

АВТОМАТИЗИРОВАННОЕ ВЫЯВЛЕНИЕ НЕЛИНЕЙНЫХ СВЯЗЕЙ ПРИ ПОСТРОЕНИИ РЕГРЕССИОННЫХ МОДЕЛЕЙ ИНДИВИДУАЛЬНОЙ ОЦЕНКИ НЕДВИЖИМОСТИ. ПРАКТИКА *

Ключевые слова: индивидуальная оценка недвижимости, компьютерное моделирование, многомерный регрессионный анализ, нелинейные зависимости, автоматизация построения регрессионной модели

Аннотация

Рассмотрено построение многофакторных регрессионных моделей с нелинейными зависимостями стоимости от ценообразующих факторов в задачах индивидуальной оценки недвижимости. Отмечено, что неумение выявлять нелинейные зависимости сдерживает применение регрессионного анализа в оценке. Представлен принцип и алгоритм выявления нелинейных зависимостей стоимости от каждого из факторов. Показана пошаговая процедура автоматизированного построения многомерной регрессионной модели с нелинейными связями в среде электронных таблиц MS Excel. Сделаны выводы о практической доступности построения адекватных регрессионных моделей на активных рынках недвижимости.

Одной из особенностей построения многомерных регрессионных моделей при оценке недвижимости является преобладание нелинейных зависимостей моделируемой величины от влияющих факторов — см., например, [1–3]. Оценщику необходимо определять вид нелинейной зависимости моделируемой величины (стоимости, рыночной арендной платы) от каждого из существенно влияющих факторов.

Умение выявлять реальные нелинейные зависимости в огромной степени определяет успех корректной спецификации регрессионных моделей. Представляется, что именно отсутствие такого умения сдерживает применение многомерного регрессионного анализа в индивидуальной оценке недвижимости. В источниках, посвященных применению регрессионного анализа в оценке недвижимости (см., например, [4, 5]), можно встретить рекомендации по выбору расчетной модели из ограниченного набора относительно простых регрессионных моделей — линейных $Y = a_0 + a_1 X_1 + \dots + a_n X_n$ либо приводимых к линейному (по коэффициентам модели) виду путем логарифмирования, например, степенной: $Y = a_0 X_1^{a_1} \dots X_n^{a_n}$, показательной: $Y = a_0 a_1^{X_1} a_2^{X_2} \dots a_n^{X_n}$, экспоненциальной: $Y = a_0 \exp\{a_1 X_1 + \dots + a_n X_n\}$ и др., где Y — модельное среднее значение зависимой переменной, X_1, \dots, X_n — значения ценообразующих признаков (факторов) объектов недвижимости, a_0, a_1, \dots, a_n — коэффициенты модели.

Моделей, из которых осуществляется выбор, может быть больше, но предлагаемый прием понятен — нужно построить все известные модели, оценить качество каждой из них по определенным статистическим критериям и выбрать лучшую. Иными словами, предлагается вести перебор известных

* ¹ Настоящий материал призван помочь в освоении практического применения регрессионных моделей, и составлен с учетом совместного использования с расчетным файлом MS Excel «ВПР и Поиск решения_2021», который доступен по адресу <http://srroo.ru/evaluators/bank/648/668/>

моделей. При этом не гарантируется, что лучшая из известных моделей корректно описывает ценообразование на рассматриваемом конкретном рынке. Она — лучшая из рассмотренных моделей, не более того. Часто используемый на практике отбор модели по единственному критерию — коэффициенту детерминации (R -квадрат) — является глубоко ошибочным, и его трудно воспринимать иначе как примитивизм в понимании многофакторного регрессионного анализа.

Альтернатива такому подходу — построение модели, учитывающей индивидуальные зависимости от каждого из факторов на рассматриваемом рынке, которые, как правило, имеют различный вид, аппроксимируемый разными аналитическими выражениями, в том числе достаточно сложными.

Десять лет назад в докладе на IV Поволжской научно-практической конференции [6] было показано, что для зависимостей, заданных дискретно (таблично), всегда существует преобразование значений влияющего фактора, которое приводит связь зависимой переменной модели с этим фактором к линейному виду (рис. 1). 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 преобразованных значений этих же факторов².

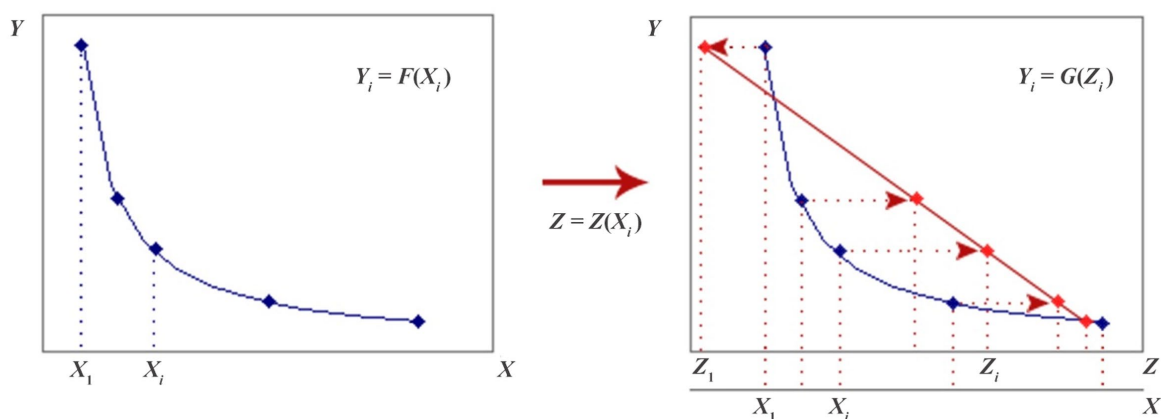


Рис. 1. Графическое представление преобразования $Z = Z(X)$ шкалы ценообразующего фактора x для линейзации зависимости $Y = G(Z)$ от преобразованных значений Z фактора X

В практике индивидуальной оценки преобразование каждой переменной может быть найдено в том числе совместным применением инструментов MS Excel: функции «ЛИНЕЙН», табличной формы задания функций и макроса «Поиск решения». Оптимизация проводится по минимуму среднеквадратического отклонения (СКО) или по максимуму коэффициента детерминации (R -квадрат), что эквивалентно [7]. Можно преобразовывать последовательно каждую из зависимостей, но этот процесс является итерационным — на каждом шаге итерации «наилучшая» нелинейная зависимость от одного фактора находится при неоптимизированных зависимостях от остальных факторов. Поэтому лучше и быстрее искать нелинейное преобразование для всех влияющих переменных³ одновременно.

Идея совместного применения функции «ЛИНЕЙН», описывающей регрессионную модель с линейными связями, и макроса «Поиск решения» состоит в том, что значения ценообразующего фактора при поиске должны изменяться таким образом, чтобы зависимость моделируемой величины от преобразованных значений фактора максимально приближалась к линейной функции, повышая показатели качества линейной регрессионной модели. Тем самым «автоматически» выявляется нелинейная зависимость моделируемой величины (оценки стоимости) от исходных значений ценообразующего фактора⁴.

В процессе описанного преобразования координат необходимо решить две сопутствующие задачи. Во-первых, все повторяющиеся значения влияющего фактора до преобразования должны также повторяться и в преобразованном виде. Во-вторых, для получения монотонных зависимостей⁵ требуется задавать порядок следования цифровых меток влияющего фактора в макросе «Поиск решения». При

² Заметим, что для массовой оценки действие этого утверждения ограничивается лишь располагаемыми вычислительными мощностями. Если таких мощностей достаточно, такое преобразование также найдется всегда.

³ Для бинарных (имеющих лишь два значения) переменных выполнять указанную процедуру не нужно, поскольку невозможно задать или выявить нелинейность на двух точках.

⁴ Моделируемая зависимая переменная линейно связана с новыми значениями фактора, поэтому вид полученной зависимости справедлив и для зависимости преобразованных значений фактора от исходных его значений (с точностью до линейного преобразования — масштаба, сдвига, а также знака).

⁵ Зависимости стоимости от ценообразующих факторов в большинстве случаев индивидуальной оценки недвижимости носит монотонный характер. Встречающиеся немонотонные зависимости можно моделировать как совокупность монотонных участков.

возрастающей зависимости необходимо, чтобы каждое последующее значение фактора было не меньше предыдущего, при убывающей зависимости каждое последующее значение должно быть не больше предыдущего. Решение обеих сопутствующих задач можно ускорить с помощью функции «ВПР» MS Excel.

Рассмотрим по шагам автоматизированную процедуру построения многофакторной регрессионной модели с нелинейными связями с использованием табличного задания функций влияния ценообразующих факторов и инструментов MS Excel — функций «ЛИНЕЙН», «ВПР» и макроса «Поиск решения».

Пусть имеется выборка двухкомнатных квартир-аналогов, расположенных в многоквартирных домах в районе новостроек с количественными (расстояние до метро, общая площадь, площадь кухни), и неколичественными (состояние, тип дома, этаж) данными (см. таблицу).

Выборка оцифрованных исходных данных для построения регрессионной модели ценообразования квартир⁶

Аналог	Ценообразующие факторы					Зависимая переменная Y
	X1	X2	X3	X4	X5	
	До метро, м	Состояние	Тип дома	Общая площадь, м ²	Площадь кухни, м ²	Цена за 1 м ²
1	1500	1	1	58,0	11,5	75 862
2	1200	2	3	65,0	9,0	90 769
3	440	2	1	52,7	10,7	94 877
4	190	1	2	62,0	12,7	87 903
5	600	2	4	56,0	9,0	103 571
6	710	1	1	58,0	7,4	78 448
7	710	2	1	52,5	7,8	91 429
8	490	1	1	52,0	8,0	86 538
9	1700	2	2	63,0	13,0	87 302
10	550	1	1	51,7	7,9	90 909
11	930	2	1	56,0	10,0	98 214
12	1180	2	3	59,0	10,5	98 305
13	380	2	1	56,4	10,1	88 652
14	80	1	1	60,0	10,5	81 333

Применяя функцию ЛИНЕЙН к полученной матрице значений влияющих переменных и вектор-столбцу зависимой переменной (удельных цен) и проводя дополнительные построения (см. файл MS Excel), получаем характеристики регрессионной модели, линейной относительно исходных координат (рис. 2).

	Площадь кухни	Общая площадь	Тип дома	Состояние	До метро	
	A5	A4	A3	A2	A1	A0
	576,8	-951,4	4732,3	8201,0	-3,6	120510,6
	746,7	368,1	1369,1	2485,8	2,6	17812,3
R-квадрат	0,846	3875,0	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	8,78	8	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	659181115	120127936	#Н/Д	#Н/Д	#Н/Д	#Н/Д
t Stat	0,772	2,585	3,457	3,299	1,418	
P-value	0,231	0,016	0,004	0,005	0,097	
Фактор значим на	77%	98%	100%	99%	90%	
Знак коэф.	логичен	логичен	логичен	логичен	логичен	
s-критерий		1,3E+07				

Рис. 2. Фрагмент файла MS Excel — характеристики регрессионной модели с линейными зависимостями

Как видно, модель с линейными зависимостями нельзя признать адекватной рынку. Два из пяти факторов, признаваемых участниками рынка существенно влияющими на цену, не значимы на требуемом ($\geq 95\%$) уровне. Кроме того, линейная зависимость стоимости квартир от расстояния до метро в интервале 80...1700 м также не соответствует известным закономерностям ценообразования на рынке. Знаки коэффициентов данной модели соответствуют экономическим гипотезам⁷ (например, с удалением от станции метро средняя удельная цена квартиры при прочих равных падает).

⁶ Исходные данные и подробности см. в файле MS Excel «ВПР и Поиск решения_2021».

⁷ Достаточно часто на этом шаге построения модели встречаются алогичные знаки коэффициентов, противоречащие зависимостям ценообразования на рынке.

Попробуем учесть нелинейные зависимости удельной стоимости Y от влияющих переменных путем преобразования с помощью макроса «Поиск решения» значений переменных таким образом, чтобы зависимости удельной стоимости от преобразованных значений были близки к линейным.

Рассматривая диапазоны изменения значений влияющих переменных, можно предположить, что все зависимости в пределах этих диапазонов можно считать монотонными, поскольку оснований предполагать отсутствие монотонности не наблюдается. Чтобы обеспечить монотонность зависимостей, соответствующих экономическим гипотезам о влиянии факторов на моделируемую величину (удельную стоимость), в макросе «Поиск решения» нужно задать ограничения на порядок следования цифровых меток (значений преобразованных переменных). Транскрипция макроса позволяет задавать условия равенства/неравенства в виде массивов, однако для этого необходимо определенным образом подготовить исходные данные.

Покажем, как это удобно сделать с помощью функции «ВПР».

Пошаговый алгоритм действий опишем на примере переменной X_1 — расстояние «До метро». С остальными переменными необходимо поступать аналогичным образом. Удобно задавать одновременное преобразование всех исходных переменных (кроме бинарных) до запуска процедуры «Поиск решения».

Шаг 1. Вначале скопируем значения исходного столбца переменных X_1 в отдельный столбец P112:P125 (см. рис. 3, а также файл MS Excel).

Шаг 2. Затем в соседнем столбце (Q112:Q124) расположим значения тех же данных, но отсортированные по возрастанию и очищенные от повторяющихся значений⁸. Эти операции можно выполнить с использованием инструментов «Сортировка» и «Удалить дубликаты» раздела «Данные» MS Excel. В данном случае уделено одно повторяющееся значение — 710.

Шаг 3. В следующем столбце (R112:R124) должны находиться значения преобразованных данных. Но сначала (до запуска процедуры «Поиск решения») сюда нужно скопировать значения предыдущего столбца (Q112:Q124).

Шаг 4. Соседний справа столбец (S112:S123) нужен для быстрого задания ограничений, обеспечивающих монотонность преобразованных переменных. Этот столбец представляет собой предыдущий столбец, сдвинутый на строку вверх. Для его задания нужно прописать в ячейке S112 «=R113» и «протаскать» ее до строки 123 включительно (в этом столбце данных на единицу меньше, чем в предыдущем, поскольку не используется первый элемент).

Таким образом, в столбце P находится ряд исходных значений фактора, в столбце Q — упорядоченные и свободные от повторений исходные значения, в столбце R — значения, которые будут изменяться при запуске макроса «Поиск решения», а в столбце S — связанные с ними «сдвинутые на шаг» значения (для задания ограничений в виде неравенства двух массивов).

	P	Q	R	S
109		До метро		
110				
111				
112	1500	80	80	190
113	1200	190	190	380
114	440	380	380	440
115	190	440	440	490
116	600	490	490	550
117	710	550	550	600
118	710	600	600	710
119	490	710	710	930
120	1700	930	930	1180
121	550	1180	1180	1200
122	930	1200	1200	1500
123	1180	1500	1500	1700
124	380	1700	1700	
125	80			

Рис. 3. Подготовка значений ценообразующего фактора «До метро» для применения функции «ВПР» при замене переменных в регрессионной модели

⁸ Повторяющиеся значения (дубликаты) неудобны при построении зависимостей преобразованных значений фактора от исходных. Удаление дубликатов не влияет на кодирование в матрице факторов для регрессионного уравнения.

Теперь, выбирая любое исходное значение из столбца Р, можно с помощью функции ВПР искать равное ему значение из столбца Q и ставить ему в соответствии рядом стоящее преобразованное значение из столбца R.

Шаг 5. Определяя в массиве исходных данных для построения регрессионной модели ячейку G112 как «=ВПР(P112;\$Q\$112:\$R\$124;2;0)» — см. рис. 4 и файл MS Excel — и «протащив» ее до конца ряда данных, получим столбец переменных G112:G125, однозначно связанный со столбцом R.

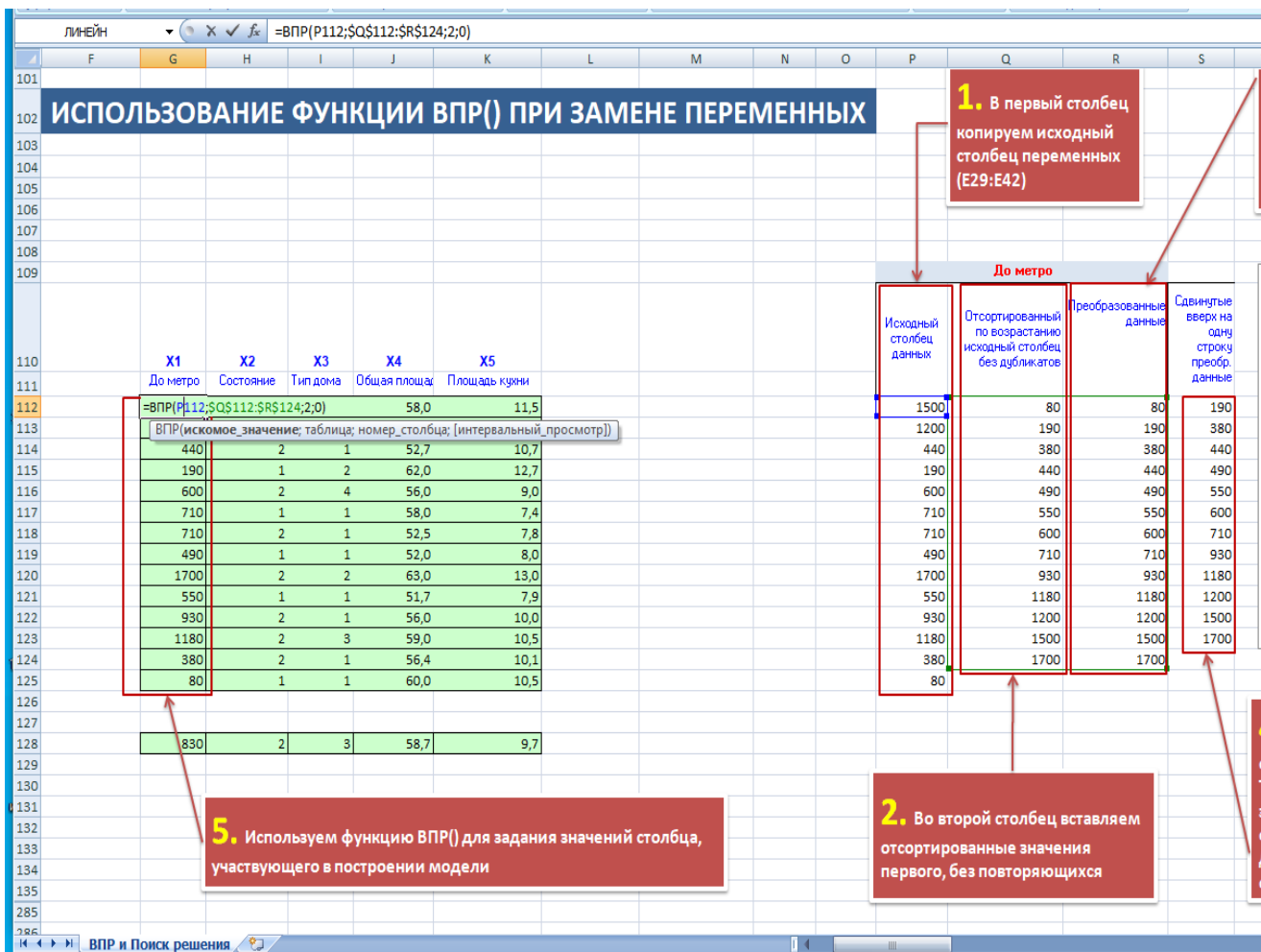


Рис. 4. Использование функции «ВПР» для задания матрицы значений влияющих факторов при замене переменных

Важно не забывать «закрепить» (знаки \$ в формуле) таблицу с данными, а также **ОБЯЗАТЕЛЬНО** поставить в формулу функции ВПР в графу «интервальный просмотр» значение «0» или «Ложь».

Таким образом, отпадает необходимость назначать значения ячеек в столбце G вручную и обеспечивается их изменение макросом «Поиск решения», причем повторяющиеся исходные значения повторяются и в преобразованном виде.

Шаг 6. Аналогичная процедура проводится с другими оставшимися влияющими переменными, кроме бинарных⁹.

Затем с помощью функции ЛИНЕЙН строим регрессионную модель (т. е. получаем коэффициенты регрессионного уравнения). Для этого указываем массив значений зависимой переменной (удельной цены) B112:B125 и массив значений влияющих переменных (ценообразующих факторов) G112:K125, т. е. =ЛИНЕЙН(B112:B125;G112:K125;1;1) — рис. 5.

Для выявления существующих нелинейных зависимостей обращаемся к инструменту «Поиск решения» и выбираем в качестве целевой ячейки ячейку D191 со значением R-квадрат (см. рис. 6, а также «Шаг 3» в файле MS Excel)¹⁰. Указываем в меню «Поиск решения» целевую ячейку, условие оптимизации (в данном случае — максимизация значения целевой ячейки — R-квадрат), вектор

⁹ Для бинарных переменных (в данном случае — «Состояние») выполнять указанную процедуру можно, но не имеет смысла — выявить нелинейность на двух точках невозможно.

¹⁰ В жизни целевой остается ячейка D142 на шаге 2, но для иллюстрации каждого отдельного шага процедуры данные повторяются на шаге 3 с неизбежным изменением адресов ячеек электронных таблиц.

изменяемых значений переменных и ограничения¹¹. Аналогичным образом задаются условия для оптимизации по всем небинарным факторам модели¹².

Как уже отмечалось, регрессионная модель с линейными связями не может быть признана адекватной рынку.

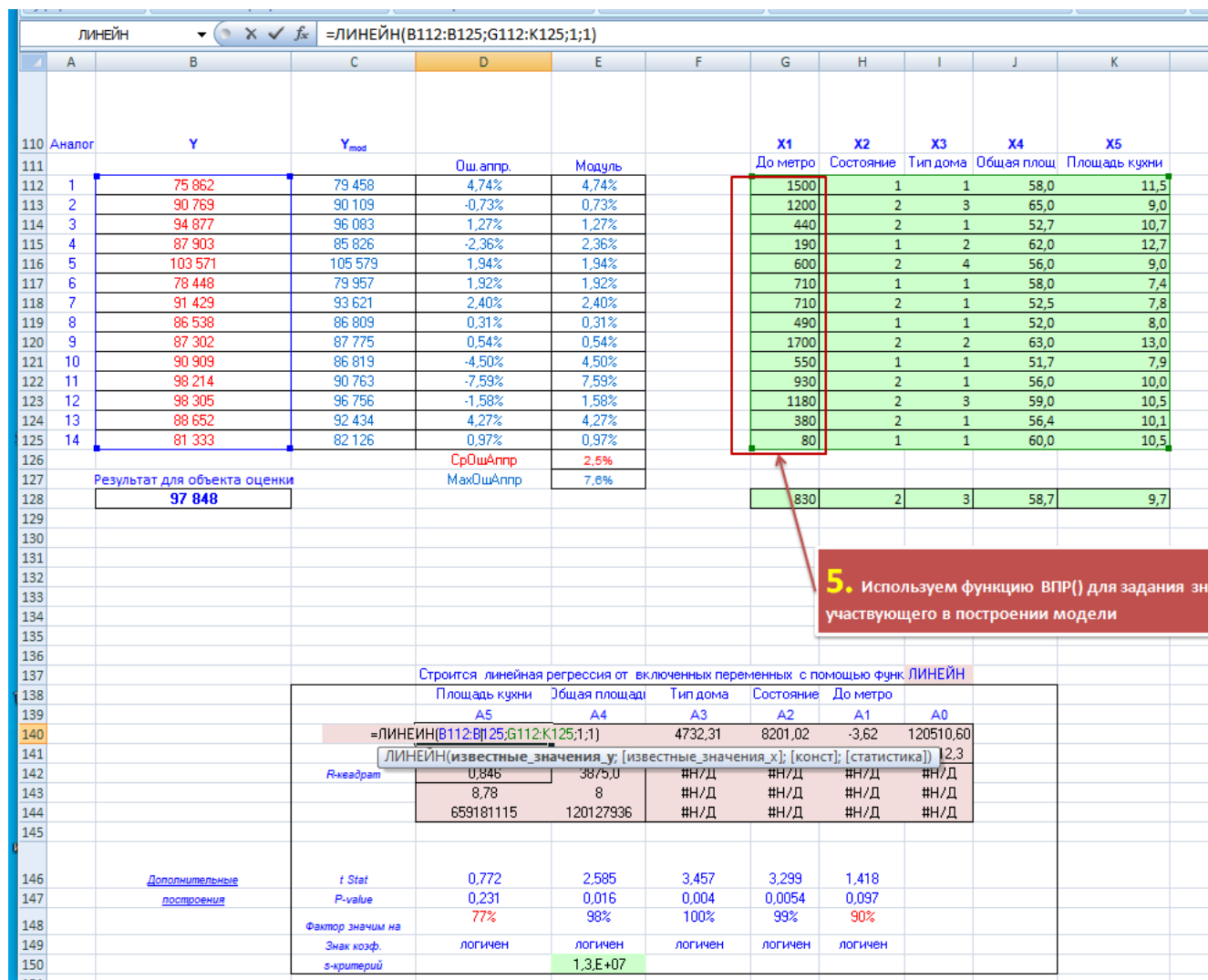


Рис. 5. Использование функции «ВПР» для задания значений влияющих факторов и показатели регрессионного уравнения с линейными связями

Запуская макрос «Поиск решения», в итоге (после оптимизации одновременно по всем факторам) получаем следующие значения для фактора «До метро» (рис. 7)¹³.

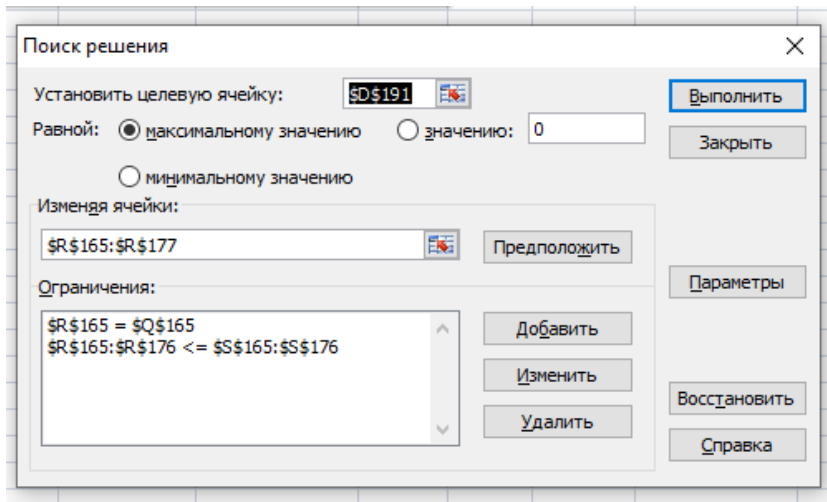
Нелинейный характер полученной зависимости (с учетом отрицательного знака коэффициента регрессии при этом факторе) вполне соответствует имеющимся рыночным данным о влиянии удаления от станций метрополитена в рассматриваемом интервале дистанций. В пешеходной доступности (примерно до 1000 м) удаление от метро почти не сказывается, затем (с необходимостью пользоваться общественным транспортом) цены заметно падают. Вместе с тем эта зависимость (см. точки на графике) не является гладкой. После применения макроса «Поиск решения» в силу наличия в ценах случайной составляющей получаются зависимости, удовлетворяющие требованию монотонности, но, как правило, не такие гладкие, как ожидается согласно имеющимся экономическим гипотезам о влиянии факторов на стоимость недвижимости¹⁴.

¹¹ Здесь для иллюстрации указаны изменяющиеся ячейки и ограничения только для фактора «До метро». При одновременной линеаризации всех необходимых зависимостей сюда нужно также добавить изменяющиеся ячейки (через знак «;» с помощью кнопки «Добавить») и неравенства для задания ограничений для остальных переменных.

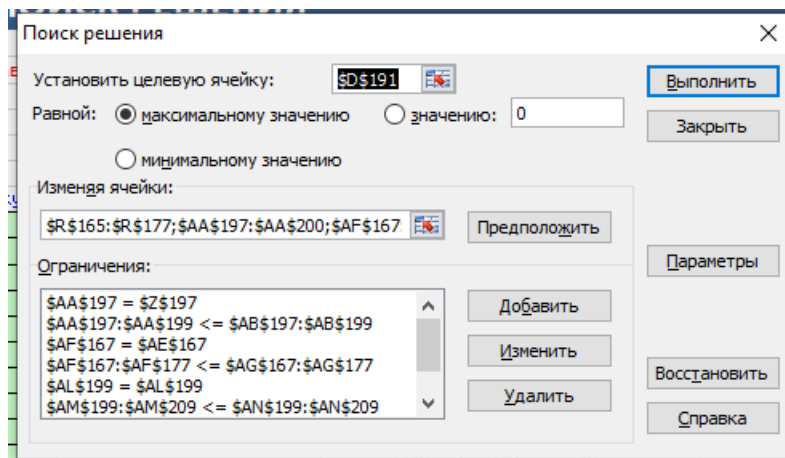
¹² Полностью данные можно увидеть в файле MS Excel.

¹³ Исходные и преобразованные значения, а также графики их зависимостей для других факторов см. файл MS Excel.

¹⁴ Пример пары таких зависимостей до и после сглаживания см. также в [1].



а



б

Рис. 6. Задание целевой ячейки, условия оптимизации, вектора изменяемых значений и ограничений: а — для фактора «До метро»; б — для всех факторов

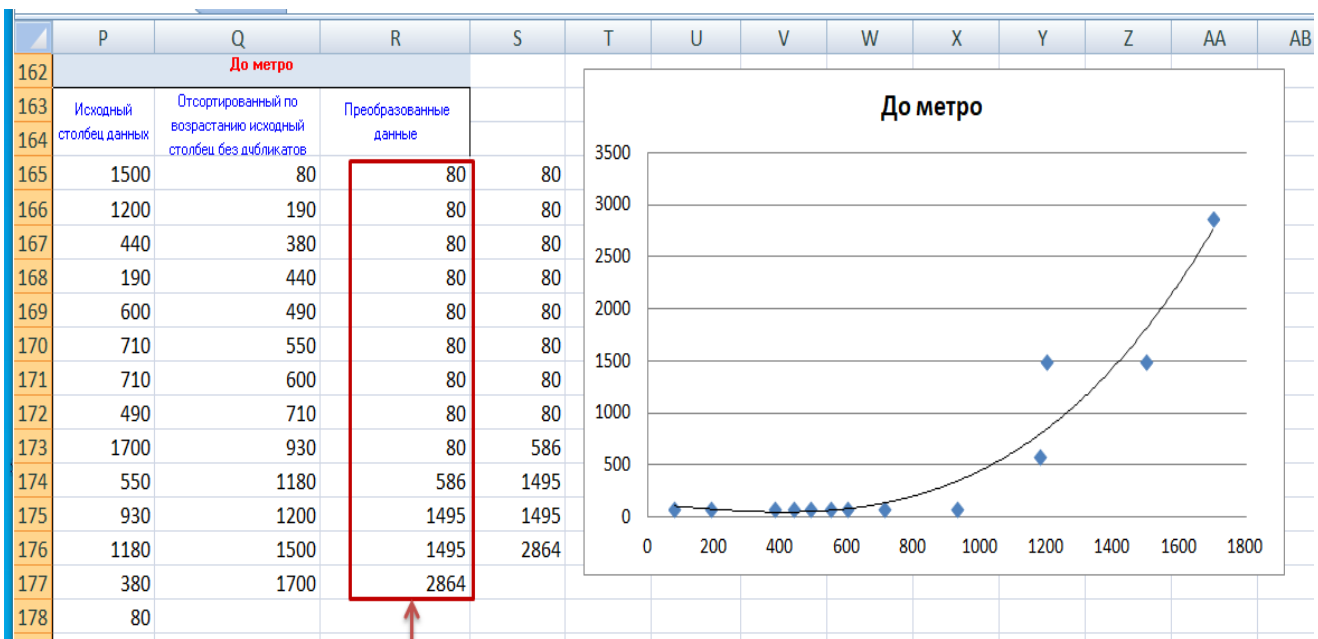


Рис. 7. Исходные и преобразованные значения фактора «Расстояние до метро» и график выявленной зависимости

Условия гладкости трудно (если вообще возможно) задать в рамках имеющегося инструментария «Поиск решения», поэтому требуемая гладкость обеспечивается ручным изменением некоторых преобразованных значений с ориентацией на график зависимости для этой переменной. В рассматриваемом случае понадобилось внести изменения лишь в два преобразованных значения для получения гладкой кривой (см. рис. 8 и файл MS Excel).

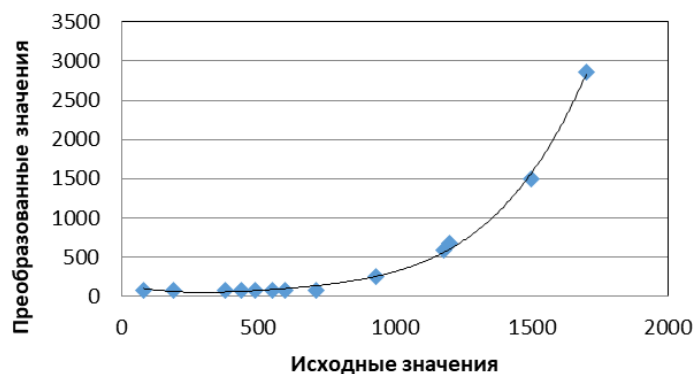


Рис. 8. График «сглаженной» вручную зависимости от расстояния до метро

Аналогичные графики строят для всех переменных, подвергнутых линейаризирующим преобразованиям, и для каждого из них, если требуется, применяют «ручное сглаживание» — см. файл MS Excel.

Необходимо помнить, что соответствие использованных в модели зависимостей закономерностям ценообразования на рынке является одним из существенных условий адекватности регрессионной модели.

Обсудим регрессионную модель, полученную в результате выявления и сглаживания нелинейных зависимостей, соответствующих рыночным данным (см. рис 9 и файл MS Excel).

	Площадь кухни	Общая площадь	Тип дома	Состояние	До метро	
	A5	A4	A3	A2	A1	A0
	1204,98	-1543,61	4587,00	6952,29	-4,12	150587,33
	160,0	111,7	307,3	601,9	0,4	6006,4
<i>R-квадрат</i>	0,991	936,3	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	176,18	8	#Н/Д	#Н/Д	#Н/Д	#Н/Д
	772295320,3	7013730	#Н/Д	#Н/Д	#Н/Д	#Н/Д
<i>t Stat</i>	7,530	13,817	14,927	11,550	10,509	
<i>P-value</i>	3,4E-05	3,6E-07	2,0E-07	1,4E-06	2,9E-06	
Фактор значим на	100%	100%	100%	100%	100%	
Знак коэф.	логичен	логичен	логичен	логичен	логичен	
<i>s-критерий</i>		7,8E+05				

Рис. 9. Показатели регрессионного уравнения с нелинейными связями после ручного сглаживания

В целом показатели модели заметно улучшились — все пять факторов модели значимы с существенным запасом, знаки при их коэффициентах логичны, полученные нелинейные зависимости не противоречат экономическим гипотезам о характере влияния факторов на стоимость, распределение остатков модели (см. файл MS Excel) не содержит каких-либо закономерностей (т. е. остатки можно считать случайными).

Отметим в заключение, что построение адекватной многофакторной регрессионной модели индивидуальной оценки недвижимости за три шага в среде электронных таблиц MS Excel с использованием встроенных в них инструментов (функций ЛИНЕЙН, ВПР и макроса ПОИСК РЕШЕНИЯ) можно считать вполне автоматизированным процессом.

Такой процесс делает оценку недвижимости методом регрессионного анализа вполне возможной и реализуемой на практике для каждого оценщика. Рыночных данных на активных рынках недвижимости для этого хватает и невозможно объяснить их неиспользование с методической точки зрения.

Литература

1. *Баринов Н.П.* Применение регрессионного анализа в задачах индивидуальной и массовой оценки недвижимости / 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> (дата обращения 14.04.2022).
2. *Баринов Н.П.* Оценка рыночной стоимости земельного участка методом многомерного регрессионного анализа // Информационно-аналитический бюллетень рынка недвижимости RWAY. 2014, июль. № 232 (часть 1); 2014, ноябрь. № 236 (часть 2). URL: <http://www.appraiser.ru/default.aspx?SectionID=35&Id=3821> (дата обращения 14.04.2022).
3. Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. Т. 3: Определение кадастровой стоимости земельных участков. СПб.: СПб ГБУ «Городское управление кадастровой оценки», 2020, 298 с. URL: <http://www.ko.spb.ru/interim-reports> (дата обращения 14.04.2022).
4. *Грибовский С.В., Сивец С.А., Левыкина И.А.* Математические методы оценки стоимости имущества. М.: Маросейка, Книжная линия, 2014. 352 с.
5. Методические указания о государственной кадастровой оценке (утверждены приказом Министерства экономического развития РФ от 12 мая 2017 г. № 226). URL: <http://base.garant.ru/71686152/#ixzz6wShdxtfY> (дата обращения 14.04.2022).
6. *Зельдин М., Баринов Н., Ситников Н.* 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, 31.03–01.04.2011. URL: <http://www.appraiser.ru/default.aspx?SectionId=35&Id=3605> (дата обращения 14.04.2022).
7. *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. URL: <http://www.appraiser.ru/default.aspx?SectionId=41&Id=1575> (дата обращения 14.04.2022).

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

Статья поступила в редакцию 06.05.2022