

Барьер на пути человеческого понимания MARS (The Barrier to Human Understanding of MARS)

Чтение модели MARS как удельных ставок, вкладов в стоимость и корректировок

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

DOI: [10.67330/qmnctv78](https://doi.org/10.67330/qmnctv78)

Это дополнение к разделу «Основы», а не статья, посвящённая концепции. Его название отражает препятствие, которое редактор заметил за годы косвенных наблюдений: оценщики, без труда интерпретирующие коэффициент регрессии, останавливаются на списке «шарнирных» терминов, выводимых программой MARS, поскольку ни один из элементов этого списка не выглядит как коэффициент. В четырёх статистических статьях, следующих за этим дополнением (*Условная математическая ожидаемая величина, Систематическая ошибка пропущенной переменной, Эндогенность и Идентификация параметров*), рассматривается ставка, которую регрессия присваивает тому или иному признаку. В собственной модели журнала используются многомерные адаптивные регрессионные сплайны (MARS; Friedman 1991)¹, которые предоставляют не одну ставку на каждый признак, а по одной на каждый диапазон, и путь от того, что выводит программа MARS, до таблицы ставок, которую оценщик может обосновать, содержит один этап, на котором таблицы, составленные вручную, дают сбой. В данном приложении приведены этот путь, указанный этап, а также программа, которая проходит этот путь и проверяет результаты своей работы.

Формальное определение. В модели MARS аппроксимированное значение представляет собой пересечение с осью абсцисс плюс сумму слагаемых, каждое из которых состоит из коэффициента, умноженного на произведение шарнирных функций признаков. Аппроксимированное значение — это оценка цены, полученная с помощью модели. Для одной характеристики, при фиксированных значениях остальных, это ряд прямых отрезков. Точки пересечения отрезков являются узлами. График вклада значения характеристики в диапазоне её значений представляет собой *график переменной*. Его наклон в каждом диапазоне является *темпом*: изменением аппроксимированного значения на единицу характеристики в данном диапазоне; *вклад значения* для данного значения признака представляет собой сумму слагаемых модели по этому признаку при данном значении; в примере все эти слагаемые равны нулю при 1 200 квадратных футов, поэтому вклад измеряется от этой отметки и является отрицательным ниже нее; модель не обязательно должна иметь такое значение, и значения имеют только различия между вкладами, которые являются корректировками; а *корректировка* от аналога к объекту оценки — это величина, добавляемая к вкладу аналога, чтобы сделать его равным вкладу объекта оценки, что представляет собой сумму, по диапазонам между двумя значениями, коэффициента, умноженного на ширину. Определённая таким образом ставка является ставкой модели; она интерпретируется как ставка, которую рынок платит только после того, как модель была признана справедливым отражением продаж, а записи, следующие за этим дополнением, касаются того, когда такое признание оправдано.

¹MARS is a registered trademark of Minitab, LLC. It is used in this Journal as the name of the statistical method introduced by Friedman (1991), and the software referred to is the R package earth, whose notation this Journal follows: hinges are written $h(x - t)$ and $h(t - x)$ and models are shown as earth prints them. Salford's MARS writes the same functions as numbered basis functions, $BF1 = \max(0, X - t)$ and so on, and differs from earth in such details of notation, not in the method.

Зачем нужно это дополнение

Оценщику, соответствующему модели MARS, в earthUI (Craytor, 2026) или непосредственно в пакете R earth (Milborrow, 2024a), предоставляется список таких терминов:

(Intercept)	1520000
h(1200-gla)	-600
h(gla-1200)	1000
h(gla-2000)	-250
h(gla-2800)	-650

Об этом никто не сообщает. В таблицу продаж вносится корректировка: вклад объекта оценки в стоимость за вычетом вклада аналога. Ставки за квадратный фут обсуждаются в отчёте как один из этапов обоснования модели MARS. Путь от напечатанных условий до этих цифр короткий, но в нём есть один этап, на котором вручную составленные таблицы дают сбой, и весь смысл использования модели теряется, если полученные с её помощью цифры невозможно объяснить рецензенту.

Простое изложение: четыре диапазона, четыре ставки

Линейная регрессия, в которой жилая площадь учитывается в качестве одного линейного члена, даёт единый коэффициент для всего рынка; именно так работают регрессии в традиционной оценке, и именно это подразумевается под термином «обычная регрессия» в данном приложении. Система MARS позволяет изменять коэффициент там, где данные о продажах указывают на его изменение. На примере прибрежного рынка жилая площадь может стоить \$600 за квадратный фут при площади до 1 200 квадратных футов, \$1,000 за квадратный фут — от 1 200 до 2 000, \$750 — от 2 000 до 2 800, и всего \$100 — выше этого уровня. Вот и вся разница, и агент понимает её сразу: рынок платит больше всего за пространство, которое превращает небольшой дом в семейный, меньше — за следующую комнату, и почти ничего — за пространство, выходящее за рамки того, что хотят покупатели в этом районе.

С учётом коэффициентов всё остальное сводится к арифметическим вычислениям, которые оценщик и так уже выполняет. Аналог с площадью 1 800 квадратных футов и объект оценки площадью 2 400 квадратных футов отличаются на 600 квадратных футов, а узел на отметке 2 000 находится между ними. Корректировка, добавляемая к стоимости объекта сравнения, равна разнице между «подгоняемым значением» объекта оценки и «подгоняемым значением» объекта сравнения ($\$2,620,000 - \$2,120,000 = \$500,000$), что также соответствует разнице между вкладом жилой площади объекта оценки и вкладом жилой площади объекта сравнения ($\$1,100,000 - \$600,000$). Ставки объясняют это: изменение превышает 200 квадратных футов при значении \$1,000, затем 400 квадратных футов при значении \$750, что в сумме составляет \$500,000. На рисунке 3 это показано на примере расчётной стоимости (цены, которую модель присваивает дому с такими характеристиками), которая варьируется от \$1,280,000 при площади 800 квадратных футов (самый маленький дом в базе данных) до примерно \$3 million при площади 3 500 (самый большой), причём большая часть роста приходится на площадь до 2 800.

Единственный коэффициент, который могла бы использовать прямая, проходящая через эти две точки, — это тот, который соединяет их: \$500,000 на 600 квадратных футов, что составляет примерно \$833 за квадратный фут: *секущая ставка* этой пары, эффективная, или смешанная, ставка для этого одного движения: среднее значение ставок модели, взвешенное по ширине каждого пересеченного диапазона. Это не один из коэффициентов модели, и для следующей пары это будет другое среднее значение. Обычная регрессия с одной ставкой для жилой площади должна обслуживать каждую пару этим единственным числом, поэтому она не может воспроизвести эти средние значения для каждой

пары; именно это и делает единая ставка: она усредняет изменение, которое, по мнению модели, производит рынок.

Читатель, остановившийся на этом месте, уже знает метод: нужно вычислить разницу между двумя вкладами, а затем объяснить её, ознакомившись со ставкой в каждом диапазоне, просчитав путь, аналог которого соответствует объекту оценки, с учетом этих ставок, и сложив результаты. То, что следует далее, предназначено для читателя, который хочет понять, почему ставки таковы, какая здесь подводная камень, что меняется при взаимодействии функций и как программа проверяет арифметику; таблица ставок и корректировки при этом не меняются.

Уточнённое изложение: шарнирные функции и знак, который упускают

График вклада значения переменной в пределах её области определения назовём *графиком переменной*. Это не кривая в повседневном смысле слова, то есть не что-то, плавно изгибающееся; это сегментированный линейный график, прямой между узлами и изгибающийся только в них. MARS строит его из *шарниров*; Фридман (1991) описал их как усечённые степенные функции, а термин «шарнир» стал общепринятым в более поздней литературе (Milboggow, без даты); Хасты, Тибширани и Фридман (2009, §9.4) называют две такие кривые, соединяющиеся в одном узле, «отражённой парой».

Определение (шарнир). Шарнир — это функция одной переменной x с единственным переломом, называемым *узлом* t , значение которой с одной стороны узла равно нулю, а с другой — равно 1

$$h(x - t) = \max(0, x - t) = \begin{cases} 0 & x \leq t, \\ x - t & x > t, \end{cases} \quad h(t - x) = \max(0, t - x) = \begin{cases} t - x & x < t, \\ 0 & x \geq t. \end{cases}$$

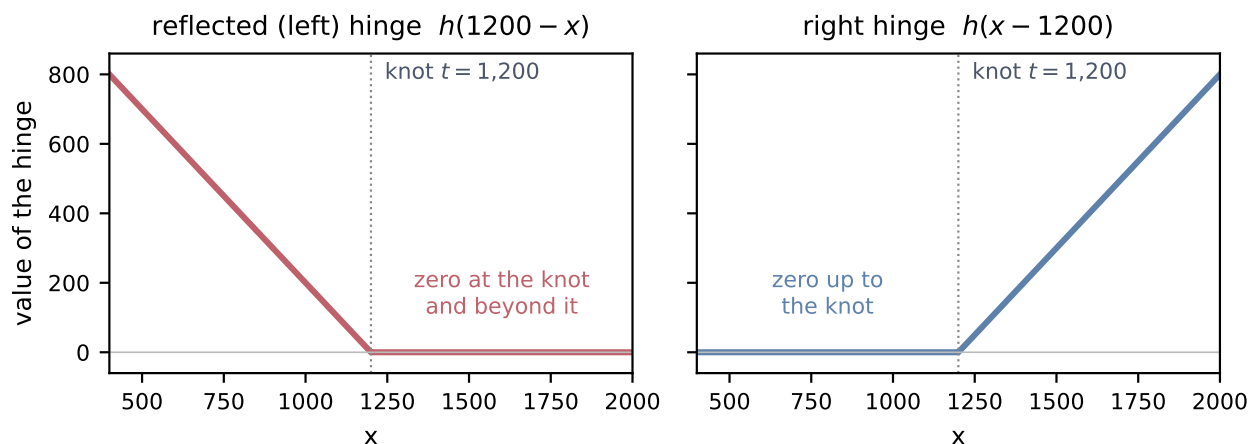


Figure 1: Две шарнирные функции с узлом в точке $t = 1,200$: отражённый (левый) шарнир $h(1200 - x)$, значение которого уменьшается слева направо до нуля в узле и равно нулю за ним; правый шарнир $h(x - 1200)$, значение которого равно нулю до узла и увеличивается слева направо за ним. Это «голые» шарниры; в модели каждый из них умножается на коэффициент, который может изменить направление на противоположное.

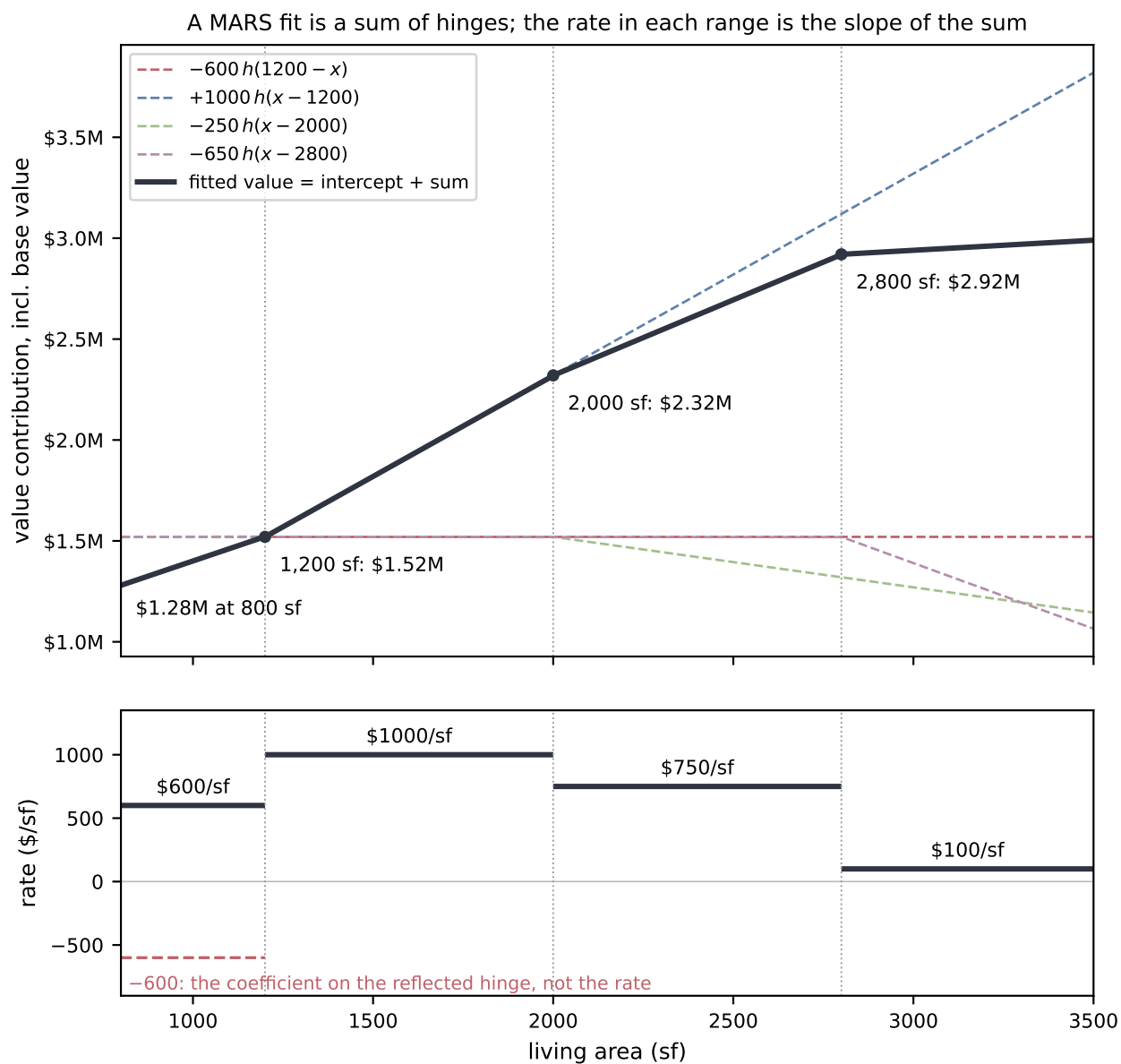
Первая, $h(x - t)$, — *правая петля*: нулевая до узла и за ним возрастающая слева направо. Вторая, $h(t - x)$, — *зеркально отражённая* (или левая) петля: убывающая слева направо до узла и нулевая в узле и за ним. Оба являются шарнирами в одном и том же смысле; они различаются лишь тем,

на какой стороне узла их значение отлично от нуля. Эти направления описывают «голые» шарниры, нанесенные на ось x на рисунке 1, и ничего больше. В модели шарнир умножается на коэффициент, который может быть отрицательным, и складывается с другими слагаемыми, поэтому слагаемое, построенное на правом шарнире, может уменьшаться слева направо, а слагаемое, построенное на отраженном шарнире, — увеличиваться, и график переменной встречает узел на той высоте, которую ему задают остальные слагаемые. Каждое направление, указанное в данном приложении, читается слева направо. Шарниры не используются отдельно. В модели MARS каждый шарнир умножается на коэффициент, а слагаемые суммируются, и именно на основе этих коэффициентов, а не какого-либо отдельного шарнира, вычисляются наклоны модели, то есть ставки: накопительно, диапазон за диапазоном, причем ставка каждого диапазона является ставкой предыдущего диапазона, скорректированной коэффициентами шарниров, узел которых лежит между ними. Шарнир — это единственный строительный блок, которым располагает MARS для одной переменной; каждый изгиб на графике переменной является узлом какого-либо шарнира, а каждый прямой отрезок — суммой слагаемых, которые в этой точке либо равны нулю, либо являются линейными. **Главная мысль, которую следует помнить при рассмотрении всего последующего: коэффициент шарнира не является наклоном. Это корректировка наклона — величина, прибавляемая в узле данного шарнира к наклону, приходящему слева.** Если читать слева направо, график переменной начинается со ставки своего первого отрезка, и каждый узел корректирует эту ставку с помощью коэффициентов, расположенных в этом узле. В этом и заключается всё правило чтения, а теоремы 1 и 2, приведенные ниже, являются его доказательством. Модель представляет собой пересечение с осью абсцисс плюс каждый коэффициент, умноженный на соответствующий шарнир, а наклон в любом интервале равен наклону этой суммы.

Рассмотрим приведенные выше печатные термины; на рисунке 2 каждый из них изображен вместе со своей суммой, а под ним — коэффициент для каждого диапазона; на рисунке 3 показано, как эти коэффициенты объясняют корректировку между аналогом и объектом оценки. Узлы на отметках 1 200, 2 000 и 2 800 образуют четыре диапазона. В каждом диапазоне правый шарнир *активен*, если его узел лежит в пределах диапазона или левее него, а зеркально отраженный шарнир — если его узел лежит в пределах диапазона или правее него; правый шарнир действует во всех диапазонах за пределами своего узла, а зеркально отраженный — во всех диапазонах перед ним. Коэффициент рассчитывается на основе коэффициентов активных шарниров: складываются коэффициенты активных правых шарниров и вычитаются коэффициенты активных зеркальных шарниров. Получившаяся таблица отражает ставку модели в каждом диапазоне, при этом остальные характеристики фиксированы на одном наборе значений (программа использует значения объекта оценки); в модели без взаимодействий это условие ничего не меняет, а в случае 2, приведенном ниже, оно меняет всё:

площадь (кв. футов)	активные петли	стоимость (\$/sf)
менее 1 200	$h(1200 - gla)$ только	$-(-600) = +\mathbf{600}$
от 1 200 до 2 000	$h(gla - 1200)$	1000
от 2 000 до 2 800	$h(gla - 1200), h(gla - 2000)$	$1000 - 250 = \mathbf{750}$
свыше 2 800	все три правые петли	$1000 - 250 - 650 = \mathbf{100}$

Первая строка — это шаг, который упускается из виду. Коэффициент -600 для отражающего шарнира не является показателем $-\$600$ на квадратный фут. Шарнир $h(1200 - gla)$ измеряет, насколько площадь дома меньше 1 200 квадратных футов, а -600 , умноженное на этот недостаток, представляет собой скидку: $\$240,000$ для дома площадью 800 квадратных футов, которая уменьшается до нуля при достижении 1 200. Каждый добавленный квадратный фут уменьшает $\$600$ скидки, поэтому ставка составляет $+\$600$. (При положительном коэффициенте тот же шарнир добавляет ценность за меньший размер, и его коэффициент отрицательный.) Правило, сформулированное однократно: в



```

Model (v = value contribution, gla = living area in sf):
v = 1,520,000 - 600·h(1200-gla) + 1000·h(gla-1200) - 250·h(gla-2000) - 650·h(gla-2800)
as earth prints it:
(Intercept) 1520000 h(1200-gla) -600 h(gla-1200) 1000 h(gla-2000) -250 h(gla-2800) -650
Rate on a segment = (coefficients of the active right hinges) - (coefficients of the active reflected hinges):
800-1200 sf: -(-600) = $600/sf only the reflected hinge at 1200 is active
1200-2000 sf: 1000 = $1000/sf
2000-2800 sf: 1000 - 250 = $750/sf
2800-3500 sf: 1000 - 250 - 650 = $100/sf all three right hinges active

```

Figure 2: Четыре значения шарнирных точек из примера, их сумма и ставка в каждом диапазоне. Коэффициент -600 для отраженной шарнирной точки $h(1200 - gla)$ выделен, чтобы читатель не перепутал его со ставкой; ставка при площади менее 1 200 квадратных футов составляет $+600$.

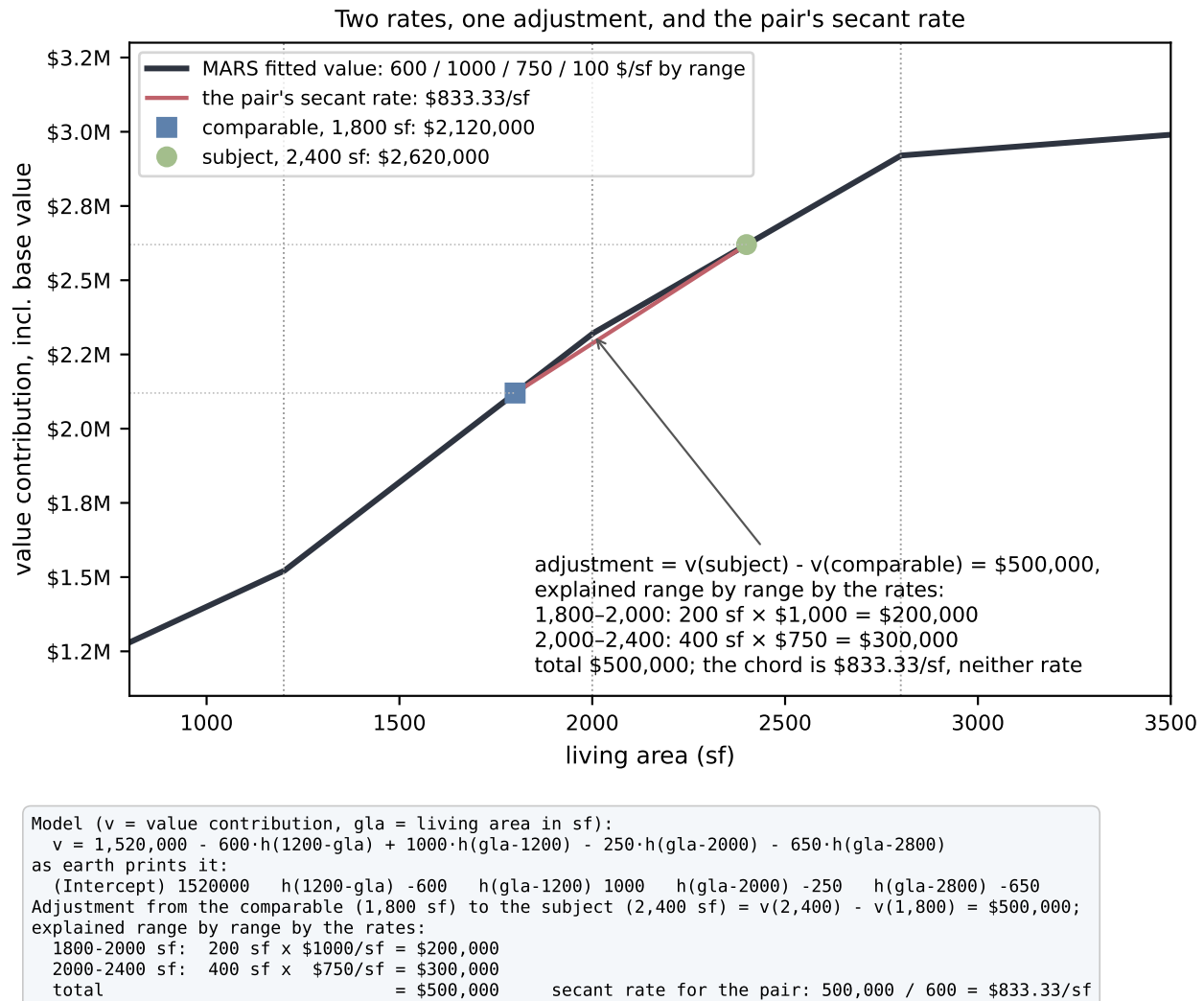


Figure 3: Расчетное значение: аналог — 1 800 квадратных футов, объект оценки — 2 400. Корректировка, то есть разница между двумя коэффициентами, составляет \$500,000 , и действующие коэффициенты объясняют её диапазоном за диапазоном; хорда, соединяющая две точки, равная примерно \$833 за квадратный фут, является секущей этой пары — единственным коэффициентом, который потребовался бы для этой пары и ни для какой другой; обычная регрессия должна использовать по одному коэффициенту для каждой пары.

диапазоне коэффициент равен сумме коэффициентов активных правых шарниров *минус* сумма коэффициентов активных зеркальных шарниров, плюс коэффициент линейного члена (элемента, введенного в виде прямой линии без узла), если он присутствует в модели. С этим связано одно соглашение: ровно в точке узла коэффициент берется справа, так что правый шарнир учитывается, начиная с его узла, а зеркально отраженный шарнир — только строго ниже своего узла; таким образом, коэффициент на площади 1 200 квадратных футов равен \$1,000 — это коэффициент диапазона, начинающегося в этой точке, и программа интерпретирует его точно так же.

Проверка правильности построения таблицы заключается в расчете приспособленных значений в узлах. При площади 1 200 квадратных футов все коэффициенты равны нулю, а значение представляет собой точку пересечения с осью абсцисс: \$1,520,000; при 2 000 — \$1,520,000 + 1,000 × 800 = \$2,320,000; при 2 800 — \$2,320,000 + 750 × 800 = \$2,920,000; а при 800 квадратных футах (самый маленький дом в наборе данных) — \$1,520,000 – 600 × 400 = \$1,280,000. От 800 до 1 200 значение возрастает с \$240,000, то есть со ставкой \$600, что соответствует первой строке; в таблице, где коэффициент интерпретировался бы как ставка, значение должно было бы уменьшаться слева направо вплоть до узла при 1 200, однако данные показывают, что этого не происходит. На рисунке 2 изображены четыре шарнирных члена, их сумма и ставка в каждом диапазоне, при этом коэффициент –600 отмечен там, где читатель мог бы принять его за темп роста.

Еще две сложности, с которыми справляется программа, и которые должна учитывать таблица рукопоятий:

- **Взаимодействия.** Программа earth на этапе прямого прохода (этап, на котором слагаются слагаемые по одному; Friedman 1991, §3) выводит результат вида $h(\text{gla} - 1500) \times \text{lot}$, когда ставка для одного параметра зависит от другого: это и есть *взаимодействие*. В этом случае коэффициент для параметра «lot» различается в зависимости от того, превышает ли жилая площадь 1 500 квадратных футов или нет, а таблица коэффициентов для одного параметра всегда представляет собой таблицу, заданную с учётом значений остальных параметров. Программа принимает точку, то есть значения всех остальных параметров, и вычисляет для неё все коэффициенты.
- **Уровень и корректировки.** Пересечение с осью абсцисс и слагаемые в других характеристиках задают уровень, от которого измеряется вклад; они компенсируются при каждой корректировке. Программа выдает отчёт о вкладах в виде слагаемых только данной характеристики, так что два вклада различаются ровно на величину корректировки.

Два случая: модели без взаимодействий и с ними

Правило, изложенное в предыдущем разделе, можно сформулировать в виде двух теорем, доказанных ниже, и они дают чёткий ответ на вопрос, который рано или поздно задаёт себе каждый читатель, знакомый с моделью MARS: от скольких из напечатанных слагаемых на самом деле зависит данная ставка? Ответ разделяется на два случая.

Случай 1: без взаимодействий

Постановка задачи. Пусть $h(u) = \max(0, u)$ для действительных чисел u . Модель MARS без взаимодействий по признакам $x, z_1, \dots, z_q$ имеет следующее значение:

$$f(x, z_1, \dots, z_q) = \beta_0 + g(x) + \sum_{k=1}^q g_k(z_k),$$

где β_0 — коэффициент отреза, каждый g_k является функцией только z_k , а g объединяет все слагаемые в x :

$$g(x) = c \cdot x + \sum_{j=1}^K a_j h(x - t_j) + \sum_{j=1}^K b_j h(t_j - x).$$

Здесь $t_1 < t_2 < \dots < t_K$ — это различные узлы шарниров в x ; a_j — коэффициент правого шарнира в t_j , а b_j — коэффициент отраженного шарнира в t_j ; оба принимают значение 0, если в модели отсутствует такой член; а c — коэффициент простого линейного члена $c \cdot x$, объекта, введённого в виде прямой линии без узла, который earth генерирует только для объектов, названных в `linpreds`, или, для любого объекта, когда включена опция `Auto.linpreds` (см. примечание после теоремы 1); $c = 0$, когда такого члена нет, как во всех моделях данного приложения. Пусть $t_0 = -\infty$ и $t_{K+1} = +\infty$, и назовем $S_i = (t_i, t_{i+1})$, для $i = 0, 1, \dots, K$, *отрезки* x ; t_i — левый узел S_i , а t_{i+1} — его правый узел. На каждом S_i каждый слагаемый g либо равен нулю, либо линейный, поэтому g линейна на S_i ; её наклон там, который также равен $\partial f / \partial x$ на S_i , поскольку ни один другой слагаемый f не зависит от x , является *ставкой* на S_i и записывается как r_i . Таблица ставок x представляет собой список $r_0, r_1, \dots, r_K$. (Если $K = 0$ существует один сегмент, $(-\infty, +\infty)$, и таблица представляет собой единственную ставку $r_0 = c$.) В самом узле g не имеет производной: его левая и правая производные равны r_{i-1} и r_i , которые различаются на $a_i + b_i$ (Теорема 2 ниже), поэтому значение, которое таблица ставок показывает в точке t_i , является условным обозначением, взятым здесь со стороны правой, а не значением g' .

Теорема 1 (ставка на отрезке). Используя обозначения данной постановки задачи, для каждого $i \in \{0, 1, \dots, K\}$,

$$r_i = c + \sum_{j=1}^i a_j - \sum_{j=i+1}^K b_j,$$

где пустая сумма определяется как 0. В частности, r_i зависит не более чем от K коэффициентов шарниров $2K$ (по одному на каждый узел) вместе с c .

Доказательство теоремы 1. Зафиксируем i и точку $x \in S_i$, так что $t_i < x < t_{i+1}$. Рассмотрим каждый слагаемый выражения g в районе x , достаточно малой, чтобы лежать внутри S_i . Простой член $c \cdot x$ имеет производную c . Для правого шарнира в точке t_j : если $j \leq i$, то $t_j \leq t_i < x$, поэтому $h(x - t_j) = x - t_j$ в окрестности, а его производная равна 1; если $j \geq i + 1$, то $x < t_{i+1} \leq t_j$, поэтому $h(x - t_j) = 0$ в этой точке, а его производная равна 0. Для зеркально отражённого шарнира в точке t_j : если $j \geq i + 1$, то $x < t_j$, поэтому $h(t_j - x) = t_j - x$, а его производная равна -1 ; если $j \leq i$, то $t_j < x$, поэтому $h(t_j - x) = 0$, а его производная равна 0. Конечная сумма функций, дифференцируемых в точке x , дифференцируема в этой точке по сумме производных, что даёт

$$g'(x) = c + \sum_{j \leq i} a_j \cdot 1 + \sum_{j \geq i+1} a_j \cdot 0 + \sum_{j \leq i} b_j \cdot 0 + \sum_{j \geq i+1} b_j \cdot (-1),$$

указанную формулу. Что касается подсчёта: множества индексов $\{j : j \leq i\}$ и $\{j : j \geq i + 1\}$ не пересекаются и вместе исчерпывают $\{1, \dots, K\}$, поэтому каждый узел t_j вносит вклад ровно в один из элементов a_j и b_j , но никогда не в оба; следовательно, в формулу входит не более K коэффициентов шарниров, а r_i одинаково во всех точках S_i . ■

Предел достигается: если $b_j = 0$ для каждого j , то крайний правый отрезок S_K использует все K из a_j .

Примечание (когда значение параметра «с» отлично от нуля). Программа earth может представить объект в виде одной прямой линии без узлов — термина « $c \cdot x$ » в настройках — двумя разными

способами, и это различие имеет значение. Параметр «`linpreds`», представляющий собой список именованных объектов, таких как «`linpreds = c("GLA")`», заставляет каждый объект из списка вводиться линейно, даже если подгонка с узлами была бы лучше; это решение аналитика, и оно может ухудшить качество подгонки. Параметр «`Auto.linpreds`», единственный переключатель, включённый по умолчанию, ничего не принуждает: он лишь изменяет то, как `earth` выводит признак, подгонка которого оказалась действительно линейной — с одним шарниром, узлом которого является наименьшее наблюдаемое значение, — выводя его в виде коэффициента и имени переменной вместо шарнира; подгонка остаётся той же. В любом случае `earth` выводит терм в виде простого имени переменной с её коэффициентом — $221.4 * GLA$ — в сводной форме `earth`, как в примере ниже. Theorem 1 обрабатывает это с помощью `c`, который добавляет коэффициент к ставке на каждом сегменте. Например, вызов

```
mod_linear <- earth(
  SalePrice ~ GLA + LotSize + BathrmCount + BedrmCount,
  data = sales,
  degree = 1,
  linpreds = c("GLA", "LotSize", "BathrmCount", "BedrmCount")
)
```

вводит каждую характеристику линейно (с параметром `degree = 1`, так что произведения линейных слагаемых также не подгоняются) и выводит в сводной форме `earth` (иллюстративный вывод, а не подстроенная модель):

```
SalePrice =
  179843
+ 221.4 * GLA
+ 11.8 * LotSize
+ 34872 * BathrmCount
+ 17641 * BedrmCount
```

Для жилой площади это настройка с параметром $c = 221.4$ и без шарниров, $K = 0$: имеется один сегмент, и Теорема 1 дает единую ставку $r_0 = c = \$221.40$ за квадратный фут по всему диапазону данных; размер участка, количество ванных комнат и спален интерпретируются аналогично: $\$11.80$ за квадратный фут участка, $\$34,872$ за ванную комнату и $\$17,641$ за спальню. Такая модель представляет собой обычную регрессию в смысле данного приложения: одна ставка на одну характеристику. На данных такой член представляет собой прямую линию с единственным правым изгибом, узел которого находится на наименьшем наблюдаемом значении, а журнал «Journal» использует формулу `Auto.linpreds = FALSE`, поэтому в его моделях признак, график переменной которого на данных является прямой, отображается как этот единственный изгиб, например, $h(gla-800) \ 700$ — узел, соответствующий самому маленькому дому в данных, с $c = 0$, $K = 1$ и $a_1 = 700$: ставка равна 0 ниже узла, на отрезке, где продаж нет, и 700 выше него, поэтому таблица ставок над данными имеет одну строку. Отсюда следует одно следствие; изменение ставки в узле является Теоремой 2 ниже.

Следствие 1 (знак отражения). Отражающий шарнир с коэффициентом b_j вносит вклад $-b_j$ в ставку на каждом отрезке слева от t_j и 0 — справа. В примере $b_1 = -600$ в точке $t_1 = 1,200$ вносит вклад $+600$ в r_0 , что и составляет суть ловушки, вокруг которой построено данное дополнение.

Доказательство. Воспользуйтесь второй суммой из теоремы 1. ■

Пример с семью узлами и допущение, на котором читатели останавливаются

Это значение имеет важное значение из-за предположения, которое почти все делают при работе с печатной моделью MARS: что коэффициент, напечатанный рядом с узлом, является наклоном отрезка, левым узлом которого он является. На самом деле это не так. **Коэффициент шарнира представляет собой корректировку к наклону, идущему слева, а не сам наклон**, и поэтому наклон отрезка равен наклону первого отрезка плюс скользящая сумма всех таких корректировок в его левом узле и в каждом узле, расположенном дальше слева. Читатель, исходящий из этого ошибочного предположения, пробует выполнить вычисления в уме, с помощью небольшой программы или формулы в электронной таблице, обнаруживает, что полученные числа не совпадают с результатами подгонки, и сдаётся, придя к выводу, что MARS — непрозрачная система. На самом деле она не непрозрачна; она носит кумулятивный характер.

Теорема 2 (накопительная сумма). Используя обозначения из постановки задачи «Случай 1», для каждого $i \in \{1, \dots, K\}$ выполняется:

$$r_i = r_{i-1} + a_i + b_i.$$

Следовательно, $r_i = r_0 + \sum_{j=1}^i (a_j + b_j)$ для каждого i , где $r_0 = c - \sum_{j=1}^K b_j$ — ставка на крайнем левом отрезке, а $r_K = c + \sum_{j=1}^K a_j$ — ставка на крайнем правом.

Доказательство теоремы 2. Пусть $i \in \{1, \dots, K\}$. Согласно теореме 1,

$$r_i = c + \sum_{j=1}^i a_j - \sum_{j=i+1}^K b_j \quad \text{и} \quad r_{i-1} = c + \sum_{j=1}^{i-1} a_j - \sum_{j=i}^K b_j.$$

При вычитании член c отменяется; две суммы членов a различаются одним членом a_i ; а поскольку $\sum_{j=i}^K b_j = b_i + \sum_{j=i+1}^K b_j$, две суммы членов b различаются одним членом b_i , который вычитается в выражении r_{i-1} , но не в выражении r_i . Следовательно, $r_i - r_{i-1} = a_i + b_i$. Добавляя это тождество для $j = 1, \dots, i$ вместо i , левые члены сжимаются: $(r_1 - r_0) + (r_2 - r_1) + \dots + (r_i - r_{i-1}) = r_i - r_0$, что даёт формулу суммы. Значения r_0 и r_K являются Теоремой 1 на $i = 0$, где сумма a 's пуста, и на $i = K$, где сумма b 's пуста. ■

Следствие 2 (правило считывания). В условиях случая 1 для каждого $i \in \{1, \dots, K\}$ ставка на отрезке, левый узел которого — t_i , равна $r_i = r_{i-1} + a_i + b_i$: коэффициенты, указанные в узле, прибавляются к ставке, поступающей слева, и сами по себе не являются ставкой отрезка, начинающегося в этом узле. Эти два значения совпадают, $r_i = a_i + b_i$, ровно тогда, когда $r_{i-1} = 0$; в частности, для модели, в которой все шарниры являются правыми и которая не содержит линейного члена, это происходит в точке $i = 1$ и, согласно теореме 2, ни в одном более позднем узле, если только сумма коэффициентов в узлах, предшествующих этому узлу, не равна нулю.

Доказательство. Первое равенство является теоремой 2. Из него следует, что $r_i = a_i + b_i$ тогда и только тогда, когда $r_{i-1} = 0$. Если $b_j = 0$ для всех j и $c = 0$, то из теоремы 2 следует, что $r_0 = 0$, следовательно, $r_1 = a_1$, и $r_{i-1} = \sum_{j<i} a_j$, что равно 0 только в том случае, когда сумма этих коэффициентов равна нулю. ■

Следствие 3 (корректировка как интеграл). В условиях случая 1 функция g является непрерывной в каждом узле, поскольку непрерывны все шарниры, и, следовательно, непрерывны также функции $f(\cdot, z)$ для каждого z ; именно это обуславливает одинаковость значения в узле независимо от того, достигается ли оно слева или справа, и позволяет переходить от одной величины к другой. Пусть x_c и x_s — величины x для аналога и объекта оценки, при этом остальные характеристики для обоих

объектов фиксированы на одинаковых значениях, так что $g(x)$, члены выражения f , содержащие x , представляют собой вклад x при величине x . Обозначим $u_1 = \min(x_c, x_s)$ и $u_2 = \max(x_c, x_s)$, а также $|I|$ — длину интервала I . Тогда корректировка $g(x_s) - g(x_c)$, величина, прибавляемая к вкладу аналога, чтобы приравнять его к вкладу объекта оценки, равна

$$g(x_s) - g(x_c) = \int_{x_c}^{x_s} g'(u) du = \text{sign}(x_s - x_c) \sum_{i=0}^K r_i \cdot |S_i \cap [u_1, u_2]|.$$

Другими словами: корректировка представляет собой интеграл ставки между двумя величинами, и поскольку ставка постоянна на каждом отрезке, интеграл равен сумме, по отрезкам, встречающимся на пути, ставки, умноженной на ширину.

Доказательство. Функция g является непрерывной и кусочно-линейной с конечным числом кусков, следовательно, абсолютно непрерывной, поэтому согласно основной теореме математического анализа имеем $g(x_s) - g(x_c) = \int_{x_c}^{x_s} g'(u) du$, причём производная существует везде, кроме узлов, которые составляют конечное множество. На каждом S_i по теореме 1 подынтегральное выражение равно r_i , поэтому $\int_{u_1}^{u_2} g'(u) du$ представляет собой сумму по i выражения r_i , умноженного на длину $S_i \cap [u_1, u_2]$, причем непересекающиеся отрезки вносят вклад длиной 0; а $\int_{x_c}^{x_s} = \text{sign}(x_s - x_c) \int_{u_1}^{u_2}$ — в силу ориентации интеграла. ■

Таблица ставок приведена по адресу g' ; корректировка, вычисляемая непосредственно по формуле $g(x_s) - g(x_c)$, равна интегралу от этой величины между двумя значениями; а теоремы 1 и 2 определяют, из чего состоит g' . На практике корректировка всегда вычисляется как разность вкладов, что не требует использования коэффициентов вообще; следствие объясняет, почему коэффициенты, рассматриваемые по диапазонам, воспроизводят эту разность, что и делает их объяснением корректировки и средством её проверки, а не способом её вычисления.

Рассмотрим набор данных о жилой площади, в котором самый маленький дом имеет площадь 800 квадратных футов и семь узлов, расположенных на отметках 1 200, 1 600, 2 000, 2 400, 2 800, 3 200 и 3 600 квадратных футов, а также девять членов в модели, два из которых отражают шарниры, каждый из которых сопряжен с правым шарниром в том же узле: при 1 200 и при 2 400:

h(1200-gla)	-500
h(gla-1200)	700
h(gla-1600)	-100
h(gla-2000)	150
h(2400-gla)	-200
h(gla-2400)	-250
h(gla-2800)	-300
h(gla-3200)	-150
h(gla-3600)	50

Семь узлов образуют восемь отрезков. Ставка на каждом из них равна сумме коэффициентов активных правых шарниров минус сумма коэффициентов активных зеркальных шарниров:

сегмент (sf)	активные термины	стоимость (\$/sf)
от 800 до 1 200	$h(1200 - gla), h(2400 - gla)$	$500 + 200 = 700$
от 1 200 до 1 600	$h(gla - 1200), h(2400 - gla)$	$700 + 200 = 900$
от 1 600 до 2 000	правые петли на отметках 1 200, 1 600; $h(2400 - gla)$	$700 - 100 + 200 = 800$
от 2 000 до 2 400	правые петли — от 1 200 до 2 000; $h(2400 - gla)$	$700 - 100 + 150 + 200 = 950$
от 2 400 до 2 800	правые петли от 1 200 до 2 400	$700 - 100 + 150 - 250 = 500$
от 2 800 до 3 200	правые петли — от 1 200 до 2 800	$500 - 300 = 200$
от 3 200 до 3 600	правые петли от 1 200 до 3 200	$200 - 150 = 50$
свыше 3 600	все семь правых петель	$50 + 50 = 100$

Три величины, которые нужно считать со стола. Ставка крайнего правого сегмента, $700 - 100 + 150 - 250 - 300 - 150 + 50 = 100$, объединяет коэффициенты всех семи правых шарниров, по одному на каждый узел, что является границей теоремы 1 с $K = 7$; два отраженных шарнира здесь не играют никакой роли, поскольку отраженный шарнир является плоским справа от своего собственного узла, а крайний правый сегмент лежит справа от каждого узла. Самый левый отрезок, от 800 до его правого узла на отметке 1 200, является зеркальным отражением: учитываются только отраженные шарниры с обратными знаками, поскольку каждый узел находится на уровне своего правого узла или правее него. И в каждом узле, несущем пару, шаг равен сумме двух коэффициентов: в точке 1 200 — $-500 + 700 = +200$, от 700 до 900; на отметке 2 400 ставка падает с 950 до 500, изменение составляет -450 , что точно равно сумме двух коэффициентов в этом узле: -250 и -200 : пересечение узла включает правый шарнир и выключает отраженный шарнир, и оба изменения действуют в одном направлении.

$+200, -100, +150, -450, -300, -150, +50$. Этот маршрут затрагивает все восемь сегментов и все девять термов и заканчивается на том же 100. Прямая сумма короче и именно её выполняет программа; прохождение — это то, что читатель должен сделать один раз вручную, чтобы убедиться, что коэффициент рядом с узлом — это шаг, а не уровень. « -450 » на отметке 2 400 состоит из двух полушагов: пересечение узла оставляет позади отраженный шарнир $h(2400 - gla)$, а вместе с ним и $+\$200$, который он добавил ниже узла; при этом он вводит правый шарнир $h(gla - 2400)$, чей -250 начинается именно там. $-200 - 250 = -450$, и ставка падает с $\$950$ до $\$500$.

На рисунке 4 приведён аналогичный пример для модели с пятью узлами, размер которой достаточно мал, чтобы оценить результат на глаз. Полоски — это коэффициенты, напечатанные на узлах; лестница — ставка на каждом отрезке; лестница начинается со ставки первого отрезка, $\$900$, заданной отражающим шарниром, и каждая ступенька представляет собой сумму полосок на данном узле. Ставка на крайнем правом сегменте, $650 - 250 + 150 - 200 - 150 = 200$, использует все пять правых шарниров; отраженный шарнир на отметке 1 200, обозначенный как -900 , устанавливает для первого сегмента значение $+900$ и ни на что больше не влияет.

Стоит отметить две особенности формы этого графика, поскольку они касаются именно домов, а не MARS. Во-первых, ось начинается с отметки 800 квадратных футов, а не с нуля, и график начинается с этой оси: каждый график модели, подвергающейся подгонке, в данном приложении, как и графики, построенные программным обеспечением для моделирования, простирается от самого маленького дома в наборе данных до самого большого и не отображает ничего за пределами этого диапазона. В каждой рыночной зоне есть самый маленький дом, ниже которого продаж не происходит, и график переменной ничего не показывает ниже этого уровня: не ноль, а отсутствие значений, подтверждаемых данными о продажах; формула все равно выдавала бы там число, и это число является экстраполяцией. На рисунках используется минимальный порог в 800 квадратных футов, но этот порог является характеристикой рыночной зоны, а не константой: в разных рыночных зонах минимальные размеры объектов недвижимости, отображаемые в системе MLS, различаются, и именно набор данных, а не оценщики, определяют, где он находится. Во-вторых, форма кривой варьируется от рынка к

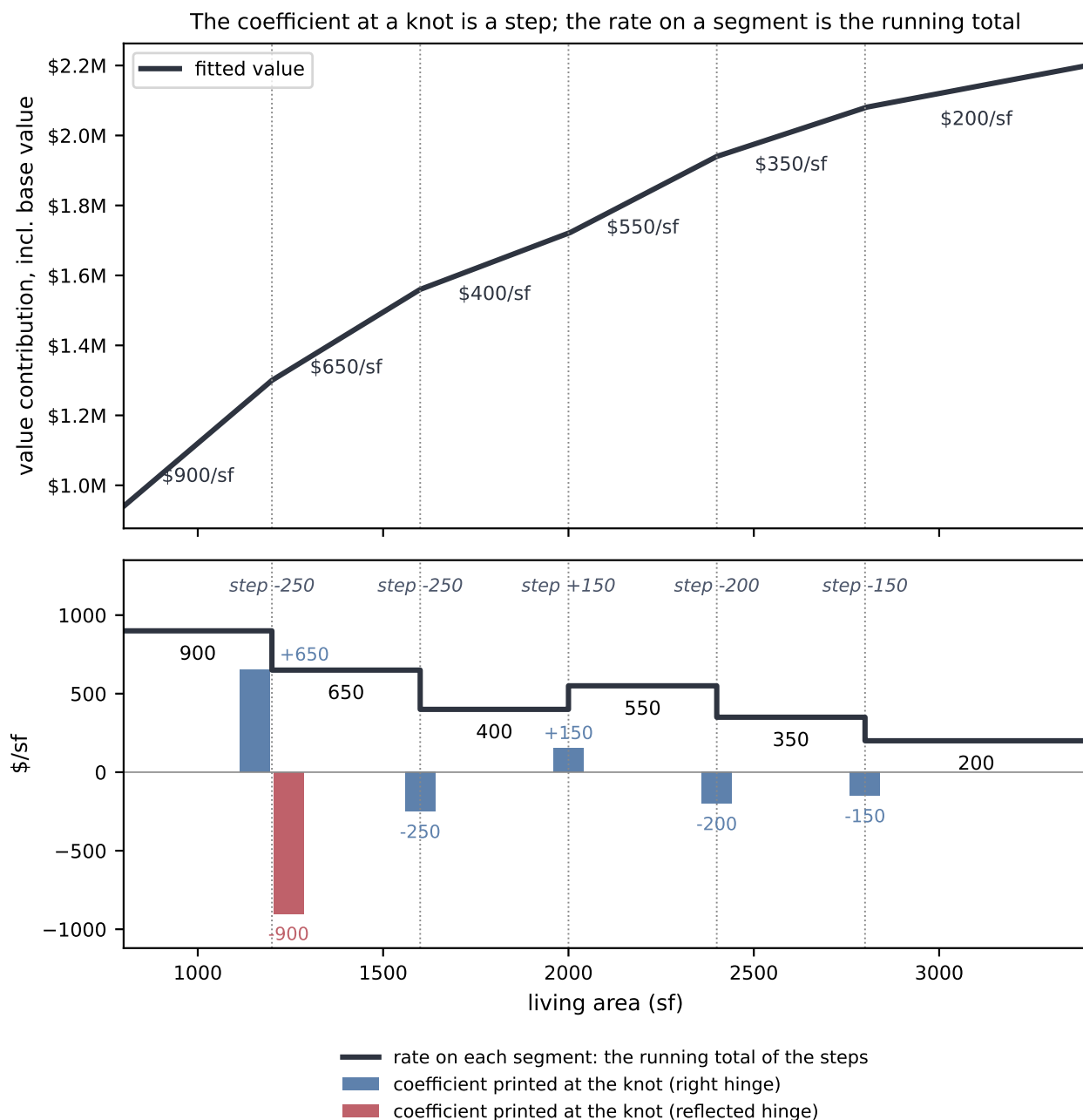


Figure 4: Модель с пятью узлами, построенная на данных, начинающихся с 800 квадратных футов: $-900 h(1200 - gla) + 650 h(gla - 1200) - 250 h(gla - 1600) + 150 h(gla - 2000) - 200 h(gla - 2400) - 150 h(gla - 2800)$. Вверху — результат подгонки со ставкой, указанной на каждом отрезке; значений ниже 800 квадратных футов нет, поскольку таких домов не существует. Ниже — коэффициент, отображённый в виде столбца на каждом узле, и темп роста на каждом отрезке в виде лестницы: каждая ступенька лестницы представляет собой сумму столбцов на данном узле (один столбец здесь, два — там, где узел содержит пару, Теорема 2), поэтому ставка на отрезке равна ставке первого отрезка плюс накопительная сумма ступеней слева от него, а не коэффициенту рядом с ним: каждый коэффициент проводит корректировку ставки, поступающей слева.

рынку: в примере с прибрежными районами наиболее крутой диапазон составляет от 1 200 до 2 000 квадратных футов, а на рисунке4 это первый сегмент. В районах с высокими ценами, таких как Атертон или Хиллсборо, самый крутой отрезок кривой жилой площади (GLA) может пролегать за отметкой 4 000 или 5 000 квадратных футов — именно там находятся покупатели. А в районе, где застройщик в определенный период возводил дома одного размера, имевшие проблемы, может наблюдаться резкое падение ставки в промежуточном диапазоне, как показано на построенном графике в диапазоне от 1 600 до 2 000. На более обширной рыночной зоне графики ставок варьируются, и таблица ставок — это то, как оценщик интерпретирует данные конкретного графика. На рисунках 5 и 6 изображены две из таких кривых в том виде, в каком они были бы напечатаны в журнале «Earth», чтобы читатель мог составить каждую таблицу ставок по правилам и сверить её с лестницей.

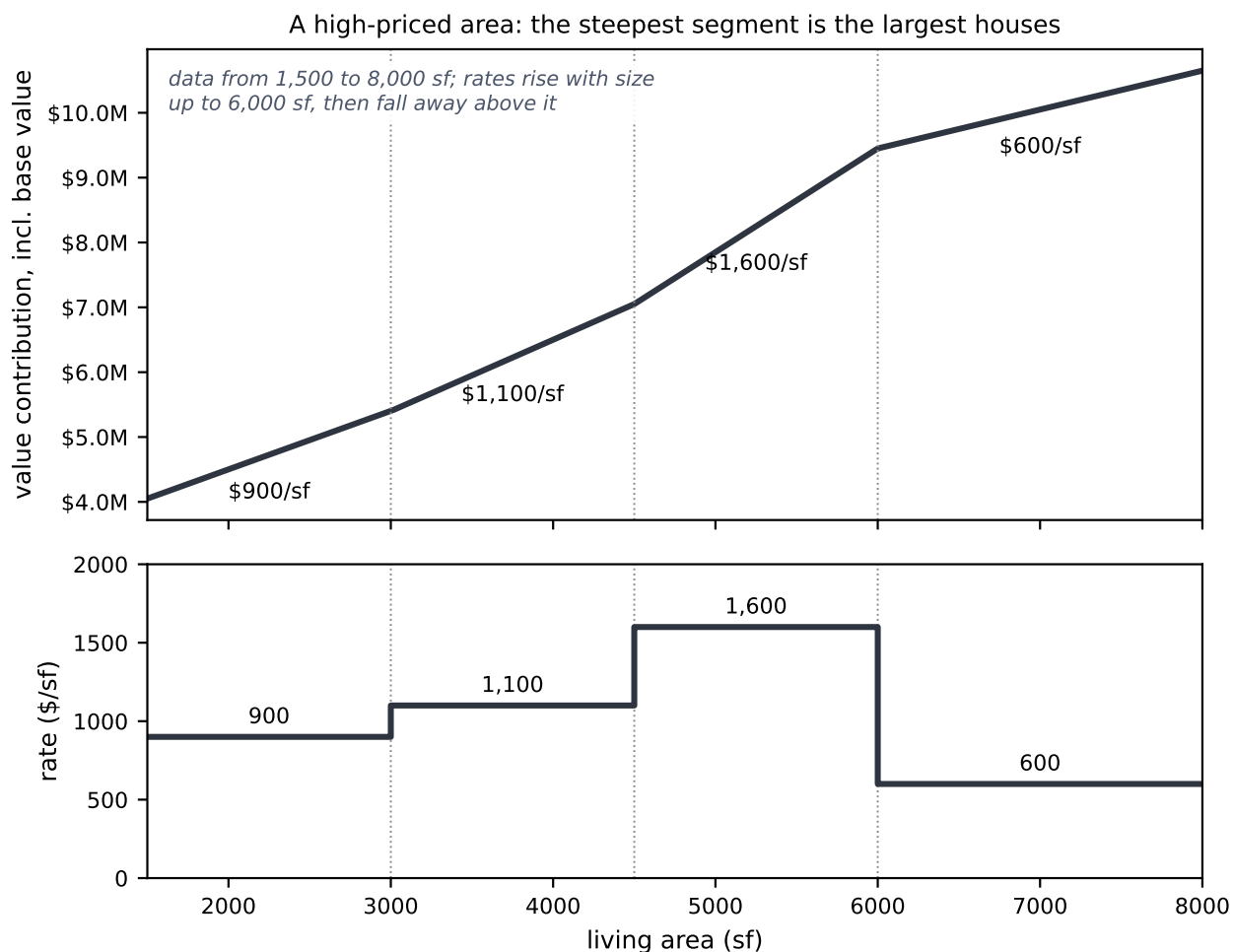


Figure 5: Район с высокой стоимостью недвижимости, площадью от 1 500 до 8 000 квадратных футов: $-900h(3000 - gla) + 1100h(gla - 3000) + 500h(gla - 4500) - 1000h(gla - 6000)$, пересечение с \$5,400,000 на первом узле. Цены \$900 , \$1,100 , \$1,600 и \$600 за квадратный фут: самый крутой участок — от 4 500 до 6 000 квадратных футов, где сосредоточены покупатели; первый участок более пологий, а выше 6 000 цена падает. Правило остаётся прежним; меняется только рынок.

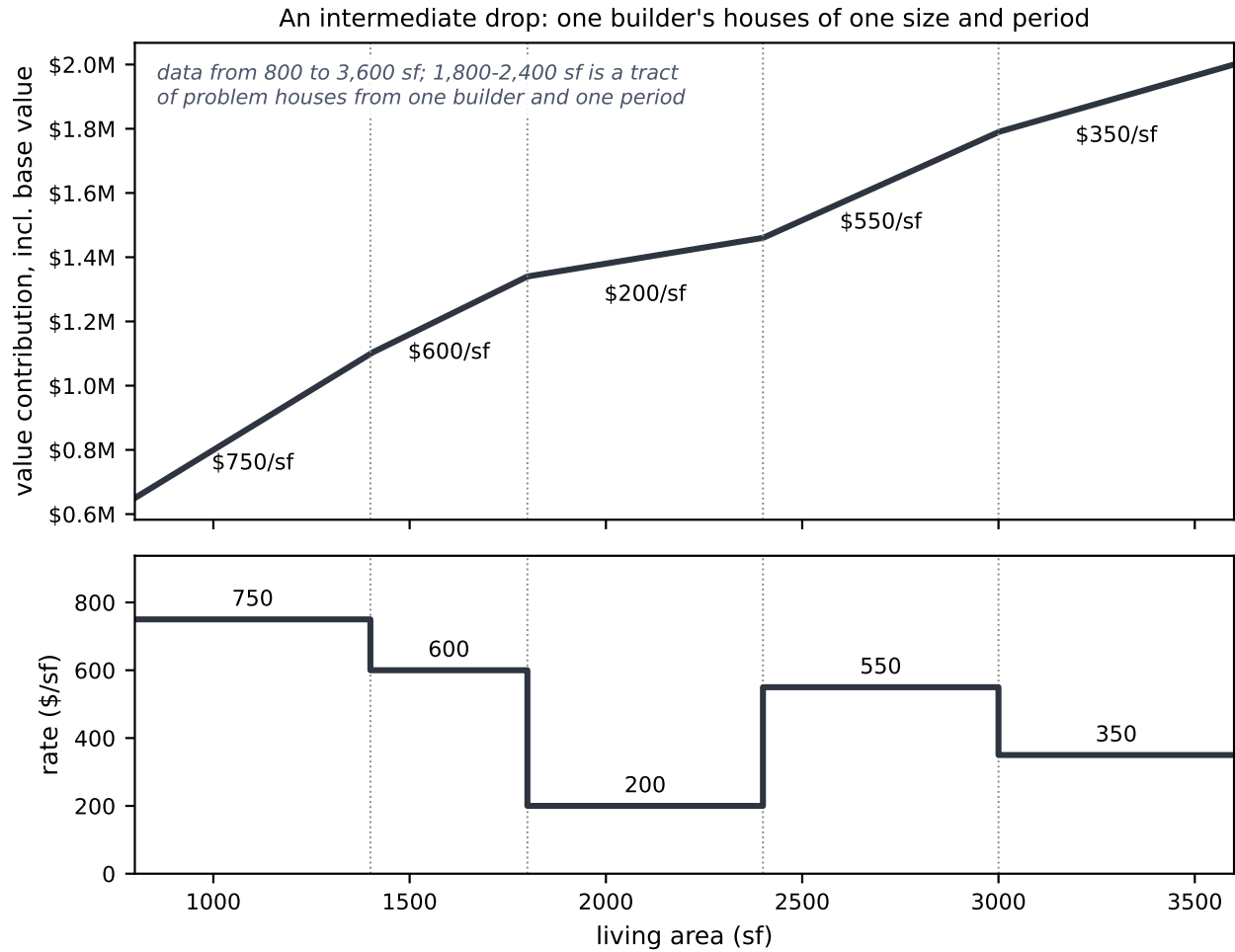


Figure 6: Промежуточное значение, по данным от 800 до 3 600 квадратных футов: $-750 h(1400 - gla) + 600 h(gla - 1400) - 400 h(gla - 1800) + 350 h(gla - 2400) - 200 h(gla - 3000)$, пересечение с \$1,100,000 в первом узле. Ставки \$750, \$600, \$200, \$550 и \$350 за квадратный фут. Предположим, что сегмент от 1 800 до 2 400 представляет собой участок с домами такого размера, построенными одним застройщиком в один период, с учетом рыночных цен (конструктивная иллюстрация); ставка падает до \$200 в этой точке и восстанавливается выше нее. Читатель, который принимает коэффициент -400 равным 1 800 для цены этого сегмента или предполагает, что цены падают только с увеличением размера, неправильно интерпретирует этот участок.

Случай 2: взаимодействия

Термин взаимодействия представляет собой коэффициент, умноженный на произведение факторов, выраженных разными переменными, причем каждый фактор является либо «шарниром», либо самой переменной, например: $\gamma \cdot h(\text{gla} - 1500) \cdot \text{lot}$. Всё, что было в случае 1, остаётся неизменным, за исключением одного слова: коэффициент на отрезке теперь является функцией других переменных, а не константой. Далее следует утверждение и его доказательство, а затем — объяснение того, как это влияет на вычисления оценщика.

Настройка. Модель MARS с взаимодействиями по признакам $x, z_1, \dots, z_q$ имеет следующее значение:

$$f(x, z) = \beta_0 + \sum_{m=1}^M \gamma_m \prod_{k \in F_m} B_{m,k}(z_k),$$

где M — количество слагаемых, кроме пересечения, β_0 — пересечение, а модель представляет собой пересечение плюс сумму слагаемых, причем слагаемое m имеет вид $\gamma_m \prod_{k \in F_m} B_{m,k}(z_k)$: его коэффициент γ_m умножен на один фактор для каждого признака в данном слагаемом. Здесь $z = (z_1, \dots, z_q)$, и, чтобы один индекс пробежал по каждой характеристике, z_0 обозначает x . Характеристики, входящие в член m , образуют множество $F_m \subseteq \{0, 1, \dots, q\}$; ни одна характеристика не встречается дважды в одном члене, и каждый фактор $B_{m,k}$ является функцией одной характеристики z_k , относящейся к одному из трёх типов: сама характеристика z_k ; правый шарнир $h(z_k - t)$; или отражённый шарнир $h(t - z_k)$. Пусть $M_x = \{m : 0 \in F_m\}$ — это термины, содержащие x . Для $m \in M_x$, $B_{m,0}(x)$ является фактором термина в x , а другие черты термина, те z_k , у которых $k \in F_m, k \neq 0$, являются *партнерами* x в данном термине. Запишите

$$P_m(z) = \prod_{k \in F_m, k \neq 0} B_{m,k}(z_k)$$

для произведения партнерских множителей, причем пустое произведение равно 1. Другими словами: как только другие характеристики объекта оценки заполнены, партнеры термина умножаются на одно обычное число; это число, умноженное на напечатанный коэффициент, является коэффициентом простого термина жилой площади, и термин читается как в случае 1. Произведение партнёров $P_m(z)$ — это то, что остаётся от члена m после вынесения его множителя в x : член m равен $\gamma_m B_{m,0}(x) P_m(z)$. В данной точке z это число, поэтому на отрезке член действует на x как единственный шарнир с коэффициентом $\gamma_m P_m(z)$, и именно это произведение определяет, как партнёры входят в ставку. Член без партнёра имеет $P_m(z) = 1$ и является членом случая 1.

Например, в выражении $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ из приведенного ниже примера, где x — жилая площадь, коэффициент в выражении x равен $h(\text{gla} - 1200)$, один из партнёров — участок, а $P_m(z) = h(\text{lot} - 6000)$. При участке площадью 5 000 квадратных футов $P_m(z) = 0$ этот член не влияет на ставку; при участке площадью 10 000 квадратных футов $P_m(z) = 4,000$ этот член действует как поворотный момент при значении 1 200 с коэффициентом $0.03 \times 4,000 = 120$, добавляя \$120 за квадратный фут к ставке при площади свыше 1 200 квадратных футов. Здесь ничто не ограничивает количество элементов, которые может содержать терм; аргумент *earth's degree* устанавливает этот максимум, и приведенные ниже теоремы справедливы для любой степени. Терм с d элементами имеет граф в $d+1$ измерениях, который никто не может нарисовать, как только d превышает два; программа показывает только срезы — по одному или двум элементам за раз, при этом остальные фиксируются, а Теорема 3 представляет собой тот же срез, выполненный в точке объекта оценки. $P_m(z)$ — это место, где происходит этот срез: сколько бы партнёров ни имел терм, в какой-то точке их произведение равно одному числу. В члене с четырьмя характеристиками $\gamma h(\text{gla} - 1200) h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$,

где $\gamma = 0.001$, график которого имеет пять измерений, $P_m(z) = h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$; при участке площадью 10 000 квадратных футов, возрасте 20 лет и 3 ванных комнатах получается $4,000 \times 10 \times 3 = 120,000$, и по жилой площади весь член представляет собой шарнир при значении 1 200 с коэффициентом $\gamma \times 120,000 = 0.001 \times 120,000 = 120$, добавляя \$120 за квадратный фут: то же самое, что и в случае с двухкомпонентным членом выше, но с коэффициентом, в тридцать раз меньшим. Именно поэтому коэффициент члена высокой степени часто отображается в виде очень маленького числа, когда его слагаемые измеряются в больших единицах: он умножает произведение нескольких значений слагаемых. А если какой-либо из компонентов находится в нулевой области своего узла — например, площадь значительно меньше 6 000 квадратных футов или возраст меньше 10 лет, — то $P_m(z) = 0$ и этот член не вносит никакого вклада, какими бы ни были остальные компоненты; чем выше степень, тем больше способов, которыми данный член может быть «отключен» при заданных объектах недвижимости. В примерах, приведенных в данном приложении, используются двусторонние взаимодействия ($\text{degree} = 2$).

Пусть $t_1 < \dots < t_K$ — это различные узлы из множества $B_{m,0}$, являющиеся шарнирами, при этом $t_0 = -\infty$, $t_{K+1} = +\infty$ и отрезки $S_i = (t_i, t_{i+1})$ — такие же, как в случае 1. Для $m \in M_x$ и $i \in \{0, \dots, K\}$ определим

$$s_{m,i} = \begin{cases} +1 & \text{если } B_{m,0} \text{ — сам признак } x \text{ или правый шарнир } h(x - t_j) \text{ при } t_j \leq t_i, \\ -1 & \text{если } B_{m,0} \text{ — отражённый шарнир } h(t_j - x) \text{ при } t_j \geq t_{i+1}, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Точка — это значение z других характеристик. Поскольку каждая функция $B_{m,0}$ является линейной или равна нулю на S_i , функция f является линейной в x на S_i при фиксированном z ; её наклон в этой точке равен *ставки* изменения S_i в точке z , что записывается как $r_i(z) = \partial f / \partial x$ для $x \in S_i$.

Чтобы увидеть все символы сразу, воспользуйтесь примером, приведённым ниже, в том виде, в каком он отображается на сайте Earth: пересечение 900 000, $600 h(\text{gla} - 1200)$, $-350 h(\text{gla} - 2400)$, $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ и 15 lot. Его характеристиками являются жилая площадь, $x = z_0 = \text{gla}$, и площадь участка, $z_1 = \text{lot}$, поэтому $q = 1$; пересечение равно $\beta_0 = 900,000$, а членов уравнения $M = 4$. Для объекта недвижимости с жилой площадью 2 000 квадратных футов на участке площадью 10 000 квадратных футов:

m	γ_m	F_m	$B_{m,0}(x)$	$B_{m,1}(z_1)$	$P_m(z)$	член для объекта
1	600	{0}	$h(\text{gla} - 1200)$	нет	1	$600 \times 800 = 480,000$
2	-350	{0}	$h(\text{gla} - 2400)$	нет	1	$-350 \times 0 = 0$
3	0.03	{0, 1}	$h(\text{gla} - 1200)$	$h(\text{lot} - 6000)$	$h(\text{lot} - 6000)$	$0.03 \times 800 \times 4,000 = 96,000$
4	15	{1}	нет	lot	(не входит в M_x)	$15 \times 10,000 = 150,000$

Итак, $M_x = \{1, 2, 3\}$: член 4 не содержит коэффициента в x , добавляет одинаковую величину независимо от жилой площади и не входит в коэффициент x . Собственное значение равняется $f = 900,000 + 480,000 + 0 + 96,000 + 150,000 = \$1,626,000$. Члены 1 и 2 не имеют партнеров, поэтому $P_1(z) = P_2(z) = 1$; член 3 имеет одного партнера, и $P_3(z) = h(10,000 - 6,000) = 4,000$ — это множитель из приведенного выше примера. Теорема 3, которая следует из этого, преобразует их в коэффициент. Жилая площадь имеет узлы $t_1 = 1,200$ и $t_2 = 2,400$, а недвижимость находится на отрезке S_1 между ними, где активны члены 1 и 3, а член 2 — нет ($s_{1,1} = s_{3,1} = 1$, $s_{2,1} = 0$); следовательно, $r_1(z) = 600 \times 1 + 0.03 \times 4,000 = 720$ — это ставка для участка площадью 10 000 квадратных футов из таблицы в примере.

Теорема 3 (ставка на отрезке с учетом взаимодействий). При использовании данной нотации для каждого i и каждой точки z выполняется равенство:

$$r_i(z) = \sum_{m \in M_x} \gamma_m s_{m,i} P_m(z).$$

Доказательство теоремы 3. Зафиксируем i , точку z и $x \in S_i$. Член с $m \notin M_x$ не содержит множителя x и является постоянным в x ; его производная в x равна 0. Член с $m \in M_x$ равен $\gamma_m B_{m,0}(x) P_m(z)$, причём $P_m(z)$ не зависит от x . Согласно анализу случаев в доказательстве теоремы 1, применённому к отдельному множителю $B_{m,0}$, производная $B_{m,0}$ в точке x равна 1 для самой характеристики, 1 или 0 для правого шарнира в зависимости от того, находится ли его узел в точке t_i или левее неё, либо в точке t_{i+1} или правее неё, и -1 или 0 для отражённого шарнира в зависимости от того, находится ли его узел в точке t_{i+1} или правее неё, либо в точке t_i или левее неё; то есть оно равно $s_{m,i}$. Таким образом, производная этого слагаемого равна $\gamma_m s_{m,i} P_m(z)$, а производная f , конечной суммы, равна сумме этих величин. Значение не зависит от того, какой из вариантов $x \in S_i$ был выбран. ■

Теорема 1 представляет собой частный случай, в котором ни один член выражения M_x не имеет партнёра, так что для каждого $P_m(z) = 1$; группировка членов суммы по их узлам даёт формулу, доказанную в данной теореме. Три следствия описывают, как меняются партнёры; последнее из них — невыполнение ограничения «один на узел» и случай, когда «член может исчезать», что иллюстрирует приведённый ниже пример.

Следствие 4 (ставка является числом только в одной точке). В рамках случая 2 для каждого i ставка r_i является функцией точки z . Она постоянна в z всякий раз, когда каждый член $m \in M_x$ с $s_{m,i} \neq 0$ не имеет партнёра. Когда какой-либо член с $s_{m,i} \neq 0$ имеет партнёра, $r_i(z)$ является функцией $z \mapsto \sum_m \gamma_m s_{m,i} P_m(z)$, которая изменяется в зависимости от значений партнёров, за исключением случаев, когда члены точно компенсируются; два свойства с одинаковым x на одном и том же отрезке имеют разные ставки x , если их партнёрские произведения различаются и эти различия не компенсируются. Именно поэтому программа принимает точку и именно поэтому не существует отдельной таблицы ставок изменения для одной характеристики.

Доказательство. Согласно теореме 3, $r_i(z) = \sum_{m \in M_x} \gamma_m s_{m,i} P_m(z)$. Если каждый член с $s_{m,i} \neq 0$ не имеет партнёра, то каждый такой P_m представляет собой пустое произведение 1, и сумма является постоянной величиной. В противном случае сумма содержит по крайней мере один счётный член $\gamma_m s_{m,i} P_m(z)$ с $\gamma_m s_{m,i} \neq 0$ и P_m — произведением одного или нескольких множителей, каждый из которых является неконстантной функцией партнёра, поэтому сумма является постоянной по z только в том случае, если эти счётные члены полностью компенсируют друг друга. ■

Следствие 5 (на отрезке x ставка представляет собой переменный график для каждого партнера). В условиях случая 2 зафиксируем i , партнера z_p ($p \geq 1$) и значения всех остальных характеристик. Тогда $z_p \mapsto r_i(z)$ является непрерывной кусочно-линейной функцией z_p , разрывы которой лежат среди узлов факторов $B_{m,p}$ слагаемых в M_x . Вдоль x при фиксированных значениях каждого партнера ставка изменяется только в узлах $t_1, \dots, t_K$ x ; именно в совместном пространстве x и его партнеров поверхность ставки пересекается как узлами x , так и узлами каждого партнера.

Доказательство. При фиксированных остальных характеристиках каждая функция $P_m(z)$ представляет собой постоянную, умноженную либо на 1 (если $p \notin F_m$), либо на единственный множитель $B_{m,p}(z_p)$, который сам по себе является либо z_p , либо $h(z_p - t)$, либо $h(t - z_p)$: непрерывная кусочно-линейная функция от z_p с не более чем одним переходом в точке узла. Конечное линейное сочетание таких функций, которое, согласно теореме 3, получается для $r_i(z)$, является непрерывным и кусочно-линейным с разрывами только там, где разрывается какой-либо слагаемый. Утверждение, приведенное на странице

x , и есть сама теорема 3: для фиксированного z , $r_i(z)$ представляет собой одно число на каждом S_i , и отрезки пересекаются только в узлах x . ■

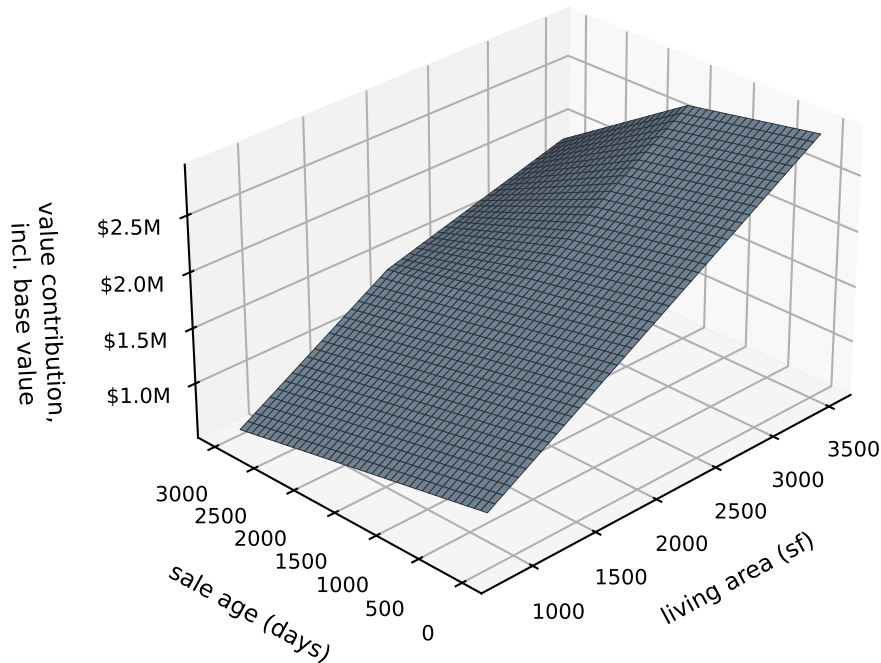
Следствие 6 (ограничение из теоремы 1 не выполняется). В условиях случая 2 число слагаемых с ненулевым счётом в формуле теоремы 3 для $r_i(z)$ ограничено лишь $|M_x|$ — числом слагаемых, содержащих x , — и может превышать K : узел вида x может появляться в нескольких слагаемых вида M_x с разными партнёрами, каждый из которых является отдельным слагаемым. И наоборот, терм с $s_{m,i} \neq 0$ не вносит никакого вклада в точку, где произведение его партнёров равно нулю.

Доказательство. Члены $\gamma_1 h(x-1500) \text{lot}$ и $\gamma_2 h(x-1500) h(\text{age}-10)$ являются различными членами выражения M_x с одним и тем же узлом, и на отрезке выше 1,500 оба имеют $s_{m,i} = 1$, поэтому оба вносят вклад, а выражение $r_i(z) = \gamma_1 \text{lot} + \gamma_2 h(\text{age}-10)$ содержит два слагаемых из одного узла; добавление дальнейших партнёров приводит к появлению новых членов в том же узле без ограничений. Для обратного случая возьмем член, партнером которого является шарнир $h(\text{lot}-6000)$ в точке с $\text{lot} < 6000$: тогда $P_m(z) = 0$ и слагаемое исчезает, хотя $s_{m,i} \neq 0$; в точке с $\text{lot} > 6000$ этого не происходит. ■

Здесь уместно привести практическое наблюдение, основанное на собственных оценках редактора. Как только допускаются взаимодействия, график жилой площади, который программа строит для переменной самостоятельно, часто получается прямой — с одной ставкой, в то время как изгибы, которые модель без взаимодействий внесла бы в жилую площадь, вместо этого появляются в слагаемых взаимодействия, как правило, в соотношении жилой площади и возраста продажи (времени между продажей и датой оценки) или жилой площади и количества ванных комнат. Теорема 3 объясняет, почему прямая линия не является коэффициентом. График по каждой переменной в earthUI отображает слагаемые без партнеров; график первой степени в plotmo, напротив, отображает срез всей модели, где другие переменные находятся на своих медианах. Слоги взаимодействия в других местах отображаются в виде поверхностей (графики первой и второй степени в plotmo, Milbortrow 2024b, которые earthUI показывает рядом). Для любого конкретного объекта недвижимости ставка за жилую площадь равна наклону прямой плюс, для каждого активного термина взаимодействия, его коэффициент, умноженный на значение его партнерского фактора; она изменяется в узлах жилой площади этих членов, и насколько — зависит от партнёров: продажа шесть лет назад и продажа в прошлом месяце, или дом с двумя ванными комнатами и дом с четырьмя — находятся в разных таблицах ставок по жилой площади, хотя на графике жилой площади отображается одна линия. Читатель, который принимает эту линию за коэффициент, снова сталкивается с препятствием, на один уровень выше; таблица коэффициентов в точке, являющейся объектом оценки, — это путь к преодолению этого препятствия. На рисунке 7 изображена одна из таких моделей так, как это делает программа, в виде поверхности, зависящей от жилой площади и срока с момента продажи, а затем с неё считывается график переменной «жилая площадь» для двух значений срока с момента продажи.

На рисунке 8 эта же идея реализована в обратном направлении, исходя из рынка, представленного на рисунке 2. Для этого рынка построена модель, в которой взаимодействуют переменные «жилая площадь» и «возраст объекта», так что основной эффект «жилой площади» представляет собой одну прямую линию, \$500 за квадратный фут, а четыре члена взаимодействия определяют изгибы кривой. График, который программа строит самостоятельно для жилой площади, представляет собой прямую линию. Если рассматривать продажи четырехлетней давности, показатели жилой площади соответствуют данным на рисунке 2 в долларах: \$600, \$1,000, \$750 и \$100; при рассмотрении новых продаж получается прямая линия; при рассмотрении продаж двухлетней давности показатели находятся между этими значениями. Модель построена для иллюстрации идеи, а не подгоняна под данные продаж, и реальная перестройка не привела бы к столь аккуратному результату, как на рисунке 2; однако механизм тот же, что и в реальной перестройке, и именно поэтому оценщик, который видит, как график жилой площади выпрямляется после добавления взаимодействий, не теряет из виду изгибы. Изгибы переместились

Living area by sale age: the surface the software draws



The living-area variable graph, read at two points

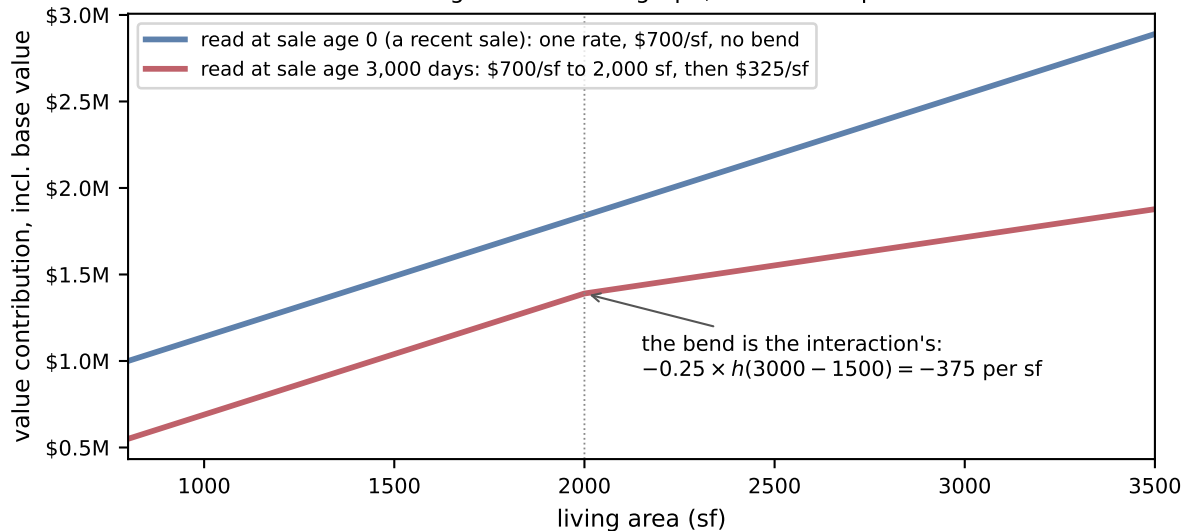


Figure 7: Взаимосвязь между жилой площадью и сроком выставления на продажу: $1,000,000 + 700 h(\text{gla} - 800) - 150 \text{saleage} - 0.25 h(\text{gla} - 2000) h(\text{saleage} - 1500)$, на основе данных о площадях от 800 до 3 500 квадратных футов и сроках выставления на продажу до 3 000 дней. Выше — график, построенный программным обеспечением для моделирования данной пары параметров. Ниже — график переменной жилой площади, отчитанный в двух точках, как того требует Теорема 3: при любой продаже, состоявшейся не более 1 500 дней назад, коэффициент партнера $h(\text{saleage} - 1500)$ равен нулю, взаимодействие не вносит вклада, и график представляет собой одну прямую линию с наклоном \$700 на квадратный фут; при сделке, заключенной 3 000 дней назад, значение фактора «партнер» равно 1 500, взаимодействие вычитает $0.25 \times 1,500 = 375$ за квадратный фут при площади свыше 2 000 квадратных футов, и ставка в этом случае составляет \$325. Основной эффект жилой площади (члены, содержащие только жилую площадь, без партнера) представляет собой одну и ту же прямую линию в обоих случаях; изгиб обусловлен взаимодействием и появляется только там, где партнер играет активную роль.

в члены взаимодействия, и восстановить их можно только по значению в конкретной точке. График основных эффектов модели с взаимодействиями отображает только слагаемые переменной, то есть срез при настройке партнеров, где каждое слагаемое взаимодействия равно нулю (в данном случае — новая сделка), а не форму рынка; проанализируйте график переменной как объект оценки.

Последнее следствие касается скорее корректировки, чем ставки, и именно оно отражается в таблице цен.

Теорема 4 (разбиение корректировки зависит от порядка). В условиях случая 2 пусть $y = z_p$ ($p \geq 1$) будет одной из остальных характеристик, а w обозначает остальные, $z_k — k \notin \{0, p\}$; обозначим $f(x, y, w)$ как f с соответствующим группированием аргументов. Пусть аналогичное значение находится по адресу (x_c, y_c, w) , а объект оценки — по адресу (x_s, y_s, w) ; для обоих используется один и тот же w . Для значения y^* из y определим *корректировку для x , при этом y фиксируется на y^** , как

$$\Delta_x(y^*) = f(x_s, y^*, w) - f(x_c, y^*, w).$$

Пусть $M_{x,y}$ — это слагаемые, содержащие как x , так и y . Каждый такой член равен произведению его коэффициента на три части: $\gamma_m A_m(x) C_m(y) Q_m(w)$: $A_m(x) = B_{m,0}(x)$ является его множителем в x , который сам по себе является x или шарниром в x ; $C_m(y) = B_{m,p}(y)$ является его множителем в y , который сам по себе является y или шарниром в y ; а $Q_m(w)$ — это произведение его множителей в остальных признаках, каждый из которых сам по себе является таким признаком или шарниром в нём, и равно 1, когда у члена нет других множителей. В приведённом ниже примере одним из таких членов является $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$, причём $A_m(x) = h(\text{gla} - 1200)$, $C_m(y) = h(\text{lot} - 6000)$ и $Q_m(w) = 1$. Запишите $\Delta_x(y_s)$ и $\Delta_x(y_c)$ для корректировки x , при этом y фиксируется на значении объекта y_s и на значении аналога y_c , то есть $\Delta_x(y^*)$ при $y^* = y_s$ и при $y^* = y_c$; в приведенном примере, где x — жилая площадь, а y — участок, $\Delta_x(y_c) = \$240,000$ при участке аналога площадью 5 000 квадратных футов, а $\Delta_x(y_s) = \$288,000$ — при участке аналога площадью 10 000. Тогда

$$\Delta_x(y_s) - \Delta_x(y_c) = \sum_{m \in M_{x,y}} \gamma_m [A_m(x_s) - A_m(x_c)] [C_m(y_s) - C_m(y_c)] Q_m(w),$$

в то время как полная разность $f(x_s, y_s, w) - f(x_c, y_c, w)$ не предполагает выбора y^* . Утверждение касается именно этого среза из двух характеристик: одна из них y , а все остальные для обоих свойств остаются неизменными w ; когда между ними различаются три или более характеристик, распределение общей суммы между ними зависит от пути таким же образом, как и вдоль более длинной цепочки перемещений по одной характеристике, и теорема применяется к каждому шагу цепочки, при этом ещё не изменённые характеристики сохраняют свои текущие значения. В частности, при наличии единственного такого члена $\gamma A(x) C(y)$ и отсутствии множителей в w две корректировки для x различаются величиной $\gamma [A(x_s) - A(x_c)] [C(y_s) - C(y_c)]$, которая не равна 0 всякий раз, когда перемещение в x изменяет A , а два свойства различаются по C . Для $A = h(x - t)$ и $C = y$ это $\gamma [h(x_s - t) - h(x_c - t)] (y_s - y_c)$; для $A = h(x - t)$ и $C = h(y - s)$ — как в приведенном ниже примере — это $\gamma [h(x_s - t) - h(x_c - t)] [h(y_s - s) - h(y_c - s)]$.

Доказательство теоремы 4. Сгруппируйте слагаемые выражения f в зависимости от того, какие из x , y они содержат. Слагаемое, не содержащее ни x , ни y , принимает одинаковое значение в (x_s, y^*, w) и (x_c, y^*, w) и аннулируется в $\Delta_x(y^*)$. Слагаемое, содержащее y , но не x , также аннулируется. Слагаемое, содержащее x , но не y , вносит вклад в $\Delta_x(y^*)$ величину, не зависящую от y^* , поэтому оно аннулируется в $\Delta_x(y_s) - \Delta_x(y_c)$. Член $m \in M_{x,y}$ вносит вклад $\gamma_m [A_m(x_s) - A_m(x_c)] C_m(y^*) Q_m(w)$ в $\Delta_x(y^*)$; при установке $y^* = y_s$ и $y^* = y_c$ и вычитании получается указанная сумма. Общая разность представляет собой разность двух значений одной функции f и не упоминает y^* . ■

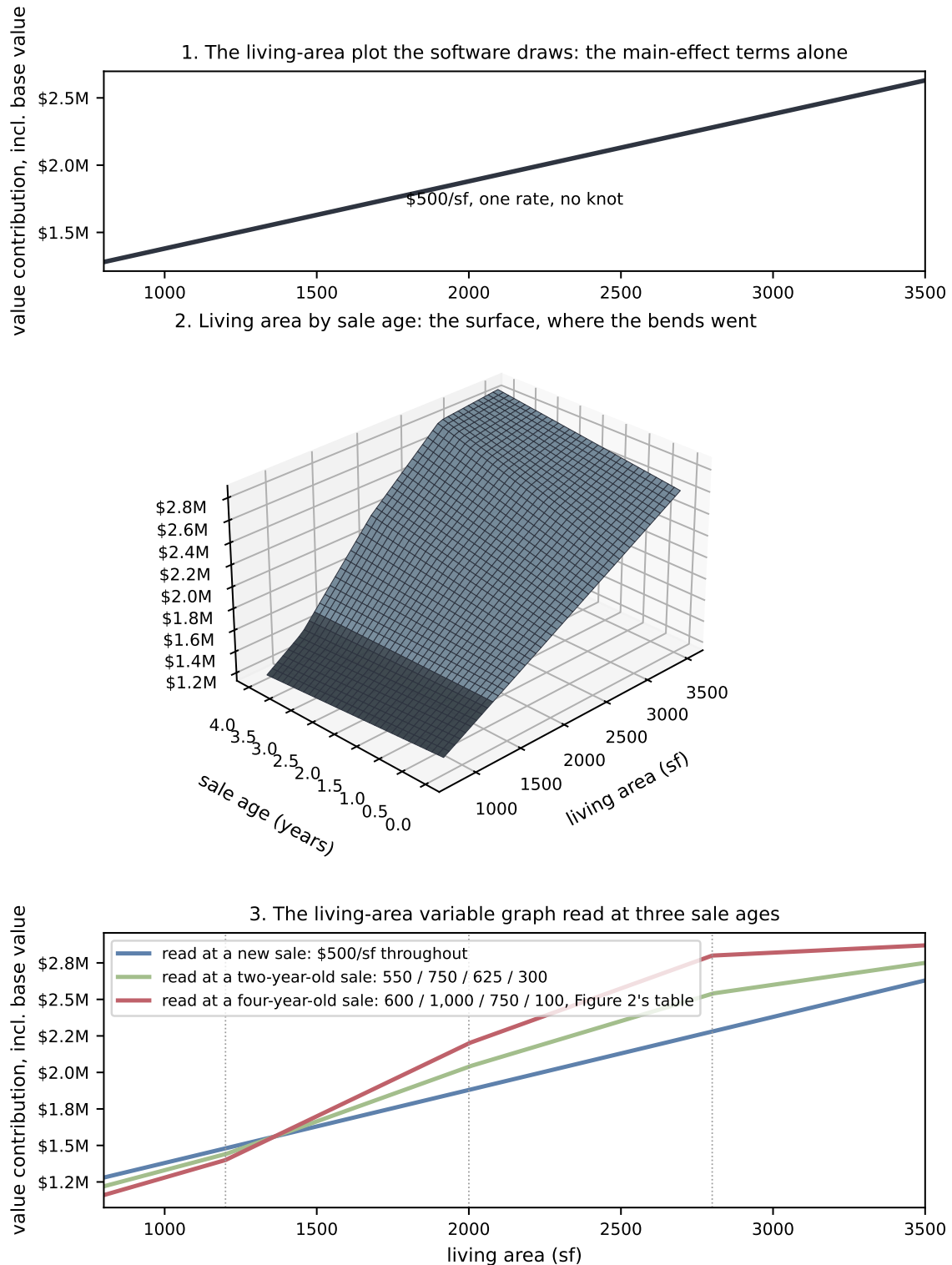


Figure 8: Рынок, представленный на рисунке 2, с учетом взаимодействия между жилой площадью и сроком владения: $1,280,000 + 500 h(\text{gla} - 800) - 20,000 \text{ saleyrs} + [-25 h(1200 - \text{gla}) + 125 h(\text{gla} - 1200) - 62.5 h(\text{gla} - 2000) - 162.5 h(\text{gla} - 2800)] \times \text{saleyrs}$, срок владения в годах от 0 до 4. Описание панелей приведено в их заголовках; для срока владения в четыре года показатели на нижнем графике точно соответствуют показателям на рисунке 2, \$120,000, но находятся на более низком уровне. График построен в демонстрационных целях, без аппроксимации.

Таким образом, значение взаимодействия относится к паре характеристик, а не к какой-либо из них в отдельности. Таблица, в которой одна характеристика подвергается корректировке за раз в соответствии со значениями аналога, а другая — в соответствии со значениями объекта оценки, совпадает по итоговому значению, но расходится в распределении значений. Удержание остальных характеристик на значениях объекта оценки является одной из обоснованных конвенций и именно её использует программа; в отчёте должно быть указано, какая именно конвенция была использована. Стандартная таблица для жилой недвижимости, такая как таблица, используемая в Едином отчёте по оценке жилой недвижимости (URAR), не содержит строки для взаимодействия. При пошаговой корректировке по одному признаку эти два шага должны удерживать «партнера» на противоположных концах — сначала жилая площадь корректируется с учетом участка объекта оценки, затем участок — с учетом жилой площади объекта аналога, или наоборот — так, чтобы взаимодействие учитывалось один раз, а сумма строк равнялась общей разнице модели. При прямом расчёте вкладов в таблице взаимодействие также можно выделить в отдельную строку — разность членов взаимодействия между объектом оценки и аналогом — рядом со строками для каждого признака; в этом случае каждая строка представляет собой одну группу членов, и порядок их расположения не имеет значения. В приведенном ниже примере три строки: жилая площадь \$240,000, участок \$75,000 и жилая площадь по участку \$120,000, сумма которых равна тому же \$435,000. При наличии взаимодействий секансная ставка пары изменяется в зависимости от пары двумя способами: во-первых, в зависимости от пересекаемых диапазонов, а во-вторых, в зависимости от того, где находится пара среди партнеров.

Разобранный пример с взаимодействием

Снова рынок недвижимости на побережье, теперь с двумя характеристиками — жилой площадью и площадью участка — и одним взаимодействием между ними. earth prints:

(Intercept)	900000
$h(gla-1200)$	600
$h(gla-2400)$	-350
$h(gla-1200)*h(lot-6000)$	0.03
lot	15

Как указано в теореме 3. Жилая площадь имеет два узла — 1 200 и 2 400 — и три слагаемых в « M_{gla} »: два простых шарнира и взаимодействие, партнёром которого является « $h(lot - 6000)$ ». На отрезке от 1 200 до 2 400 активен простой шарнир в точке 1 200, а также шарнир взаимодействия в точке 1 200, поэтому ставка в этом диапазоне составляет $600 + 0.03 h(lot - 6000)$: переменный график в зависимости от площади участка, постоянная величина 600 для участков площадью до 6 000 квадратных футов и увеличивающаяся на 3 цента за каждый дополнительный квадратный фут участка. При площади свыше 2 400 вступает в действие поворотная точка на отметке 2 400, и ставка составляет $250 + 0.03 h(lot - 6000)$. Ставка, зависящая от площади участка, составляет $15 + 0.03 h(gla - 1200)$ при площади участка свыше 6 000 квадратных футов и 15 при меньшей площади, поэтому ставка каждой составляющей зависит от величины другой.

Две точки, две таблицы. Таблица ставок программы для жилой площади в диапазоне от 800 до 3 500 квадратных футов: один случай для участка площадью 5 000 квадратных футов и один — для участка площадью 10 000 квадратных футов. Каждый столбец отражает наклон кривой модели по жилой площади при фиксированном значении площади участка, и только при этом значении:

жилая площадь (кв. футы)	стоимость за 5 000 лотов (\$за квадратный фут)	стоимость за 10 000 лотов (\$за квадра
от 800 до 1 200	0	
от 1 200 до 2 400	600	
от 2 400 до 3 500	250	

На меньшем участке коэффициент взаимодействия $h(5000-6000)$ равен нулю, поэтому взаимодействие не вносит никакого вклада, и таблица соответствует результатам модели без этого взаимодействия; то есть это случай из «Следствия 6», когда «член может исчезнуть». На более крупном участке коэффициент взаимодействия составляет 4,000, взаимодействие добавляет $0.03 \times 4,000 = 120$ к каждому сегменту выше 1 200, и один и тот же узел — 1 200 — питает два активных слагаемых, что свидетельствует о несостоятельности ограничения «одно на узел».

Теперь о корректировке. Объект аналога имеет площадь 1 800 квадратных футов на участке площадью 5 000 квадратных футов, а его оценочная стоимость составляет \$1,335,000; объект оценки имеет площадь 2 200 квадратных футов на участке площадью 10 000 квадратных футов, его оценочная стоимость составляет \$1,770,000. Общая сумма, подлежащая объяснению, составляет \$435,000. Сначала скорректируйте жилую площадь с учётом участка аналога, затем — участок с учётом жилой площади объекта оценки. Каждая корректировка представляет собой разницу между двумя сравниваемыми величинами; если корректировка остаётся в пределах одного сегмента, эта разница равна ширине, умноженной на коэффициент, что и отражено в таблице:

	сначала жилая площадь, потом участок	Сначала участок, потом жилая площадь
корректировка жилой площади	$400 \times 600 = 240,000$	$400 \times 720 = 288,000$
корректировка участка	195,000	147,000
всего	435,000	435,000

Корректировки по участку представляют собой разницу в размере взносов по той же схеме, и эти разницы объясняются соответствующими ставками; ставка для данного участка меняется при площади 6 000 квадратных футов: при площади объекта оценки в 2 200 квадратных футов ставка для участка составляет \$15 ниже порогового значения и \$45 выше него, поэтому переход от 5 000 до 10 000 составляет $1,000 \times 15 + 4,000 \times 45 = 195,000$, а не $5,000 \times 45$; при площади аналога в 1 800 квадратных футов она составляет $1,000 \times 15 + 4,000 \times 33 = 147,000$. Итоговые значения совпадают, как и предписывает Теорема 4, а разбивки различаются на \$48,000, что точно соответствует формуле теоремы: $0.03 \times [h(2200 - 1200) - h(1800 - 1200)] \times [h(10000 - 6000) - h(5000 - 6000)] = 0.03 \times 400 \times 4,000$. Эта величина \$48,000 представляет собой ценность наличия дополнительных 400 квадратных футов жилой площади и дополнительных 4 000 квадратных футов участка вместе взятых; она не относится ни к одной из этих характеристик, и та из них, которая подвергается второй корректировке, получает эту сумму. Согласно правилам программы, при фиксированных значениях остальных характеристик, в правой колонке для жилой площади указывается \$288,000 по цене \$720 за квадратный фут, а в отчёте указывается, в какой точке был зафиксирован этот показатель:

```
Feature: gla
Rates by range (data from 800 to 3500; nothing is reported outside it):
  800 .. 1200   rate 0.0
  1200 .. 2400   rate 720.0
  2400 .. 3500   rate 370.0
Fitted value at the knots (intercept and partners included):
```

```

1200    1050000.0
2400    1914000.0
Comparable at 1800: contribution 432000.0
Subject at 2200:    contribution 720000.0
Adjustment, range by range:
  1800 .. 2200    400 at 720.0 = 288000.0
Adjustment added to the comparable: 288000.0
Secant rate for the pair (one straight-line rate): 720.0
Check: segments sum to the adjustment.

```

Оба показателя рассчитаны исходя из площади участка объекта оценки в 10,000 квадратных футов; при площади собственного участка аналога в 5,000 квадратных футов показатель жилой площади составляет \$360,000.

Рецензент, который повторно запускает эту программу на участке с аналогичными данными, получает результат \$240,000 и не ошибается; рецензент и программа считывают данные с одной и той же поверхности в двух точках, и в отчёте должно быть указано, в каких именно.

Короче говоря: без взаимодействий модель MARS представляет собой таблицу ставок по каждому признаку, и все утверждения о ставке, сделанные для обычной регрессии, переносятся сегмент за сегментом. При наличии взаимодействий ставка представляет собой поверхность, и теорема 3 точно определяет, как она изгибается, а оценщик должен указать, в какой точке она была зафиксирована.

Программа

Функция `mars_rates.pl`, полный текст которой приведён в приложении, считывает таблицу `earth` в печатном виде или модель, записанную в виде терма языка Prolog, и выдает таблицу ставок, значения в узлах, два вклада и их разность, корректировку с пояснениями по каждому диапазону, текущую ставку пары, а также проверку того, что сумма значений по диапазонам равна корректировке. Учитывая диапазон данных — самый маленький и самый большой дом из списка продаж — программа не предоставляет отчёта о ставках за пределами этого диапазона: алгебра шарнира простирается до бесконечности в одну сторону, а данные о продажах — нет, и ставка за пределами этого диапазона является экстраполяцией. Например, для данных от 800 до 3 500 квадратных футов:

```

Feature: gla
Rates by range (data from 800 to 3500; nothing is reported outside it):
  800 .. 1200    rate 600
  1200 .. 2000   rate 1000
  2000 .. 2800   rate 750
  2800 .. 3500   rate 100
Fitted value at the knots (intercept and partners included):
  1200    1520000
  2000    2320000
  2800    2920000
Comparable at 1800: contribution 600000
Subject at 2400:    contribution 1100000
Adjustment, range by range:
  1800 .. 2000    200 at 1000 = 200000
  2000 .. 2400    400 at 750 = 300000
Adjustment added to the comparable: 500000

```

Secant rate for the pair (one straight-line rate): 833.3333333333334
Check: segments sum to the adjustment.

В этом блоке представлены два вида чисел. Вклад определяется исключительно счётами, связанными с жилой площадью, при этом счёты, связанные с пересечением и всеми остальными характеристиками, исключены: \$600,000 и \$1,100,000 различаются именно на величину корректировки. Значения в узлах являются аппроксимированными значениями, включая пересечение, а в модели с взаимодействиями — также и члены, относящиеся к партнерам в данной точке; здесь \$1,520,000 при 1 200 квадратных футов является пересечением, поскольку все члены взаимодействия в этой точке равны нулю. Подгоняемые значения в точках аналогов и объекта оценки — \$2,120,000 и \$2,620,000 — также содержат пересечение. В отчёте значения в узлах обозначены как подгоняемые, чтобы читатель не принял их за вклады.

Именно этот блок и должен быть включен в приложение к отчёту: коэффициенты, узлы и арифметические вычисления, связывающие одно с другим, в таком виде, чтобы рецензент мог повторить их вручную. 69 тестов программы проверяют правило знаков, правило взаимодействия, согласованность сегментов с корректировками, а также синтаксический анализатор для вывода `earth` и проходят тестирование в Trealla, Prolog журнала, а также в SWI-Prolog и Stryer Prolog. Для читателей, работающих в R, тот же материал доступен без необходимости выходить из этой среды: файл `mars_rates.R`, опубликованный вместе с программой, принимает объект `earth` или его распечатанную таблицу и выводит тот же блок, с коэффициентами, ограниченным диапазоном данных; Prolog остаётся проверенным эталоном, а файл R содержит собственные средства самопроверки по примерам из данного приложения. Программа, тесты к ней и скрипт для создания рисунка публикуются вместе с этим выпуском в виде файла для скачивания рядом с PDF-файлом в соответствии с лицензией GNU Affero General Public License, версии 3 или более поздней.

Почему MARS, и что это дополнение оставляет другим статьям

Журнал использует модели MARS по трем причинам, и именно они делают возможным существование данного приложения. Объективность: узлы размещаются там, где их размещает система продаж, в соответствии с одной заданной процедурой и одним заданным правилом обрезки (поздний этап, на котором удаляются термы, не заслуживающие своего места), и ничего, касающегося формы переменного графа, не определяется заранее, помимо указанных настроек (максимальная степень взаимодействия, штраф за обрезку, минимальное расстояние между узлами). Обнаруживаемость: прямой проход просматривает каждую признаковую величину и каждый кандидат в узел, включая взаимодействия до заданной степени, и предоставляет отчёт о том, что он обнаружил, чего не может сделать ни один метод, начинающий с заранее выбранной формы. Интерпретируемость: постоянные коэффициенты в диапазонах, выбранных на основе данных, и взаимодействия, построенные на основе тех же «шарниров», позволяют записать модель в виде таблицы коэффициентов, а корректировку объяснить как произведение коэффициента на ширину. Два других метода также дают коэффициенты, которые изменяются в пределах диапазона признака, и они отличаются друг от друга. Первый — сегментированная (типа «ломанной палки») регрессия — принимает переменную и количество избранных аналитиком точек разрыва и оценивает, где они находятся, с указанием стандартной ошибки (показателя того, насколько оценка может измениться при других объемах продаж) для каждой из них; между точками разрыва коэффициент остается постоянным, как в модели MARS. Второй метод — аддитивная модель с ограничениями по форме — осуществляет подгонку плавной кривой вместо прямых отрезков, принудительно придавая ей форму, заданную аналитиком — например, что значение никогда не снижается по мере увеличения жилой площади или что каждый добавленный квадратный фут прибавляет меньше, чем предыдущий. Ни один из этих методов здесь не используется. Каждая из них отвечает на вопрос, который в данном приложении не задается, и каждая теряет по крайней мере одно из трёх свойств: сегментированная

регрессия теряет обнаруживаемость, поскольку переменная и количество переломов выбираются заранее; модель с ограничениями по форме теряет интерпретируемость, поскольку кривая не имеет постоянных коэффициентов, которые можно было бы представить в таблице, и объективность, поскольку её форма навязана. В данном приложении рассматривается модель, создаваемая программой MARS; выбор метода, а также стабильность узлов и коэффициентов, которые MARS находит при переходе от одного набора данных о продажах к другому, являются объектами оценки.

Чего это дополнение не решает

Показатели относятся к модели, и к ним в неизменном виде применимы четыре статистических положения: показатель приближается к условному математическому ожиданию лишь настолько, насколько это позволяет форма модели (*условное математическое ожидание*); он может отражать значение пропущенной переменной (*смещение пропущенной переменной*); она может искажаться, когда переменная изменяется вместе со структурной ошибкой (*Эндогенность*); и она может вообще не определяться данными о продажах, каким бы большим ни было их количество, если две переменные всегда изменяются синхронно (*Идентификация параметров*). MARS обнаруживает изменение показателя только там, где данные о продажах содержат контраст, позволяющий его продемонстрировать; «узел» представляет собой утверждение о рынке и достоверен ровно настолько, насколько достоверны данные о продажах по обе стороны от него. Дополнение данного материала носит более узкий характер: исходя из модели, здесь объясняется, как интерпретировать её выводы.

References.

- Craytor, W. B. (2026). earthUI: An Interactive Interface for the earth (MARS) Package. *Valuation Engineer Journal*, 1(1). <https://doi.org/10.67330/7fj41b65>
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.), §9.4, Multivariate Adaptive Regression Splines. Springer.
- Milborrow, S. (2024a). earth: Multivariate Adaptive Regression Splines. R package. <https://CRAN.R-project.org/package=earth>
- Milborrow, S. (2024b). plotmo: Plot a Model's Residuals, Response, and Partial Dependence Plots. R package. <https://CRAN.R-project.org/package=plotmo>
- Milborrow, S. (n.d.). Notes on the earth Package. <http://www.milbo.org/doc/earth-notes.pdf>

Cross-references: conditional expectation; omitted variable bias; endogeneity; parameter identification; hedonic regression; implicit price.

Addendum: Prolog Formalization

The following machine-readable companion (`mars_rates.pl`) formalizes this entry in the journal's Prolog knowledge base. It is also published as a browsable galley alongside this article.

```
% SPDX-License-Identifier: AGPL-3.0-or-later
% Copyright (C) 2026 William Bert Craytor. Published by The Valuation Engineer
% Journal (Pacific Vista Net). This program is free software: you can
% redistribute it and/or modify it under the terms of the GNU Affero General
% Public License as published by the Free Software Foundation, either version 3
% of the License, or (at your option) any later version. It is distributed
```

```
% WITHOUT ANY WARRANTY, without even the implied warranty of MERCHANTABILITY or
% FITNESS FOR A PARTICULAR PURPOSE; see the COPYING file beside it or
% <https://www.gnu.org/licenses/agpl-3.0.html>.
```

```
/* mars_rates.pl -- mars-rates: from a fitted MARS model to the rates, value
contributions and adjustments an appraiser reports.
```

A MARS model, as the R package earth prints it, is an intercept plus a sum of terms, each a coefficient times a product of basis functions:

```
h(x - t)    a hinge, zero for x below the knot t, rising with slope +1 above
h(t - x)    a reflected hinge, zero above t, rising to the LEFT: slope -1 in x
x           a linear term
h(x-t)*h(y-s) an interaction, the product of two factors
```

The appraiser does not report basis functions. The appraiser reports, for a feature, the RATE the market pays per unit in each range (the slope), the VALUE CONTRIBUTION of the quantity a property has (the level), and the ADJUSTMENT: the amount added to a comparable's contribution to make it equal to the subject's. This program turns the one into the other, and checks that they agree.

The sign rule that hand-built tables get wrong: in a range, the rate is the sum of the coefficients of the active right hinges MINUS the sum of the coefficients of the active reflected hinges (plus any linear term's coefficient). A negative coefficient on $h(5 - x)$ contributes a POSITIVE slope below 5.

Representation. A model is

```
model(Intercept, Terms)
```

where each term is `term(Coef, Factors)` and each factor is

```
hinge(Var, Knot)      h(Var - Knot)
rhinge(Var, Knot)     h(Knot - Var)
linear(Var)           Var
```

A point is a list of `Var = Value` pairs, e.g. `[gla = 1800, lot = 6000]`.

Entry points:

```
value(+Model, +Point, -Y)           the fitted value at a point
contribution(+Model, +Var, +Point, -C) the part of the value that the
terms involving Var contribute
rate(+Model, +Var, +Point, -R)       slope in Var at the point (other
variables held at the point)
knots(+Model, +Var, -Knots)          the sorted knots on Var
rate_table(+Model, +Var, +Point, -Rows)
rows range(Lo, Hi, Rate), Lo/Hi
numbers or the atoms minus_inf,
plus_inf; other variables fixed
at the point
rate_table(+Model, +Var, +Point, +data(Lo, Hi), -Rows)
the same table clamped to the
data range: the first row starts
at Lo, the last ends at Hi, rows
outside the data are dropped.
Rates outside the data are
extrapolation and are not
reported
knot_values(+Model, +Var, +Point, -Pairs)
Knot-Value at each knot
adjustment(+Model, +Var, +Point, +CompQ, +SubjQ, -A)
amount added to the comparable's
```

```

                                contribution (Var = CompQ) to
                                equal the subject's (Var = SubjQ)
segments(+Model, +Var, +Point, +CompQ, +SubjQ, -Segs)
                                the adjustment as the appraiser
                                explains it: seg(Lo, Hi, Rate,
                                Width, Amount) per range crossed
pair_rate(+Model, +Var, +Point, +CompQ, +SubjQ, -R)
                                the pair's secant rate: the one rate a straight line
↪ would
                                need between the two quantities:
                                adjustment over width
consistent(+Model, +Var, +Point, +CompQ, +SubjQ)
                                the segments sum to the adjustment
parse_earth(+Atom, -Model)      read earth's printed coefficient
                                table (see parse_earth below)
print_report(+Model, +Var, +Point, +CompQ, +SubjQ)

Plain ISO Prolog: runs under Trealla, SWI-Prolog and Scryer Prolog. No
modules, no format/2, no double-quoted strings. Consult this file, then
tests/harness.pl and tests/test_mars_rates.pl, and call run_tests.      */

/* ===== */
/* 1. Evaluation */
/* ===== */

point_value(Var, Point, X) :- mr_member(Var = X, Point), !.
point_value(Var, _, _) :- throw(error(missing_variable(Var), point_value/3)).

factor_value(hinge(Var, T), Point, V) :- point_value(Var, Point, X), D is X - T, V is max(0, D).
factor_value(rhinge(Var, T), Point, V) :- point_value(Var, Point, X), D is T - X, V is max(0, D).
factor_value(linear(Var), Point, V) :- point_value(Var, Point, V).

term_value(term(Coef, Factors), Point, V) :-
    factors_product(Factors, Point, P), V is Coef * P.

factors_product([], _, 1).
factors_product([F | Fs], Point, P) :- factor_value(F, Point, V), factors_product(Fs, Point, P0), P
    ↪ is V * P0.

value(model(Intercept, Terms), Point, Y) :-
    terms_sum(Terms, Point, S), Y is Intercept + S.

terms_sum([], _, 0).
terms_sum([T | Ts], Point, S) :- term_value(T, Point, V), terms_sum(Ts, Point, S0), S is S0 + V.

/* The contribution of a variable at a point: the terms that involve it. The
   intercept and the terms in other variables are the level the feature is
   measured against; an adjustment never depends on them. */

involves(term(_, Factors), Var) :- mr_member(F, Factors), factor_var(F, Var), !.
factor_var(hinge(V, _), V). factor_var(rhinge(V, _), V). factor_var(linear(V), V).

contribution(model(_, Terms), Var, Point, C) :-
    findall(V, ( mr_member(T, Terms), involves(T, Var), term_value(T, Point, V) ), Vs),
    mr_sum(Vs, C).

/* ===== */
/* 2. Rates (slopes) */
/* ===== */

/* d/dVar of a term = Coef * (slope of the factor in Var) * (product of the

```

```

other factors' values). A hinge is active strictly above its knot, a
reflected hinge strictly below; at the knot itself the slope is taken from
the right, which is the convention of the rate table below.          */

factor_slope(hinge(Var, T), Var, Point, S) :- !, point_value(Var, Point, X), ( X >= T -> S = 1 ; S
    ↪ = 0 ).
factor_slope(rhinge(Var, T), Var, Point, S) :- !, point_value(Var, Point, X), ( X < T -> S = -1 ; S
    ↪ = 0 ).
factor_slope(linear(Var), Var, _, 1)          :- !.
factor_slope(_, _, _, 0).

term_slope(term(Coef, Factors), Var, Point, S) :-
    term_slope_(Factors, Var, Point, [], Coef, S).

term_slope_([], _, _, _, 0).
term_slope_([F | Fs], Var, Point, Seen, Coef, S) :-
    ( factor_var(F, Var) ->
        factor_slope(F, Var, Point, S1),
        mr_append(Seen, Fs, Others), factors_product(Others, Point, P),
        S1 is Coef * S1 * P
    ; S1 = 0 ),
    term_slope_(Fs, Var, Point, [F | Seen], Coef, S0),
    S is S0 + S1.

rate(model(_, Terms), Var, Point, R) :-
    findall(S, ( mr_member(T, Terms), term_slope(T, Var, Point, S) ), Ss),
    mr_sum(Ss, R).

/* ===== */
/* 3. Knots, ranges and the rate table */
/* ===== */

knots(model(_, Terms), Var, Knots) :-
    findall(K, ( mr_member(term(_, Fs), Terms), mr_member(F, Fs), knot_of(F, Var, K) ), Ks),
    sort(Ks, Knots).

knot_of(hinge(Var, K), Var, K).
knot_of(rhinge(Var, K), Var, K).

/* One row per range between consecutive knots, the rate measured at a point
inside the range (the midpoint, or one unit past the last knot). */

rate_table(Model, Var, Point, Rows) :-
    knots(Model, Var, Knots),
    bounds(Knots, Bounds),
    rows(Bounds, Model, Var, Point, Rows).

bounds([], [minus_inf-plus_inf]).
bounds([K | Ks], [minus_inf-K | Rest]) :- bounds_(K, Ks, Rest).
bounds_(K, [], [K-plus_inf]).
bounds_(K, [K2 | Ks], [K-K2 | Rest]) :- bounds_(K2, Ks, Rest).

/* The table over the observed data only. A hinge's algebra runs to infinity
on one side; the sales do not. Rates beyond the smallest and largest
observed quantity are extrapolation, so the report clamps the end rows to
the data range and drops any row that lies wholly outside it. */

rate_table(Model, Var, Point, data(DLo, DHi), Rows) :-
    rate_table(Model, Var, Point, Rows0),
    clamp_rows(Rows0, DLo, DHi, Rows).

```

```

clamp_rows([], _, _, []).
clamp_rows([range(Lo, Hi, R) | Rs], DLo, DHi, Out) :-
    clamp_lo(Lo, DLo, Lo1), clamp_hi(Hi, DHi, Hi1),
    ( Lo1 < Hi1 -> Out = [range(Lo1, Hi1, R) | Rest] ; Out = Rest ),
    clamp_rows(Rs, DLo, DHi, Rest).

clamp_lo(minus_inf, DLo, DLo) :- !.
clamp_lo(Lo, DLo, X) :- ( Lo < DLo -> X = DLo ; X = Lo ).
clamp_hi(plus_inf, DHi, DHi) :- !.
clamp_hi(Hi, DHi, X) :- ( Hi > DHi -> X = DHi ; X = Hi ).

rows([], _, _, _, []).
rows([Lo-Hi | Bs], Model, Var, Point, [range(Lo, Hi, R) | Rows]) :-
    probe(Lo, Hi, X), set_point(Var, X, Point, P), rate(Model, Var, P, R),
    rows(Bs, Model, Var, Point, Rows).

probe(minus_inf, plus_inf, 0) :- !.
probe(minus_inf, Hi, X) :- !, X is Hi - 1.
probe(Lo, plus_inf, X) :- !, X is Lo + 1.
probe(Lo, Hi, X) :- X is (Lo + Hi) / 2.

set_point(Var, X, Point, [Var = X | Rest]) :- mr_delete_var(Point, Var, Rest).

knot_values(Model, Var, Point, Pairs) :-
    knots(Model, Var, Knots),
    findall(K-Y, ( mr_member(K, Knots), set_point(Var, K, Point, P), value(Model, P, Y) ), Pairs).

/* ===== */
/* 4. Adjustments and their explanation */
/* ===== */

adjustment(Model, Var, Point, CompQ, SubjQ, A) :-
    set_point(Var, CompQ, Point, PC), set_point(Var, SubjQ, Point, PS),
    value(Model, PC, YC), value(Model, PS, YS), A is YS - YC.

/* The path from the comparable's quantity to the subject's, cut at every knot
   between them; each piece is a width at a rate. Going down (subject smaller
   than comparable) the widths are negative, so the amounts are, and the sum
   is still the adjustment as defined: added to the comparable. */

segments(Model, Var, Point, CompQ, SubjQ, Segs) :-
    knots(Model, Var, Knots),
    Lo is min(CompQ, SubjQ), Hi is max(CompQ, SubjQ),
    findall(K, ( mr_member(K, Knots), K > Lo, K < Hi ), Inner),
    mr_append([Lo | Inner], [Hi], Cuts),
    ( SubjQ >= CompQ -> Dir = 1 ; Dir = -1 ),
    pieces(Cuts, Model, Var, Point, Dir, Segs).

pieces([], _, _, _, _, []).
pieces([A, B | Cs], Model, Var, Point, Dir, [seg(A, B, R, W, Amt) | Segs]) :-
    probe(A, B, X), set_point(Var, X, Point, P), rate(Model, Var, P, R),
    W is Dir * (B - A), Amt is R * W,
    pieces([B | Cs], Model, Var, Point, Dir, Segs).

/* The pair's secant rate. Between two quantities the only single rate
   consistent with the MARS contributions is the chord, adjustment over
   width: the model's rates averaged by the width of each range crossed. An
   ordinary regression, one rate for the feature, must serve every pair with
   one number and so cannot reproduce these. Printed beside the
   MARS rates it shows what a single rate would have had to be for THIS
   comparable and THIS subject, and how far each range's rate departs from

```

```

it. Undefined when the two quantities are equal.                                */

pair_rate(Model, Var, Point, CompQ, SubjQ, R) :-
    CompQ = \= SubjQ,
    adjustment(Model, Var, Point, CompQ, SubjQ, A), R is A / (SubjQ - CompQ).

consistent(Model, Var, Point, CompQ, SubjQ) :-
    adjustment(Model, Var, Point, CompQ, SubjQ, A),
    segments(Model, Var, Point, CompQ, SubjQ, Segs),
    findall(Amt, mr_member(seg(_, _, _, Amt), Segs), Amts), mr_sum(Amts, S),
    abs(S - A) < 0.000001.

/* ===== */
/* 5. Reading earth's printed coefficients                                     */
/* ===== */

/* earth prints its model as a table, one term per line, e.g.

    (Intercept)      10
    h(x-2)           3
    h(5-x)          -1.5
    h(x-8)           0.8
    h(x-2)*h(y-3)    0.4
    x                1.2

parse_earth/2 takes that table as one atom (lines separated by newline
characters) and returns the model. Whitespace is free; a coefficient is a
decimal number with an optional sign; variable names are letters, digits,
dots and underscores, as R prints them.                                     */

parse_earth(Atom, model(Intercept, Terms)) :-
    atom_chars(Atom, Chars), split_lines(Chars, Lines),
    parse_lines(Lines, 0, Intercept, Terms).

split_lines(Chars, Lines) :- split_lines_(Chars, [], Lines).
split_lines_([], Acc, [Line]) :- mr_reverse(Acc, Line).
split_lines_(['\n' | Cs], Acc, [Line | Lines]) :- !, mr_reverse(Acc, Line), split_lines_(Cs, [],
    ↪ Lines).
split_lines_([C | Cs], Acc, Lines) :- split_lines_(Cs, [C | Acc], Lines).

parse_lines([], I, I, []).
parse_lines([L | Ls], IO, I, Terms) :-
    strip(L, S),
    ( S == [] -> parse_lines(Ls, IO, I, Terms)
    ; split_ws(S, Name, CoefChars), number_chars_safe(CoefChars, Coef),
      ( Name = ['(', 'I', 'n', 't', 'e', 'r', 'c', 'e', 'p', 't', ')'] ->
        parse_lines(Ls, Coef, I, Terms)
      ; parse_factors(Name, Factors),
        Terms = [term(Coef, Factors) | Rest], parse_lines(Ls, IO, I, Rest) ) ).

strip(L, S) :- drop_ws(L, L1), mr_reverse(L1, R1), drop_ws(R1, R2), mr_reverse(R2, S).
drop_ws([C | Cs], S) :- ws(C), !, drop_ws(Cs, S).
drop_ws(S, S).
ws(' '). ws('\t'). ws('\r').

/* the coefficient is the LAST whitespace-separated token; the name is what
   precedes it and may itself contain spaces (earth prints h(x- 3) for a
   negative knot)                                                           */
split_ws(S, Name, Coef) :-
    mr_reverse(S, R), split_ws_(R, [], Coef, NameR), % consuming the reversed line puts the
    ↪ coefficient back in order

```

```

    mr_reverse(NameR, Name0), strip(Name0, Name).
split_ws_([C | Cs], Acc, Coef, Rest) :- ws(C), !, Coef = Acc, Rest = Cs.
split_ws_([C | Cs], Acc, Coef, Rest) :- split_ws_(Cs, [C | Acc], Coef, Rest).
split_ws_([], Acc, Acc, []).

number_chars_safe(Chars, N) :-
    catch(number_chars(N, Chars), _, fail), !.
number_chars_safe(Chars, N) :- % "-.5" and ".5", which earth can print
    ( Chars = ['-', '.' | R] -> number_chars(N, ['-', '0', '.' | R])
    ; Chars = ['.' | R] -> number_chars(N, ['0', '.' | R]) ).

/* factors are separated by '*' */
parse_factors(Chars, [F | Fs]) :-
    split_star(Chars, First, Rest), parse_factor(First, F),
    ( Rest == [] -> Fs = [] ; parse_factors(Rest, Fs) ).

split_star(Chars, First, Rest) :- split_star_(Chars, [], First, Rest).
split_star_(['*' | Cs], Acc, First, Cs) :- !, mr_reverse(Acc, First).
split_star_([C | Cs], Acc, First, Rest) :- split_star_(Cs, [C | Acc], First, Rest).
split_star_([], Acc, First, []) :- mr_reverse(Acc, First).

/* h(x-2) / h(2-x) / x -- inside the parentheses, the side that is a number
    is the knot and the other side the variable */
parse_factor(['h', '(' | Cs], F) :- !,
    mr_reverse(Cs, [' ' | RevInner]), mr_reverse(RevInner, Inner),
    split_minus(Inner, Left, Right),
    ( number_chars_safe(Right, K) -> atom_chars(Var, Left), F = hinge(Var, K)
    ; number_chars_safe(Left, K) -> atom_chars(Var, Right), F = rhinge(Var, K) ).
parse_factor(Cs, linear(Var)) :- atom_chars(Var, Cs).

/* the minus that separates the two sides is the last one not at position 0
    and not following an 'e' (so "x-2", "5-x", "gla-1500" and "h(x- -3)" all
    read); a leading '-' belongs to a negative knot */
split_minus(Chars, Left, Right) :-
    mr_length(Chars, N), N1 is N - 1,
    minus_positions(Chars, 0, Ps), mr_member(P, Ps), P > 0, P < N1,
    mr_take(P, Chars, Left0), strip(Left0, Left),
    P2 is P + 1, mr_drop(P2, Chars, Right0), strip(Right0, Right),
    Left \== [], Right \== [], !.

minus_positions([], _, []).
minus_positions(['-' | Cs], I, [I | Ps]) :- !, I1 is I + 1, minus_positions(Cs, I1, Ps).
minus_positions([_ | Cs], I, Ps) :- I1 is I + 1, minus_positions(Cs, I1, Ps).

/* ===== */
/* 6. A report */
/* ===== */

print_report(Model, Var, Point, CompQ, SubjQ) :-
    write('Feature: '), write(Var), nl,
    write('Rates by range:'), nl,
    rate_table(Model, Var, Point, Rows),
    print_report_(Model, Var, Point, CompQ, SubjQ, Rows).

/* The same report over the observed data: print_report/6 takes data(Lo, Hi),
    the smallest and largest observed quantity of Var, and reports no rate
    outside it. */

print_report(Model, Var, Point, data(DLo, DHi), CompQ, SubjQ) :-
    write('Feature: '), write(Var), nl,
    write('Rates by range (data from '), write(DLo), write(' to '), write(DHi), write('; nothing is

```

```

    ↪ reported outside it):'), nl,
    rate_table(Model, Var, Point, data(DLo, DHi), Rows),
    print_report_(Model, Var, Point, CompQ, SubjQ, Rows).

print_report_(Model, Var, Point, CompQ, SubjQ, Rows) :-
    ( mr_member(range(Lo, Hi, R), Rows), write(' '), write(Lo), write(' .. '), write(Hi), write('
    ↪ rate '), write(R), nl, fail ; true ),
    write('Fitted value at the knots (intercept and partners included):'), nl,
    knot_values(Model, Var, Point, KVs),
    ( mr_member(K-Y, KVs), write(' '), write(K), write(' '), write(Y), nl, fail ; true ),
    set_point(Var, CompQ, Point, PC), set_point(Var, SubjQ, Point, PS),
    contribution(Model, Var, PC, CC), contribution(Model, Var, PS, CS),
    write('Comparable at '), write(CompQ), write(': contribution '), write(CC), nl,
    write('Subject at '), write(SubjQ), write(': contribution '), write(CS), nl,
    segments(Model, Var, Point, CompQ, SubjQ, Segs),
    write('Adjustment, range by range:'), nl,
    ( mr_member(seg(A, B, R2, W, Amt), Segs), write(' '), write(A), write(' .. '), write(B), write('
    ↪ (' '), write(W), write(' at '), write(R2), write(' = '), write(Amt), nl, fail ; true ),
    adjustment(Model, Var, Point, CompQ, SubjQ, Adj),
    write('Adjustment added to the comparable: '), write(Adj), nl,
    ( pair_rate(Model, Var, Point, CompQ, SubjQ, OR) ->
        write('Secant rate for the pair (one straight-line rate): '), write(OR), nl ; true ),
    ( consistent(Model, Var, Point, CompQ, SubjQ) -> write('Check: segments sum to the adjustment
    ↪.') ; write('CHECK FAILED') ), nl.

/* ===== */
/* 7. List helpers (own names; no library assumed) */
/* ===== */

mr_member(X, [X | _]).
mr_member(X, [_ | T]) :- mr_member(X, T).

mr_append([], L, L).
mr_append([H | T], L, [H | R]) :- mr_append(T, L, R).

mr_reverse(L, R) :- mr_reverse_(L, [], R).
mr_reverse_([], R, R).
mr_reverse_([H | T], A, R) :- mr_reverse_(T, [H | A], R).

mr_sum([], 0).
mr_sum([X | Xs], S) :- mr_sum(Xs, S0), S is S0 + X.

mr_length([], 0).
mr_length([_ | T], N) :- mr_length(T, N0), N is N0 + 1.

mr_take(0, _, []) :- !.
mr_take(N, [H | T], [H | R]) :- N1 is N - 1, mr_take(N1, T, R).

mr_drop(0, L, L) :- !.
mr_drop(N, [_ | T], R) :- N1 is N - 1, mr_drop(N1, T, R).

mr_delete_var([], _, []).
mr_delete_var([V = _ | T], V, R) :- !, mr_delete_var(T, V, R).
mr_delete_var([H | T], V, [H | R]) :- mr_delete_var(T, V, R).

```