообразующих факторов, скорее всего, может оказаться другим для рынка недвижимости в другом городе или, даже, районе. Кроме того, он тоже может меняться с течением времени. Значительно большую практическую ценность имеют проведенный в работе анализ степени влияния различных факторов на ценообразование, анализ степени корреляции этих факторов и сравнительный анализ различных, линейных и нелинейных, спецификаций модели. Хотелось бы еще раз отметить возможную высокую

прогностическую точность классических регрессионных моделей, а также необходимость интервального оценивания для объективного выявления недооцененных и переоцененных квартир. Естественно, как, наверное, и любая другая, предложенная в настоящей работе модель может быть улучшена. Здесь прослеживаются, по крайней мере, два пути. Во-первых, выявление новых, значимых для ценообразования, факторов позволило бы повысить коэффициент детерминации модели. Хотя, как представляется автору, одним из наиболее существенных для ценообразования является весьма субъективный и весьма сложно поддающийся оценке фактор престижности проживания в конкретном доме или микрорайоне. Во-вторых, рисунок 1 демонстрирует явную гетероскедастичность цены квартиры по году постройки. С одной стороны, гетероскедастичность не влияет на несмещенность оценок параметров, и, при достаточно больших объемах выборки, не приводит к заметной ошибке в их оценке. С другой – учет гетероскедастичности в рамках, скажем, взвешенного метода наименьших квадратов, позволил бы уточнить интервальные оценки средней стоимости, а вместе с этим, повысил бы и уровень объективности в выявлении недооцененных и переоцененных объектов недвижимости.

Список источников

1. Грибовский С.В., Табала Д.Н., Мурашов В.С., Громкова О.Н. Теория и практика массовой оценки недвижимости на примере города Санкт-Петербурга. Теоретические аспекты. // Имущественные отношения в РФ. – 2005. - №7. – С.72-95.
2. Баринов Н.П. Применение регрессионного анализа в задачах индивидуальной и массовой оценки объектов недвижимости // Вопросы оценки. – 2022. - №1. – С.34-46.
3. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. М.: Дело, 2004. 576 с.

4. Беляева А.В. Использование пространственных моделей в массовой оценке стоимости объектов недвижимости // Компьютерные исследования и моделирование. – 2012. – Т.4. - №3. – С.639-650. DOI: 10.20537/2076-7633-2012-4-3-639-650.
5. Горобцова А.Б. Оценка рыночной стоимости квартир с помощью методов регрессионного анализа // Моделирование и анализ данных. – 2019. - №2. – С.63-72.
6. Богданова Т.К., Камалова А.Р., Кравченко Т.К., Полтораки А.И. Проблемы моделирования оценки стоимости жилой недвижимости // Моделирование социальных и экономических систем. – 2020. – Т.14. - №3. – С.7-23. DOI: 10.17323/2587-814X.2020.3.7.23.
7. Нестерова С.И. Моделирование оценки прогнозной стоимости жилой недвижимости на вторичном рынке (на примере г.о. Самара) // Вестник международного института рынка. – 2017. - №1. – С.36-42.
8. Домнина С.В., Савоскина Е.В., Солопова Н.А. Использование регрессионных моделей для анализа и прогнозирования рынка жилой недвижимости // Фундаментальные исследования. – 2024. - №4. – С.36-41. DOI:10.17513/fr.43591.
9. Хлюпина М.А., Исавнин А.Г. Моделирование зависимости и анализ цен на квартиры от ряда факторов на примере города Елабуга // Фундаментальные исследования. – 2016. - №5. – С.213-217.
10. Лейфер Л.А., Черная Е.В. Массовая оценка объектов недвижимости на основе технологии машинного обучения. Анализ точности различных методов на примере определения рыночной стоимости квартир // Имущественные отношения в РФ. – 2020. - №3. – С.32-42.
11. Острикова А.Л., Селютин В.В. Инновационные технологии массовой оценки жилой недвижимости // Системный анализ и моделирование экономических и экологических систем. – 2023. - №8. – С.147-154. DOI: 10.23885/2500-395X-2023-1-8-147-154.

12. Шалагин А.А. Пространственные методы оценки стоимости объектов недвижимости // Финансы и бизнес. – 2023. – Т.19. - №2. – С.59-73. DOI 10.31085/1814-4802-2023-19-2-112-59-73.

13. Ferlan N., Bastic M., Psunder I. Influential factors on the market value of residential properties // Engineering Economics. – 2017. – V.28. – N.2. – P.135-144. DOI:10.5755/j01.ee.28.2.13777.

14. Narula S.C., Wellington J.F., Lewis S.A. Valuating residential real estate using parametric programming // European Journal of Operational Research. – 2012. – V.217. – N.1. – P.120-128. DOI:10.1016/j.ejor.2011.08.014.

References

1. Gribovskij S.V., Tabala D.N., Murashov V.S., Gromkova O.N. Teoriya i praktika massovoj ocenki nedvizhimosti na primere goroda Sankt-Peterburga. Teoreticheskie aspekty`. // Imushhestvenny`e otnosheniya v RF. – 2005. - №7. – S.72-95.

2. Barinov N.P. Primenenie regressionnogo analiza v zadachax individual`noj i massovoj ocenki ob`ektov nedvizhimosti // Voprosy` ocenki. – 2022. - №1. – S.34-46.

3. Magnus Ya.R., Katy`shev P.K., Pereseczkij A.A. E`konometrika. Nachal`ny`j kurs. M.: Delo, 2004. 576 s.

4. Belyaeva A.V. Ispol`zovanie prostranstvenny`x modelej v massovoj ocenke stoimosti ob`ektov nedvizhimosti // Komp`yuterny`e issledovaniya i modelirovanie. – 2012. – Т.4. - №3. – S.639-650. DOI: 10.20537/2076-7633-2012-4-3-639-650.

5. Gorobczova A.B. Ocenka ry`nochnoj stoimosti kvartir s pomoshh`yu metodov regressionnogo analiza // Modelirovanie i analiz danny`x. – 2019. - №2. – S.63-72.

6. Bogdanova T.K., Kamalova A.R., Kravchenko T.K., Poltorak A.I. Problemy` modelirovaniya ocenki stoimosti zhiloy nedvizhimosti // Modelirovanie social`ny`x i e`konomicheskix sistem. – 2020. – Т.14. - №3. – S.7-23. DOI: 10.17323/2587-814X.2020.3.7.23.

7. Nesterova S.I. Modelirovanie ocenki prognoznoj stoimosti zhiloy nedvizhimosti na vtorichnom ry`nke (na primere g.o. Samara) // Vestnik mezhdunarodnogo instituta ry`nka. – 2017. - №1. – S.36-42.
8. Domnina S.V., Savoskina E.V., Solopova N.A. Ispol`zovanie regressionny`x modelej dlya analiza i prognozirovaniya ry`nka zhiloy nedvizhimosti // Fundamental`ny`e issledovaniya. – 2024. - №4. – S.36-41. DOI:10.17513/fr.43591.
9. Xlyupina M.A., Isavnin A.G. Modelirovanie zavisimosti i analiz cen na kvartiry` ot ryada faktorov na primere goroda Elabuga // Fundamental`ny`e issledovaniya. – 2016. - №5. – S.213-217.
10. Lejfer L.A., Chernaya E.V. Massovaya ocenka ob`ektov nedvizhimosti na osnove texnologii mashinnogo obucheniya. Analiz tochnosti razlichny`x metodov na primere opredeleniya ry`nochnoj stoimosti kvartir // Imushhestvenny`e otnosheniya v RF. – 2020. - №3. – S.32-42.
11. Ostrikoval A.L., Selyutin V.V. Innovacionny`e texnologii massovoj ocenki zhiloy nedvizhimosti // Sistemny`j analiz i modelirovanie e`konomicheskix i e`kologicheskix sistem. – 2023. - №8. – S.147-154. DOI: 10.23885/2500-395X-2023-1-8-147-154.
12. Shalagin A.A. Prostranstvenny`e metody` ocenki stoimosti ob`ektov nedvizhimosti // Finansy` i biznes. – 2023. – T.19. - №2. – S.59-73. DOI 10.31085/1814-4802-2023-19-2-112-59-73.
13. Ferlan N., Bastic M., Psunder I. Influential factors on the market value of residential properties // Engineering Economics. – 2017. – V.28. – N.2. – P.135-144. DOI:10.5755/j01.ee.28.2.13777.
14. Narula S.C., Wellington J.F., Lewis S.A. Valuating residential real estate using parametric programming // European Journal of Operational Research. – 2012. – V.217. – N.1. – P.120-128. DOI:10.1016/j.ejor.2011.08.014.