

ПРИМЕНЕНИЕ РЕГРЕССИОННОГО АНАЛИЗА В ЗАДАЧАХ ИНДИВИДУАЛЬНОЙ И МАССОВОЙ ОЦЕНКИ ОБЪЕКТОВ НЕДВИЖИМОСТИ*

Ключевые слова: индивидуальная оценка недвижимости, регрессионный анализ, многофакторная регрессионная модель, нелинейные связи, критерии адекватности регрессионной модели, предпосылки регрессии и последствия их нарушения, распространенные ошибки построения регрессионных моделей

Аннотация

Автором предпринята попытка в краткой форме осветить основные аспекты корректного применения метода регрессионного анализа при оценке недвижимости. Даны определения регрессии и регрессионного анализа, показаны преимущества и особенности метода при оценке недвижимости, различия его применения в индивидуальной и массовой оценке. Сформулированы требования к объектам-аналогам, их минимальному числу и числу ценообразующих факторов, предложен метод выявления нелинейных связей факторов с моделируемой величиной. Описаны критерии оценки качества регрессионной модели, последствия невыполнения предпосылок регрессии, а также распространенные ошибки, допускаемые оценщиками на этапах построения модели.

Регрессия — бабушка контролируемого искусственного интеллекта. Она — старушка, но в этом ее сила: у нее была масса времени, чтобы развить во всех областях своего применения математическую точность.

Джон Форман. «Много цифр. Анализ больших данных при помощи Excel»

В условиях повышения ответственности оценщиков за результаты оценки стоимости объектов становится актуальной задача выбора методов оценки, позволяющих уменьшить субъективную составляющую в суждении оценщика и повысить защищенность результата оценки. На активных рынках одним из таких методов является метод регрессионного анализа. В публикации предпринята попытка в краткой, почти тезисной форме осветить основные аспекты корректного применения метода при оценке недвижимости.

ЧТО ТАКОЕ РЕГРЕССИЯ И РЕГРЕССИОННЫЙ АНАЛИЗ

Регрессия — это зависимость среднего значения какой-либо случайной величины от некоторой другой величины или от нескольких величин¹.

* Материал впервые опубликован в монографии: Современные тенденции рынка и оценка рыночной стоимости: коллективная монография / под ред. Н.Ю. Пузыня. СПб.: Изд-во СПбГЭУ, 2020. 250 с. С. 90–122. URL: <https://elibrary.ru/item.asp?id=44872630>

¹ Математическая энциклопедия: URL: http://enc-dic.com/enc_math/Regressija-3920.html (дата обращения 13.09.2021).

Регрессионный анализ — раздел математической статистики, объединяющий *практические методы* исследования регрессионной зависимости между величинами по статистическим данным.

Регрессионный анализ позволяет решать следующие основные задачи:

- 1) выбирать модель регрессии, т. е. выдвигать предположения о зависимости функций регрессии от влияющих переменных и параметров регрессии;
- 2) оценивать параметры регрессии в выбранной модели методом наименьших квадратов;
- 3) проверять статистические гипотезы о регрессии.

ПОЧЕМУ ОЦЕНКЕ СТОИМОСТИ НУЖЕН РЕГРЕССИОННЫЙ АНАЛИЗ

Многомерный регрессионный анализ (МРА) — сегодня единственный защищаемый количественный метод сравнительного подхода к оценке недвижимости и движимого имущества². Для реализации метода не требуется привлекать данные и экспертные мнения, полученные на других рынках. Метод соответствует пониманию рыночной стоимости как средней по всему рассматриваемому рынку цены на аналогичный актив (объект) [1], а также восприятию результирующего воздействия ценообразующих факторов на стоимость как суммы (или произведения) воздействий каждого фактора в отдельности.

Регрессионный анализ позволяет получать точечную оценку стоимости, а также расчетную оценку интервала, в котором может находиться стоимость (интервала неопределенности).

Метод регрессионного анализа является общепризнанным методом массовой (кадастровой) оценки. Как метод индивидуальной оценки регрессионный анализ в последние годы признан федеральными стандартами оценки РФ (ФСО № 7), требованиями квалификационного экзамена по оценке недвижимости (2017), зарубежной и отечественной практикой оценки, а также судебными системами РФ и США [2].

Главный недостаток метода МРА состоит в невозможности его применения на трех-пяти аналогах, поэтому он используется при оценке на активных рынках. В условиях «ценообразующей» оценки (уникальных объектов, на пассивных рынках) уповать на метод не приходится.

ОСОБЕННОСТИ ПРИМЕНЕНИЯ МЕТОДА РЕГРЕССИОННОГО АНАЛИЗА ПРИ ОЦЕНКЕ НЕДВИЖИМОСТИ

Регрессионный анализ применяется во многих отраслях знания, каждая из которых имеет свои особенности, диктует свои правила. При построении регрессионных моделей индивидуальной и массовой оценки можно выделить общие черты, отличающие проведение регрессионного анализа при оценке недвижимости:

– строятся пространственные модели, которые описывают влияние территориальных факторов (location, location & location), относительно стабильных во времени;

– фактор времени включается в регрессионные модели скорее как исключение, что во многом упрощает построение моделей. Изменение цен во времени согласно известной динамике рынка учитывается вне модели приведением цен к одной дате оценки;

– имеются «качественные» факторы, подлежащие упорядочению и оцифровке для учета в модели. При этом должна быть обеспечена оцифровка в шкале отношений, т. е. с сохранением количественных соотношений между уровнями влияния каждого из факторов на моделируемую величину (стоимость, рыночную арендную плату);

– преобладают нелинейные зависимости моделируемой величины от влияющих факторов. Необходимо выявлять вид нелинейной зависимости от каждого из существенно влияющих факторов.

РАЗЛИЧИЯ В ПРИМЕНЕНИИ МЕТОДА В ИНДИВИДУАЛЬНОЙ И МАССОВОЙ ОЦЕНКЕ

Индивидуальная оценка:

– моделируется ценообразование на малых генеральных совокупностях (локальных территориях «вокруг» объекта оценки), при этом малые выборки составляют относительно большие доли генеральной совокупности аналогов (большие доли отбора). Эта ситуация нехарактерна для «классической» статистики и требует учета своей специфики;

² Модель среднего значения рассматривается здесь как вырожденный (в отсутствие различий по всем влияющим факторам) случай регрессионной модели.

– для каждой оценки (одного или нескольких схожих объектов) строится своя регрессионная модель, которая используется, как правило, однократно по принципу «здесь и сейчас».

Массовая (кадастровая) оценка:

– модели строятся для оценки множества объектов одного сегмента и используются в течение продолжительного периода (2–5 лет) по принципу «везде и надолго»;

– генеральные совокупности объектов оценки, как и выборки аналогов в каждом сегменте, достаточно многочисленны, что приближает задачи оценки к задачам «классической» статистики;

– как правило, моделируемые территории обширны, поэтому территорию оцениваемого рынка часто разбивают на «однородные» части с построением отдельных моделей. Единая регрессионная модель в таких случаях является составной («матрешкой»);

– в подавляющем большинстве случаев информация о характеристиках объектов-аналогов является неполной по сравнению с необходимой для оценки и требует доработки для введения ее в модель.

ТРЕБОВАНИЯ К ЧИСЛУ ЦЕНООБРАЗУЮЩИХ ФАКТОРОВ

В регрессионной модели должны быть учтены *все ценообразующие факторы*, по которым различаются объекты сравнения и вариация которых способна объяснить наблюдаемую вариацию цен.

Речь идет о реальных ценообразующих факторах, проявляющихся на рынке. Необходимо помнить о явлении ложной корреляции и не использовать факторы, которые, возможно, и могут влиять на цены, но не проявляются на рассматриваемом рынке. Также нельзя рассматривать в качестве реальных ценообразующих факторов регрессионной модели величины, построенные на основе экспертных оценок, которые невозможно подтвердить рыночными данными³.

Пропуск (невключение в модель) существенно влияющего фактора — *грубая ошибка* спецификации модели, приводящая к *смещению оценок* модели. Такие модели можно встретить у оценщиков, «соблазненных» простотой модели, включающей два-три основных фактора, с хорошими отдельными показателями (R^2 , погрешности аппроксимации).

При построении моделей учет «лишнего» слабо влияющего фактора — гораздо меньшее зло⁴. Однако таких «лишних» факторов в модели должно быть немного. На практике «для гарантии» несмещенности получаемой оценки стоимости можно оставлять в модели один-два фактора, не вполне удовлетворяющие статистическим тестам⁵.

Значения учитываемых ценообразующих факторов должны быть представлены для всех аналогов (так называемая полная матрица свойств). В реальной жизни выполнить это требование часто сложнее, чем найти нужное число аналогов. И если для какого-то аналога нет сведений о значении хотя бы одного из включенных в модель факторов, необходимо дополнить эти сведения косвенным путем либо отказаться от включения аналога в выборку.

КАКИЕ АНАЛОГИ НУЖНЫ ДЛЯ ПОСТРОЕНИЯ РЕГРЕССИОННОЙ МОДЕЛИ?

Отвечая кратко — лишь те, которые образуют однородную выборку. В свою очередь, однородной для регрессионного анализа считается выборка, все элементы которой, включая оцениваемый объект, принадлежат одному сегменту рынка.

Отметим, помня определение аналога в п. 22б ФСО № 7⁶, что все объекты сравнения должны характеризоваться единым набором ценообразующих факторов и при этом влияние каждого из них на зависимую величину (стоимость, арендную плату) должно быть единообразным. Это определение опирается на логичное и ясное положение — одной моделью можно корректно описать лишь одну зависимость от каждого фактора. Поэтому включение в модель объектов из разных сегментов существенно затрудняет либо делает невозможным построение корректной модели (рис. 1).

Выход здесь один — необходимо формировать однородную выборку, содержащую объекты одного сегмента рынка. Подчеркнем — для корректного построения регрессионной модели не существует «офисно-торговой» недвижимости, помещений «свободного назначения» — каждый объект должен быть отнесен по своему наилучшему использованию к соответствующему сегменту рынка (офисы либо торговля в первом случае, торговля, услуги либо офис или склад — во втором).

³ В таких случаях проведенный «регрессионный анализ», как правило, исполняет роль «ширмы объективности», прикрывающей субъективное мнение экспертов или ошибки, присущие экспертным методам.

⁴ Если зло вообще, так как смещения оценок нет, растет лишь неопределенность оценки, которая контролируется расчетом доверительного интервала.

⁵ Невысокая значимость фактора по статистическим тестам может быть обусловлена, в том числе, малым объемом выборки и/или отличным от нормального распределением остатков модели.

⁶ В этом пункте дано гораздо более точное определение, нежели общее для всех направлений оценки определение аналога в п. 10 ФСО № 1. Тем не менее оно часто игнорируется оценщиками недвижимости.

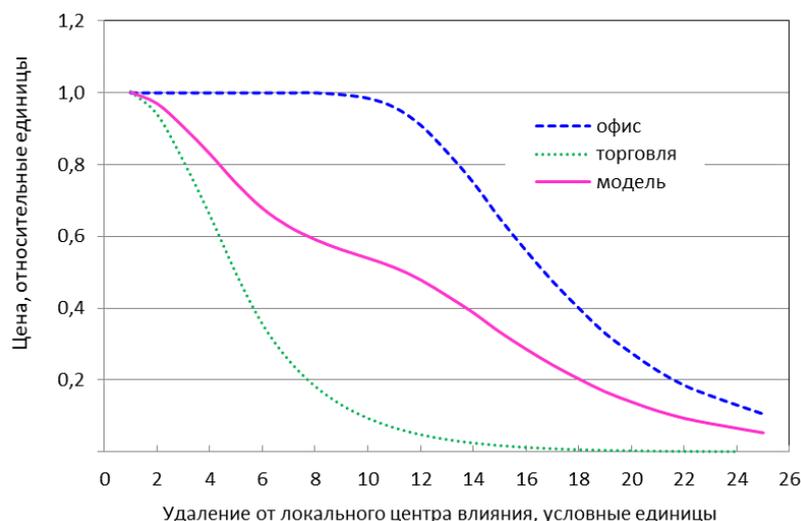


Рис. 1. Отклик модели при установлении зависимости по аналогам из разных сегментов рынка

В этой связи уместно упомянуть об одной из проблем кадастровой оценки. Принятая в ней сегодня группировка объектов оценки по видам разрешенного использования часто не совпадает с разделением рынка на сегменты в смысле определения п. 226 ФСО № 7. Так, в пределах соответствующих групп можно выделить объекты разных сегментов, ценообразование которых существенно различается: ритейл-торговля и масштабные торговые объекты форматов «макси», «гипер», «мега»; складские объекты на производстве, логистические комплексы и т. д. Выходом здесь может служить построение нескольких моделей в пределах одной группы — своей для каждого сегмента рынка.

В индивидуальной оценке при формировании выборки аналогов для расчетов стремятся, по возможности, отбирать аналоги «хуже» и «лучше» объекта оценки по ценообразующим свойствам. Это уменьшает неопределенность оценки⁷ и необходимость введения поправок вне модели, когда параметры объекта оценки выходят за границы интервалов варьирования соответствующих параметров объектов-аналогов.

СКОЛЬКО НУЖНО АНАЛОГОВ?

Краткий ответ — ровно столько, сколько нужно для построения адекватной модели. Аналогов (в смысле п. 226 ФСО № 7) не бывает много — чем больше аналогов, тем лучше проявляются связи и тем меньше неопределенность оценки стоимости. Но в практике оценки недвижимости число аналогов и их близость объекту оценки и друг другу — обменные показатели. Чем больше аналогов, тем по большему числу факторов они могут отличаться друг от друга и объекта оценки. В этом смысле корректно обсуждать требования лишь к минимальному числу аналогов, необходимому для построения регрессионной модели.

При определенных наборах аналогов адекватные модели могут быть построены при числе аналогов $n \geq 2(k + 1)$, где k — число учитываемых в модели ценообразующих факторов [3, 4]. Поэтому при часто встречающихся выборках аналогов, различающихся по четырем-пяти факторам, корректные модели могут быть построены на 10–12 аналогах.

Разность между числом используемых аналогов и числом рассчитываемых моделью коэффициентов принято называть числом степеней свободы — ЧСС = $n - k - 1$. Выражая число аналогов через ЧСС: $n = \text{ЧСС} + (k + 1)$, получаем неравенство $\text{ЧСС} \geq k + 1$, определяющее минимально требуемое число степеней свободы. То есть аналогов нужно столько, чтобы на каждый фактор приходилось не менее одной степени свободы с небольшим запасом.

В практике индивидуальной оценки аналоги ищут начиная с «точно таких же» или «очень похожих». Считают факторы, по которым различаются объекты сравнения, и добавляют аналоги до тех пор, пока не обеспечивается требуемое число степеней свободы. Если минимальные требования удовлетворить не удастся — не стоит и начинать строить модель.

В массовой оценке собирают все доступные аналоги. Их, как правило, не хватает, поскольку объектов оценки, для которых применяется модель стоимости, значительно больше числа возможных аналогов.

⁷ См. далее расчет доверительного интервала.

КАК «ЭКОНОМИТЬ» ТРЕБУЕМЫЕ СТЕПЕНИ СВОБОДЫ ПРИ ИНДИВИДУАЛЬНОЙ ОЦЕНКЕ НЕДВИЖИМОСТИ?

Очевидный ответ — отобрать аналоги, «такие же, как объект», т. е. отличающиеся от него по минимальному числу факторов. Но на неактивных рынках любых аналогов не хватает, не говоря о близко схожих.

Можно рекомендовать несколько практических приемов:

– использовать табличное представление функций влияния факторов на моделируемую величину вместо аналитического их представления;

– минимизировать применение бинарных переменных как самостоятельных факторов;

– применять в качестве ценообразующих факторов «обобщенные» показатели вместо нескольких «элементарных» (например, площадь, объем вместо длины, ширины, высоты);

– рассматривать возможность разумного объединения нескольких факторов в градации одного. Например, вместо двух факторов «удаленность от трассы» и «вид покрытия подъездных путей» использовать фактор «условия подъезда» с расширенным числом градаций, ранжированных по степени влияния. Прием следует применять с осторожностью, соблюдая здравый смысл и обеспечивая экономическое обоснование выбранных градаций обобщенного фактора;

– при отборе аналогов следить за тем, чтобы рост числа аналогов опережал рост числа факторов, по которым они различаются. Если этого не наблюдается, дальнейшее увеличение числа аналогов теряет смысл.

ВЫБОР ЕДИНИЦЫ СРАВНЕНИЯ

Общая рекомендация состоит в том, что в качестве единицы сравнения следует выбирать удельные показатели (цена/м², цена/м³, цена/машиноместо, и др.). Этому есть несколько причин:

– участники рынка недвижимости используют именно такие показатели для сравнения объектов;

– переход к удельным показателям часто уменьшает гетероскедастичность (неравномерность дисперсии) остатков модели, которая искажает интервальные оценки;

– при использовании абсолютных значений цен сильная «положительная» зависимость цены от фактора масштаба (площади, объема, числа машиномест и др.) затрудняет, маскирует выявление зависимости (как правило, «отрицательной») удельного показателя цены от этого же фактора масштаба;

– иногда модель в «абсолютных» ценах не строится, зависимости от факторов не видны, однако с переходом к удельным показателям они проявляются.

НЕЛИНЕЙНЫЕ СВЯЗИ НА РЫНКЕ

Большинство зависимостей стоимости от влияющих факторов на рынке — нелинейные [5]. Более того, часть из них — немонотонные (рис. 2).

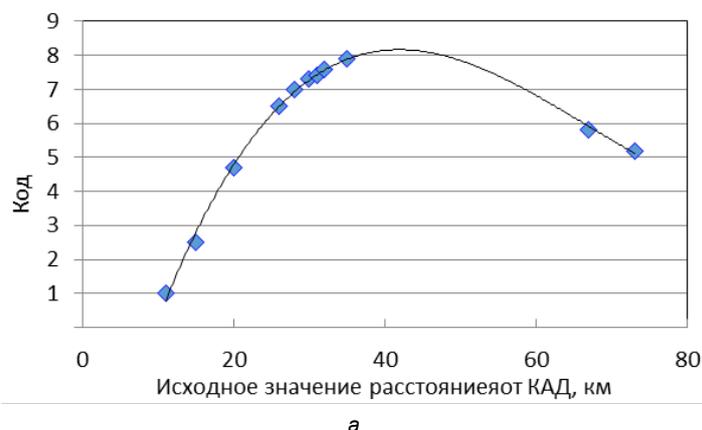
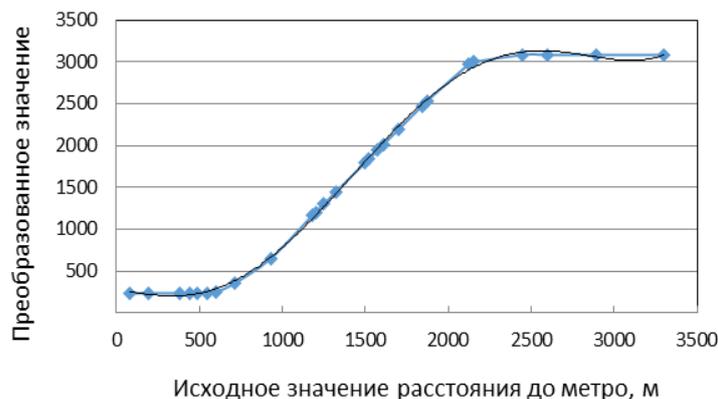
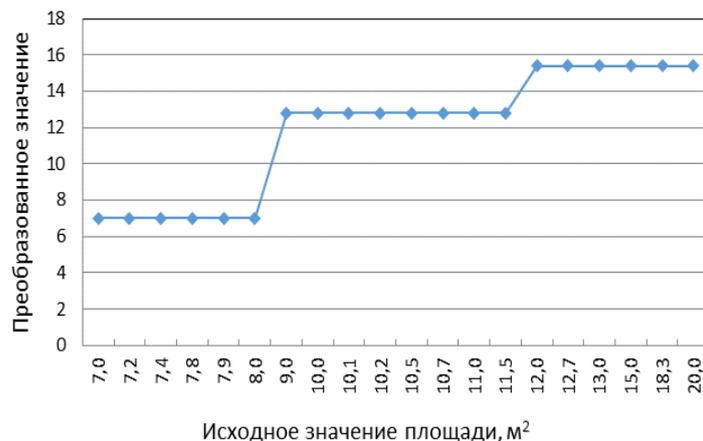


Рис. 2 (начало). Примеры нелинейных зависимостей стоимости недвижимости от отдельных ценообразующих факторов:

а — расстояние до КАД



б



в

Рис. 2 (окончание). Примеры нелинейных зависимостей стоимости недвижимости от отдельных ценообразующих факторов: б — расстояние до метро; в — площадь кухни

КАК УЧЕСТЬ НЕЛИНЕЙНЫЕ СВЯЗИ

Известно, что теория регрессионного анализа наиболее полно разработана для аддитивных моделей, линейных относительно рассчитываемых коэффициентов регрессии. Такие модели часто называют линейными, а при наличии линейных же зависимостей от влияющих факторов — собственно линейными или сугубо линейными. Но как учесть преобладающие в реальности нелинейные зависимости?

Рекомендации, которые можно найти в литературе: строить нелинейные регрессионные модели с аналитическим представлением зависимостей. Как правило, это модели мультипликативного вида, приводимые к аддитивному виду логарифмированием, и к линейному виду — заменой переменных. Недостатки реализации таких рекомендаций заключаются в однотипности вида нелинейной зависимости для всех факторов или большинства из них, сложности подбора вида аналитических зависимостей.

В определенной степени эти недостатки купируются использованием семейства параметрических кривых, описываемых одним аналитическим выражением, но различающихся по виду при изменении значений влияющего параметра. Например, в моделях кадастровой оценки недвижимости Санкт-Петербурга [6] используются семейства вида

$$f(x) = \exp \left[- \left(\frac{x}{R} \right)^N \right].$$

Подбирая значение параметра N для каждого учитываемого в модели фактора, добиваются максимальной близости зависимости от расстояния рыночным данным. Преимуществом такого приема является получение гладких зависимостей, соответствующих зависимостям стоимости, а недостатком — сложность реализации немонотонных зависимостей, когда в этом возникает необходимость (рис. 3).

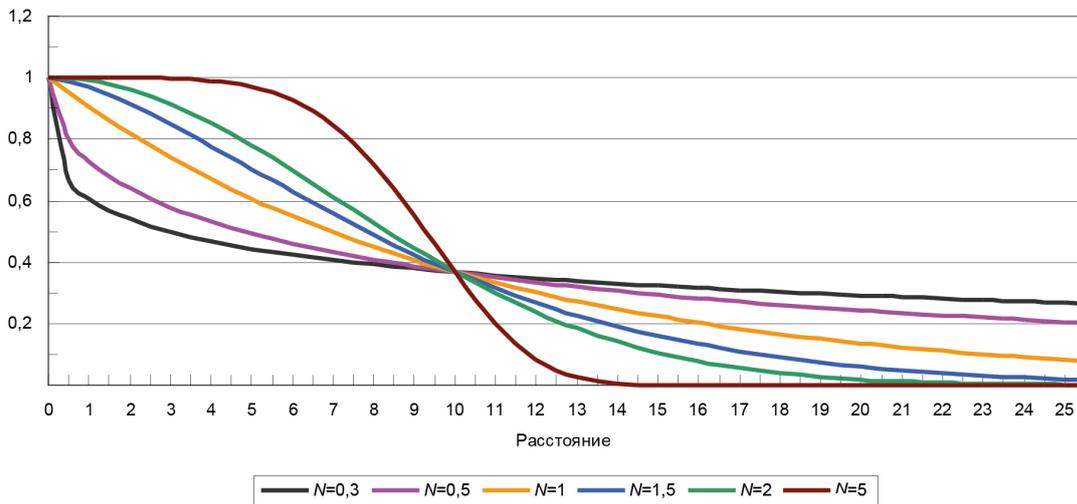


Рис. 3. Графики параметрического семейства кривых, изменяющих вид в зависимости от параметра N

В индивидуальной оценке существует доступная альтернатива известным приемам учета нелинейных связей в линейных регрессионных моделях.

Доказано [7], что для зависимостей, заданных дискретно (таблично), всегда существует преобразование координат (значений влияющего фактора), линеаризующее связь зависимой переменной модели с фактором (рис. 4). Иными словами, в индивидуальной оценке (с конечным и относительно небольшим числом аналогов и значений ценообразующих факторов) всегда найдется такая замена переменных, которая приведет модель с нелинейными связями к модели с линейными зависимостями от новых значений этих же факторов.

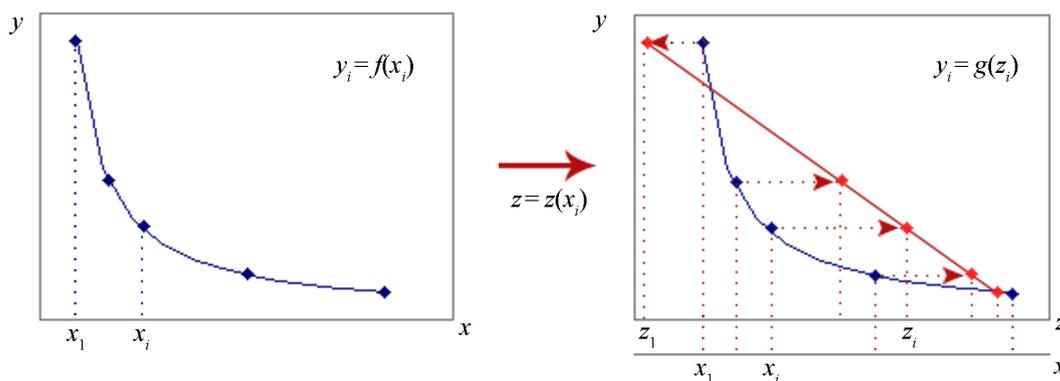


Рис. 4. Графическое представление преобразования $z = z(x)$ шкалы ценообразующего фактора x для линеаризации зависимости $y = g(z)$ от преобразованных значений z фактора x

На практике преобразование каждой из переменных может быть найдено, в том числе, совместным применением инструментов MS Excel: функции «ЛИНЕЙН», табличной формой задания функций и макроса «Поиск решения». Оптимизация проводится по минимуму остаточной суммы или по максимуму коэффициента детерминации (R^2), что эквивалентно [8]. При этом заданием порядка следования цифровых меток влияющего фактора в макросе «Поиск решения» обеспечивается монотонность каждой зависимости или ее отдельных участков. Процесс ускоряется с помощью функции «ВПР». Можно преобразовывать последовательно каждую из зависимостей, но лучше (и быстрее) — для всех переменных одновременно [8].

В процессе проверяют вид получаемых зависимостей на соответствие экономическим гипотезам, отражающим ценообразование на рынке. Для этого строят графики зависимости преобразованных значений фактора от исходных либо зависимости моделируемой величины (средней удельной цены) от исходных значений фактора⁸.

⁸ Эти зависимости подобны с точностью до знака, так как стоимость линейно зависит от преобразованных значений.

При выявлении нелинейности с помощью макроса «Поиск решения» его инструментами удается задать лишь условие монотонности зависимости. Поэтому получаемые кривые чаще всего не являются гладкими в силу наличия случайной составляющей в ценах аналогов. Поскольку выявлению подлежит зависимость не цены, а стоимости от ценообразующего фактора, отклонения точек кривой относительно гладкой кривой тренда нельзя признать соответствующими представлениям о ценообразовании на рынке. Потому полученные кривые «сглаживаются» вручную изменением преобразованных значений фактора (рис. 5).

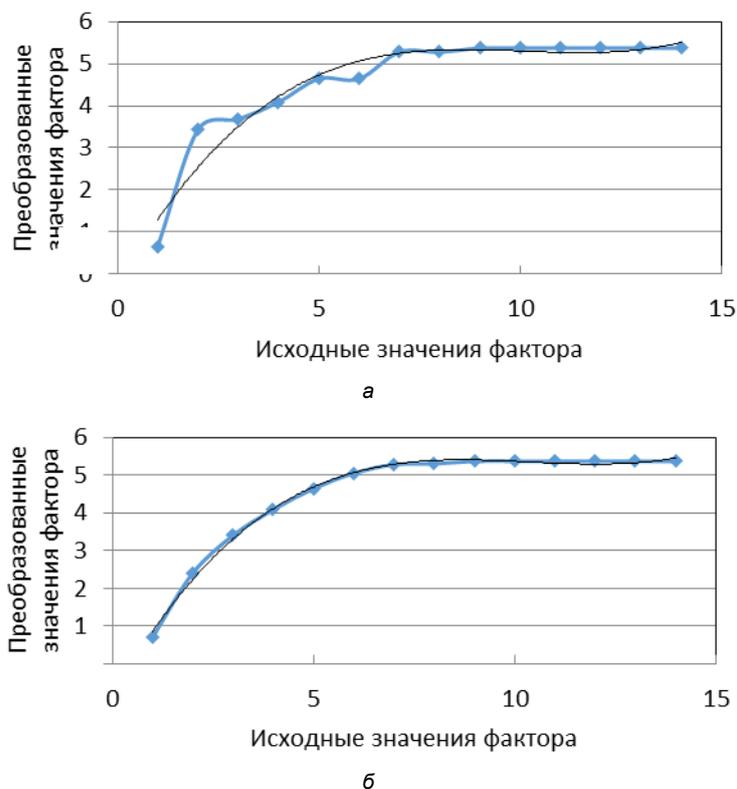


Рис. 5. Пример выявленного с помощью макроса «Поиск решения» нелинейного преобразования значений $X_{\text{преобр}} = f(X_{\text{исходн}})$ влияющего фактора: а — до сглаживания; б — после сглаживания «вручную»

Заметим напоследок, что умение выявлять реальные нелинейные зависимости в огромной степени определяет успех построения корректных регрессионных моделей. Представляется, что именно отсутствие такого умения сдерживает применение многомерного регрессионного анализа в оценке.

АНАЛИЗ ОСТАТКОВ МОДЕЛИ

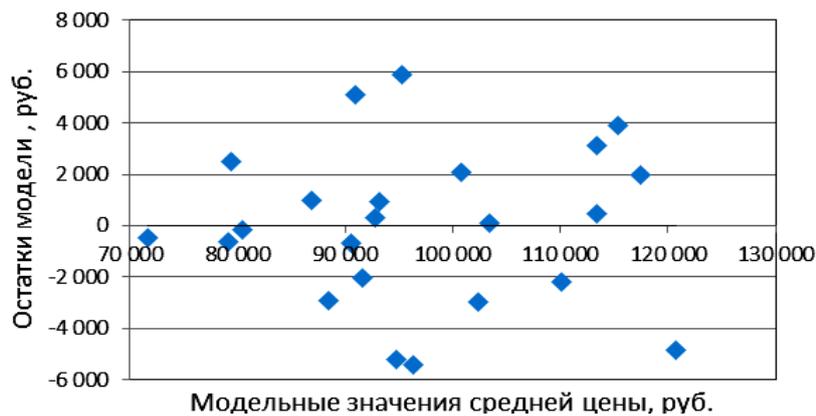
«Вы будете видеть в этих остатках потенциально неограниченный источник для зарождения новых идей, а возможно, и новых теорий. Остатки, взятые по отдельности, указывают, когда, где и в какой степени существующая модель не смогла объяснить наблюдаемые события» [9].

Анализ остатков модели является важным элементом проверки качества построения модели и позволяет выявить ряд отклонений от предпосылок регрессии. В качестве иллюстрации приведены диаграммы остатков двух моделей — адекватной и с пропущенным ценообразующим фактором (рис. 6).

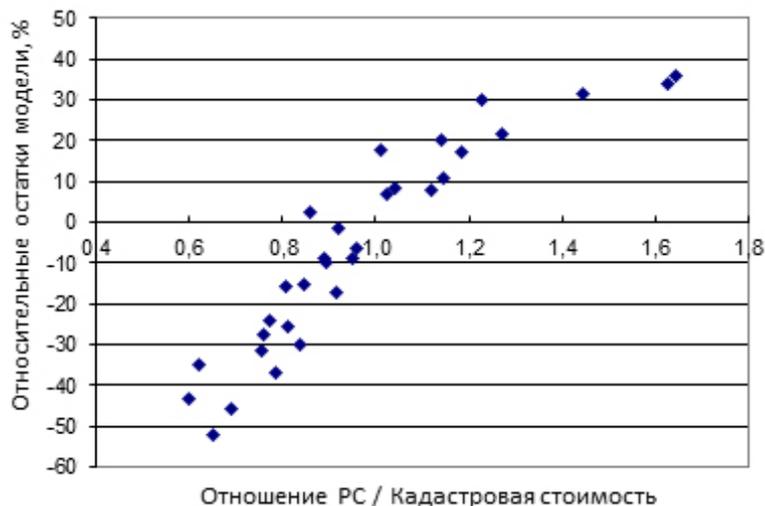
КРИТЕРИИ ОЦЕНКИ АДЕКВАТНОСТИ (КОРРЕКТНОСТИ) МОДЕЛИ

Единственного критерия адекватности (корректности) регрессионной модели, к сожалению, не существует. Можно говорить о группе (комплексе) показателей, обеспечивающих адекватность модели рыночным данным и закономерностям:

- соответствие знаков коэффициентов модели экономическим гипотезам;
- соответствие вида функции влияния каждого из факторов на зависимую переменную экономическим гипотезам;



а



б

Рис. 6. Диаграммы распределения остатков регрессионной модели от прогнозных значений зависимой переменной:
а — адекватная модель; б — модель с пропущенным влияющим фактором

- статистическая значимость большинства коэффициентов модели;
- случайность распределения остатков модели от модельных значений зависимой переменной;
- средняя и максимальная ошибки аппроксимации;
- (доверительный) интервал неопределенности для точечной оценки;
- устойчивость (балансировка) регрессионной модели к присутствию/удалению отдельных аналогов и чувствительность моделируемой величины к изменению факторов;
- для массовой оценки дополнительно — корректная работа модели за пределами интервала изменения входных параметров по выборке аналогов.

На практике, как правило, строят несколько моделей, и на последних этапах отладки лучшую из них выбирают не столько по формальным статистическим показателям, сколько по соображениям соответствия рынку.

РАСЧЕТ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ СРЕДНЕГО ЗНАЧЕНИЯ ЦЕНЫ (ОЦЕНКИ СТОИМОСТИ)

Внутренним свойством регрессионной модели является возможность расчетной оценки доверительного интервала для оценки среднего значения моделируемой величины. Для корректных моделей оценки недвижимости доверительный интервал можно рассматривать как интервал, в котором может находиться значение оцениваемой стоимости.

Расчету интервала для методов сравнительного подхода посвящена публикация [10], к которой мы отсылаем заинтересованного читателя. Здесь ограничимся основными формулами.

Однофакторная модель:

$$\tilde{y} = a_0 + a_1 f(z) \rightarrow \tilde{y} = b_0 + b_1 x.$$

Доверительный интервал для оценки стоимости:

$$\tilde{y}_0 \pm t_{\alpha} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

Многофакторная модель:

$$\tilde{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k.$$

Решение многофакторного уравнения в матричном виде:

$$B = (X^T X)^{-1} X^T Y.$$

Доверительный интервал оценки стоимости в матричном виде:

$$\tilde{y}_0 \pm t_{\alpha} s \sqrt{x_0^T (X^T X)^{-1} x_0}.$$

Подробно процедура расчета изложена в упомянутой выше публикации, примеры реализации в среде электронных таблиц MS Excel для реальных моделей оценки недвижимости можно найти на ветках дискуссий интернет-портала Appraiser.ru⁹.

Приведенные выше «классические» формулы для расчета интервалов получены для неограниченно больших генеральных совокупностей, малых долей отбора и допущения о нормальном распределении остатков модели. Однако результаты проведенного моделирования [11] методом статистических испытаний (Монте-Карло) показали, что в реальных условиях оценки недвижимости (малые генеральные совокупности, большие доли отбора, отсутствие нормального распределения остатков) «классические» соотношения дают оценки интервалов с некоторым запасом, т. е. являются оценками сверху для его величины. Это позволяет получать на практике оценки интервала с несколько большей уверенностью, чем с объявленной доверительной вероятностью.

О ПРЕДПОСЫЛКАХ РЕГРЕССИИ (УСЛОВИЯ ГАУССА — МАРКОВА)

Классический регрессионный анализ базируется на следующих предположениях [12].

1. На вектор неизвестных параметров (коэффициентов) регрессии не наложено никаких ограничений.
2. Вектор случайных отклонений (остатков) ε — случайный. Отсюда следует, что и вектор зависимой переменной Y — также случайный.
3. Математическое ожидание отклонений равно нулю: $E(\varepsilon_i) = 0$. Это условие выполняется автоматически в уравнении регрессии с постоянным членом.
4. Для любых $i \neq j$ $E(\varepsilon_i \varepsilon_j) = 0$ и $E(\varepsilon_i^2) = \sigma^2$, т. е. остатки независимы между собой и имеют постоянную дисперсию (гомоскедастичны).
5. Матрица влияющих переменных X детерминирована, т. е. x_i не являются случайными переменными. Предполагается, что уровни влияющих факторов известны с необходимой точностью.
6. $\text{rank}(X) = m$, где m — число влияющих переменных, т. е. все влияющие переменные линейно независимы между собой.

Некоторые источники включают в число предпосылок регрессии предположение о нормальном распределении остатков и зависимой величины. Однако это удобное свойство не является необходимым условием для построения адекватных регрессионных моделей: «Требование нормальности распределения (зависимой величины) Y необходимо лишь для проверки значимости уравнения и его параметров, а также для интервального оценивания. Для получения точечных оценок этого условия не требуется [13].

Не требуется проверка нормальности распределения значений зависимой переменной (удельной цены, арендной ставки) еще и потому, что в условиях нелинейных связей даже при нормальных остатках модели распределение зависимой переменной не может быть нормальным.

Нарушения предпосылок регрессии и условий построения регрессионных моделей, а также последствия таких нарушений сведены в таблицу.

⁹ URL: <http://www.appraiser.ru/default.aspx?SectionId=32&g=posts&t=5828&p=2> (сообщение Н.П. Баринаова (NB) от 01.02.2013), а также <http://www.appraiser.ru/default.aspx?SectionId=32&g=posts&t=5350&p=16> (сообщение Н.П. Баринаова (NB) от 01.02.2010).

Предпосылки регрессии и последствия их нарушений

Нарушение условия построения модели либо предпосылки регрессии	Негативные последствия, возможные действия для устранения эффекта
Объем выборки не превышает числа варьируемых переменных $n \leq k$ либо между переменными существует строгая линейная зависимость	<i>Невозможно вычислить коэффициенты уравнения регрессии.</i> Необходимо увеличить объем выборки, обеспечить ее однородность, а также исключить из рассмотрения строго зависимые переменные
Малый объем (и малая доля отбора) выборки	<i>Снижение эффективности, увеличение дисперсии и интервалов неопределенности оценок. Затруднено применение статистических критериев.</i> Необходимо увеличивать объем выборки
Выборка неоднородна (включены объекты из другого сегмента рынка, не подчиняющиеся закономерностям основной совокупности)	<i>Регрессионные оценки коэффициентов смещены. Также смещена оценка моделируемой величины. Модель неадекватна.</i> Необходимо удалять из выборки наблюдения других совокупностей
Неправильная спецификация модели — недобор влияющих переменных	<i>Смещение и несостоятельность оценок. Модель неадекватна.</i> Необходимо пересмотреть спецификацию модели
Неправильная спецификация модели — перебор влияющих переменных	<i>Увеличение дисперсии оценок.</i> Оценки коэффициентов регрессии и зависимой величины не смещены
Сопряженность (мультиколлинеарность) влияющих переменных	<i>Снижение эффективности оценок. Оценки коэффициентов регрессии смещены, оценка зависимой величины (уравнения в целом) остается несмещенной. Затруднена интерпретация модели в предметной области — оценка вклада каждого из факторов</i>
Значения остатков неслучайны относительно зависимой переменной	Значения зависимой переменной также неслучайны, <i>оценки смещены, модель неадекватна.</i> Необходимо пересмотреть спецификацию модели — вид функциональной связи и/или состав влияющих факторных переменных
Значения остатков зависят от влияющих переменных (для пространственных моделей — нетипично)	<i>Регрессионные оценки смещены, модель неадекватна.</i> Необходимо пересмотреть спецификацию модели — вид функциональной связи и/или состав влияющих факторных переменных
Гетероскедастичность (различная дисперсия) остатков	Коэффициенты регрессии остаются несмещенными, <i>но становятся неэффективными. Затруднена проверка значимости коэффициентов по <i>t</i>-критерию, а также оценка доверительных интервалов моделируемой величины.</i> В качестве первого шага рекомендуется переход к удельным показателям
Отсутствие нормальности распределения остатков модели	<i>Затруднено построение интервальных оценок, а также проверка значимости коэффициентов регрессии.</i> При симметричных генеральных совокупностях могут быть получены корректные оценки сверху для интервальных оценок в отсутствие нормальности распределения остатков
Наличие автокорреляции остатков (как правило, для временных рядов). Для пространственных моделей нетипично	Коэффициенты регрессии остаются несмещенными, <i>но становятся неэффективными.</i> Рекомендуется пересмотреть спецификацию модели

РАСПРОСТРАНЕННЫЕ ОШИБКИ ПОСТРОЕНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ

На этапе отбора аналогов:

– использование объектов, принадлежащих разным сегментам (субсегментам) рынка, в том числе включение «разномасштабных» аналогов;

– построение модели по выборке с объемом, меньшим минимально необходимого.

При определении состава ценообразующих факторов:

– отсутствие анализа ценообразующих свойств, по которым различаются объекты сравнения;

– предварительный отсев факторов по результатам построения корреляционной матрицы связей;

– невключение фактора, указанного в анализе рынка как влияющего, по которому отобранные объекты сравнения различаются;

– включение в состав факторов субъективных показателей (развитость социальной структуры, оптимальность местоположения и др.), ранжирование которых не поддается объективному контролю;

– включение в состав факторов модели свойств объектов, влияние которых не подтверждено рынком;

– удаление факторов из модели по статистическим тестам без учета рыночных закономерностей и условий несмещенности оценок;

– сохранение в модели факторов с алогичными (не соответствующими экономическим гипотезам) знаками коэффициентов.

При оцифровке уровней ценообразующих факторов:

– использование числа градаций фактора, превышающего возможность надежного разнесения по ним объектов сравнения;

- использование градаций, не соответствующих ценообразованию на рынке;
- при линеаризации модели изменение порядка следования цифровых меток, соответствующих преобразованным уровням факторов, относительно первоначального порядка, заданного согласно экономическим гипотезам.

При определении функций влияния факторов:

- игнорирование нелинейности зависимостей, необоснованное использование линейных функций;
- произвольный (без обоснования) выбор вида нелинейной функции влияния фактора из числа простейших;
- использование негладких связей между зависимой переменной и фактором без надлежащего обоснования;
- отсутствие графиков влияния факторов на зависимую переменную (либо преобразованных значений каждого фактора от исходных значений этого фактора);
- выбор по наибольшему значению R^2 лучшей модели из нескольких простейших нелинейных моделей.

При оценке корректности регрессионной модели:

- использование критерия R^2 (как самостоятельного либо в совокупности с F -критерием) в качестве доказательства адекватности (корректности) полученной модели;
- отсутствие статистических оценок значимости факторов, использованных в модели;
- отсутствие контроля адекватности знаков коэффициентов регрессии;
- отсутствие графиков влияния факторов на зависимую величину (связи преобразованных и исходных значений факторов);
- отсутствие оценок погрешностей аппроксимации модели;
- отсутствие анализа распределения остатков модели.

ВЫВОДЫ

Реальных задач при построении регрессионных моделей индивидуальной оценки недвижимости шесть.

1. Обеспечение однородности выборки аналогов. На практике объекты должны отбираться из одного сегмента (субсегмента) рынка с единообразным ценообразованием.

2. Отбор требуемого числа аналогов (числа степеней свободы). Как минимум, должно быть обеспечено число аналогов, с небольшим запасом, вдвое превышающим число факторов модели (т. е. чтобы число степеней свободы превышало число учитываемых факторов на несколько единиц).

3. Недопущение пропуска (неучета) существенно влияющих факторов. Реальный прием — сохранение в модели одного-двух факторов, значимость которых недостаточна по статистическим тестам.

4. Корректная спецификация модели в части учета реальных нелинейных связей факторов с моделируемой величиной. Практический прием — использование оптимизационных процедур для осуществления преобразования влияющих переменных с табличным заданием функций и контролем вида полученных зависимостей по графикам.

5. Контроль случайности остатков модели от значений моделируемой зависимой переменной (и влияющих факторов). Реальный прием — построение точечных диаграмм остатков модели от модельных значений зависимой переменной. Для пространственных моделей оценки недвижимости зависимость остатков от влияющих факторов нетипична.

6. Финишный контроль чувствительности модели к изменению уровней факторов. Для моделей массовой оценки дополнительно — проверка адекватности модели по всему множеству объектов оценки, охватываемых моделью. Контроль позволяет на практике проверять соответствие полученной модели рыночным соотношениям.

ЗАКЛЮЧЕНИЕ

Все перечисленные задачи регрессионного анализа при оценке недвижимости можно успешно решать на активных рынках. Неиспользование метода в этих условиях методически необъяснимо.

Представляется, что растущая потребность в оценках стоимости, способных выдержать защиту в спорах, в том числе судебных, неизбежно приведет к расширению применения метода многомерного регрессионного анализа при индивидуальной оценке недвижимости на рынках, способных предоставить необходимое число аналогов для построения моделей.

Литература

1. *Баринов Н.П.* Сравнительный подход к оценке недвижимости. Современный взгляд // Вопросы оценки. 2019. № 1 (95). С. 2–16.
2. *Колвелл П.Ф., Хеллер Д.А., Трефзгер Д.В.* Экспертные свидетельства: регрессионный анализ и другие систематические методы // Вопросы оценки. 2014. № 3 (77). С. 11–21.
3. *Анисимова И.Н., Баринов Н.П., Грибовский С.В.* О требованиях к количеству сопоставимых объектов при оценке недвижимости сравнительным подходом // Вопросы оценки. 2003. № 1. С. 2–7.
4. *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.
5. *Баринов Н.П.* Оценка рыночной стоимости земельного участка методом многомерного регрессионного анализа / Информационно-аналитический бюллетень рынка недвижимости RWAY. 2014. № 232 (июль), № 236 (ноябрь). URL: <http://www.appraiser.ru/default.aspx?SectionID=35&Id=3821> (дата обращения 12.04.2022).
6. *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 (46). С. 60–72.
7. *Зельдин М.А., Баринов Н.П., Ситников Н.А.* Линеаризация нелинейных связей в регрессионной модели или еще раз об оцифровке влияющих переменных // Матер. IV Поволжской науч.-практ. конф. «Статистические методы массовой и индивидуальной оценки». Нижний Новгород, 31 марта — 1 апреля 2011 г. URL: <http://www.appraiser.ru/default.aspx?SectionId=35&Id=3605> (дата обращения 12.04.2022).
8. *Анисимова И.Н., Баринов Н.П., Грибовский С.В.* Учет разнотипных ценообразующих факторов в многомерных регрессионных моделях оценки недвижимости // Вопросы оценки. 2004. № 2. С. 2–15.
9. *Дугерти К.* Введение в эконометрику / пер. с англ. М.: ИНФРА-М, 2009. 465 с.
10. *Баринов Н.П.* Расчет интервала неопределенности оценки стоимости методами сравнительного подхода // Вопросы оценки. 2019. № 4. С. 2–10
11. *Зельдин М.А., Баринов Н.П., Аббасов М.Э.* Неопределенность оценки рыночной стоимости, получаемой по модели множественной регрессии // Бюллетень рынка недвижимости RWAY. 2013. № 221 (август). URL: <http://www.appraiser.ru/default.aspx?SectionId=35&Id=3750> (дата обращения 12.04.2022).
12. *Демиденко Е.З.* Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981. 303 с.
13. *Дубров А.М., Мхитарян В.С., Трошин Л.И.* Многомерные статистические методы. М.: Финансы и статистика, 2003. 352 с.

Баринов Николай Петрович, e-mail: n.barinov@bk.ru

Статья поступила в редакцию 06.05.2022