

Die Hürde für das menschliche Verständnis von MARS (The Barrier to Human Understanding of MARS)

Ein MARS-Modell als Einheitspreise, Wertbeiträge und Anpassungen lesen

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

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

Dies ist eine Ergänzung zum Abschnitt „Grundlagen“ und kein Konzeptbeitrag. Der Titel bezeichnet eine Hürde, die der Herausgeber im Laufe jahrelanger indirekter Beobachtung erkannt hat: Gutachter, die einen Regressionskoeffizienten problemlos lesen können, bleiben bei der Liste der „Hinge-Terme“ stehen, die ein MARS-Programm ausgibt, da nichts in dieser Liste wie ein Einheitspreis aussieht. Die vier statistischen Einträge, die auf diese Ergänzung folgen (*Bedingte Erwartung*, *Verzerrung durch ausgelassene Variablen*, *Endogenität* und *Parameteridentifikation*), befassen sich mit dem Wert, den eine Regression einem Merkmal zuweist. Die Modellierung der Zeitschrift selbst nutzt multivariate adaptive Regressionssplines (MARS; Friedman 1991)¹, die nicht einen Wert pro Merkmal, sondern einen pro Bereich ausgeben, und der Weg von den Ausgabedaten eines MARS-Programms zur Wertetabelle, die ein Gutachter verteidigen kann, weist einen Schritt auf, an dem manuell erstellte Tabellen fehlerhaft werden. Diese Beilage beschreibt den Weg, den Schritt und ein Programm, das diesen Weg durchläuft und seine eigene Arbeit überprüft.

Formale Definition. In einem MARS-Modell setzt sich der angepasste Wert aus einem Achsenabschnitt und einer Summe von Termen zusammen, wobei jeder Term aus einem Koeffizienten multipliziert mit einem Produkt aus Scharnierfunktionen der Merkmale besteht. Der angepasste Wert ist die Preisschätzung des Modells. Für ein Merkmal, während die anderen festgehalten werden, handelt es sich um eine Folge gerader Segmente. Die Punkte, an denen sich die Segmente treffen, sind die Knoten. Die Kurve des Wertbeitrags des Merkmals über seinen Bereich ist die *Variablenkurve*. Ihre Steigung in jedem Bereich ist die: die Änderung des angepassten Werts pro Einheit des Merkmals in diesem Bereich; der *Wertbeitrag* einer Menge des Merkmals ist die Summe der Modellterme für dieses Merkmal bei dieser Menge; im Beispiel sind alle diese Terme bei 1.200 Quadratfuß gleich Null, sodass der Beitrag ab diesem Wert gemessen wird und darunter negativ ist; Ein Modell muss eine solche Menge nicht aufweisen, und nur Unterschiede zwischen den Beiträgen – also Anpassungen – sind von Bedeutung; und die *Anpassung* von einem Vergleichsobjekt auf das Bewertungsobjekt ist der Betrag, der zum Beitrag des Vergleichsobjekts addiert wird, um ihn dem des Bewertungsobjekts anzugleichen, was der Summe aus Einheitspreis mal Breite über die Bereiche zwischen den beiden Mengen entspricht. Die so definierte Einheitspreis ist die des Modells; sie wird erst dann als die vom Markt gezahlte Einheitspreis interpretiert, wenn das Modell als faire Darstellung der Umsätze akzeptiert wurde, und die auf diesen Nachtrag folgenden Einträge befassen sich damit, wann diese Akzeptanz gerechtfertigt ist.

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.

Warum es diese Ergänzung gibt

Einem Schätzer, der Anpassungen an einem MARS-Modell in earthUI (Craytor 2026) oder direkt im R-Paket earth (Milborrow 2024a) vornimmt, wird eine Liste mit Begriffen wie den folgenden bereitgestellt:

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

Das wird von niemandem angegeben. Was in die Verkaufstabelle einfließt, ist die Anpassung: der Wertbeitrag des Bewertungsobjekts abzüglich des Wertbeitrags des Vergleichsobjekts. Die Preise pro Quadratfuß werden im Bericht erörtert, als Schritt zur Begründung des MARS-Modells. Der Weg von den gedruckten Bedingungen zu diesen Zahlen ist kurz, aber es gibt einen Schritt, bei dem manuell erstellte Tabellen Fehler aufweisen können, und der Sinn der Verwendung eines Modells geht verloren, wenn die daraus resultierenden Zahlen einem Prüfer nicht erklärt werden können.

Die einfache Fassung: vier Bereiche, vier Einheitspreise

Eine lineare Regression, bei der die Wohnfläche als linearer Term einfließt, ergibt einen einheitlichen Preis pro Quadratfuß für den gesamten Markt; genau das tun Regressionen in der traditionellen Wertermittlung, und genau das ist in dieser Ergänzung unter „gewöhnlicher Regression“ zu verstehen. MARS lässt den Preis pro Quadratfuß dort variieren, wo die Verkaufszahlen eine Veränderung anzeigen. Auf dem Küstenmarkt des Beispiels kann die Wohnfläche bei einer Fläche von bis zu 1.200 Quadratfuß einen Wert von \$600 pro Quadratfuß haben, \$1,000 pro Quadratfuß von 1.200 bis 2.000, \$750 von 2.000 bis 2.800 und nur \$100 darüber. Das ist der gesamte Unterschied, und ein Makler versteht ihn auf Anhieb: Der Markt zahlt am meisten für den Raum, der aus einem kleinen Haus ein Einfamilienhaus macht, weniger für den nächsten Raum und fast nichts für Raum, der über das hinausgeht, was die Käufer in der Nachbarschaft wollen.

Angesichts der Werte ist alles andere reine Arithmetik, die ein Gutachter ohnehin bereits durchführt. Ein Vergleichsobjekt mit 1.800 Quadratfuß und ein Bewertungsobjekt mit 2.400 Quadratfuß unterscheiden sich um 600 Quadratfuß, und der Mittelwert bei 2.000 liegt dazwischen. Die Anpassung, die zum Vergleichsobjekt hinzugerechnet wird, ist der angepasste Wert des zu bewertenden Objekts abzüglich desjenigen des Vergleichsobjekts, $\$2,620,000 - \$2,120,000 = \$500,000$, was auch dem Wertbeitrag der Wohnfläche des zu bewertenden Objekts abzüglich desjenigen des Vergleichsobjekts entspricht, $\$1,100,000 - \$600,000$. Die Einheitspreise erklären dies: Die Verschiebung überschreitet 200 Quadratfuß bei \$1,000, dann 400 Quadratfuß bei \$750, was insgesamt \$500,000 ergibt. Abbildung 3 veranschaulicht dies anhand eines geschätzten Werts (des Preises, den das Modell für ein Haus mit diesen Merkmalen angibt), der von \$1,280,000 bei 800 Quadratfuß – dem kleinsten Haus in den Daten – bis zu etwa \$3 million bei 3.500 – dem größten – reicht, wobei der größte Teil des Anstiegs vor 2.800 stattfindet.

Der einzige Wert, den eine Gerade durch die beiden Punkte für diese Bewegung hätte verwenden können, ist derjenige, der sie verbindet: \$500,000 bei 600 Quadratfuß, was etwa \$833 pro Quadratfuß entspricht: die *Sekantenrate* des Paares, die effektive oder gemittelte Einheitspreis für diese eine Bewegung: der Durchschnitt der Einheitspreise des Modells, gewichtet nach der Breite jedes

durchquerten Bereichs. Er ist kein Einheitspreis des Modells, und für das nächste Paar wäre es ein anderer Durchschnitt. Eine gewöhnliche Regression mit einem einzigen Einheitspreis für die Wohnfläche muss für jedes Paar diesen einen Wert verwenden und kann daher diese paarweisen Durchschnittswerte nicht wiedergeben; genau das bewirkt ein einziger Einheitspreis: Er bildet den Durchschnitt über eine Veränderung, die laut Modell auf dem Markt stattfindet.

Ein Leser, der an dieser Stelle innehält, hat die Methode verstanden: Man berechnet die Differenz der beiden Beiträge, erklärt diese dann, indem man den Einheitspreis für jeden Bereich abliest, den Vergleichswert entsprechend diesen Einheitspreisen auf das Bewertungsobjekt anwendet und die Werte addiert. Was nun folgt, richtet sich an den Leser, der wissen möchte, warum die Geschwindigkeiten so sind, wie sie sind, wo die einzige Falle liegt, was sich ändert, wenn Funktionen zusammenwirken, und wie ein Programm die Berechnungen überprüft; die Geschwindigkeitstabelle und die Anpassung ändern sich dabei nicht.

Die genauere Fassung: Scharnierfunktionen und das Vorzeichen, das übersehen wird

Bezeichnen wir den Graphen des Wertbeitrags einer Variablen über ihren Definitionsbereich als *Variablengraphen*. Es handelt sich dabei nicht um eine Kurve im alltäglichen Sinne, also um etwas, das sich sanft biegt; vielmehr ist es ein segmentierter linearer Graph, der zwischen den Knoten geradlinig verläuft und sich nur an diesen biegt. MARS baut ihn aus *Scharnieren* auf; Friedman (1991) beschrieb sie als abgeschnittene Potenzfunktionen, und „Scharnier“ ist der Name, auf den sich die spätere Literatur geeinigt hat (Milborrow, o. J.); Hastie, Tibshirani und Friedman (2009, §9.4) bezeichnen die beiden an einem Knoten als „reflektiertes Paar“.

Definition (Scharnier). Ein Scharnier ist eine Funktion einer Variablen x mit einer einzigen Bruchstelle, die als *Knoten* t bezeichnet wird und deren Wert auf der einen Seite des Knotens null ist und auf der anderen Seite dem Abstand vom Knoten entspricht. An jedem Knoten gibt es zwei davon, einen für jede Seite, wie in Abbildung 1 dargestellt:

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

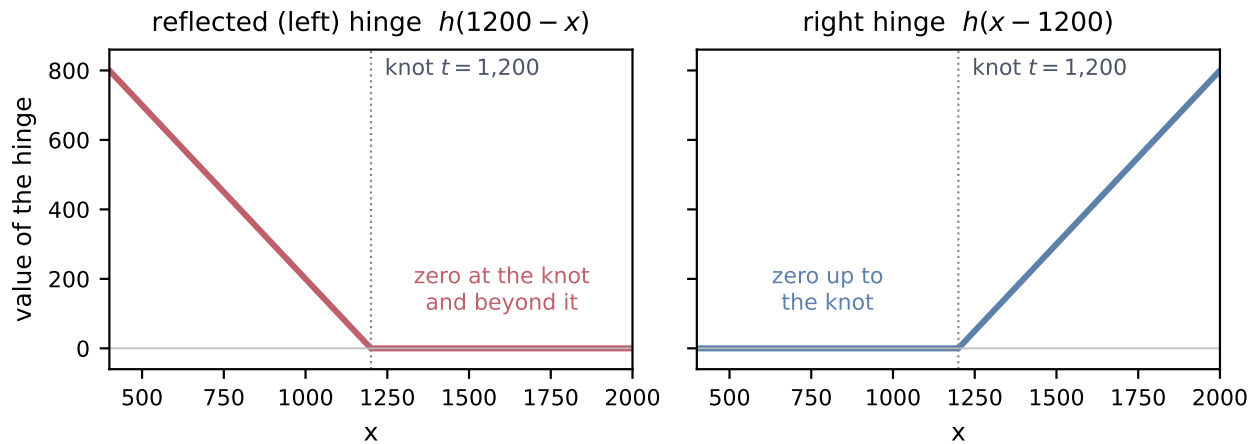
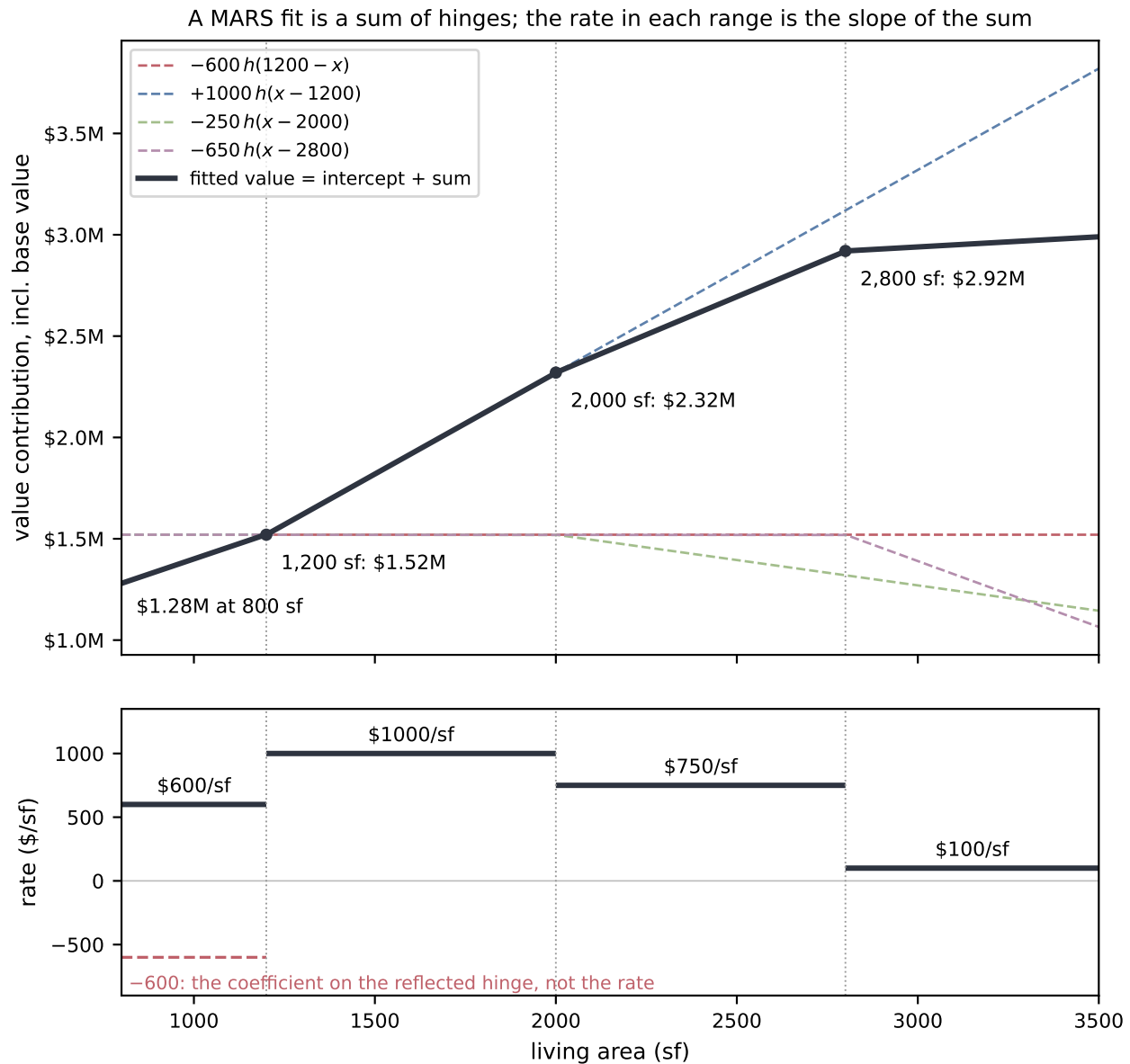


Figure 1: Die beiden Scharnierfunktionen mit dem Knoten bei $t = 1,200$: das gespiegelte (linke) Scharnier $h(1200 - x)$, das von links nach rechts bis zum Knoten auf Null abnimmt und jenseits des Knotens den Wert Null annimmt; das rechte Scharnier $h(x - 1200)$, das bis zum Knoten den Wert Null annimmt und jenseits des Knotens von links nach rechts ansteigt. Dies sind die reinen Scharnierfunktionen; in einem Modell wird jede mit einem Koeffizienten multipliziert, der die Richtung umkehren kann.

Das erste, $h(x - t)$, ist das *rechte* Scharnier: Null bis zum Knoten und jenseits davon von links nach rechts ansteigend. Das zweite, $h(t - x)$, ist das *gespiegelte* (oder linke) Scharnier: von links nach rechts bis zum Knoten abfallend und am Knoten sowie jenseits davon Null. Beide sind im gleichen Sinne Scharniere; sie unterscheiden sich lediglich darin, auf welcher Seite des Knotens sie ungleich Null sind. Diese Richtungen beschreiben die reinen Scharniere, die in Abbildung1 aus der Achse „ x “ abgeleitet wurden, und nichts weiter. In einem Modell wird ein Scharnier mit einem Koeffizienten multipliziert, der negativ sein kann, und zu anderen Termen addiert; daher kann ein auf einem rechten Scharnier basierender Term von links nach rechts abnehmen und ein auf einem gespiegelten Scharnier basierender Term zunehmen, und der Variablengraphen trifft auf einen Knoten in der Höhe, die ihm die anderen Terme zuweisen. Jede in diesem Anhang angegebene Richtung wird von links nach rechts gelesen. Scharniere werden nicht für sich allein verwendet. In einem MARS-Modell wird jedes Scharnier mit einem Koeffizienten multipliziert und die Terme addiert, und aus diesen Koeffizienten – nicht aus einem einzelnen Scharnier – werden die Steigungen des Modells, die Einheitspreise, berechnet: kumulativ, Bereich für Bereich, wobei der Einheitspreis jedes Bereichs der Einheitspreis des vorherigen Bereichs ist, angepasst um die Koeffizienten der Scharniere, deren Knoten zwischen ihnen liegen. Das Scharnier ist der einzige Baustein, über den MARS für eine einzelne Variable verfügt; jede Biegung in einem Variablendiagramm ist ein Knoten eines Scharniers, und jeder gerade Abschnitt ist eine Summe von Termen, die an dieser Stelle entweder null oder linear sind. **Der entscheidende Punkt, der sich durch alles Folgende zieht: Der Koeffizient eines Scharniers ist keine Steigung. Es ist eine Anpassung der Steigung, der Betrag, der am Knoten dieses Scharniers zu der Steigung addiert wird, die von links ankommt.** Von links nach rechts gelesen beginnt ein Variablengraf mit der Steigung seines ersten Segments, und jeder Knoten passt diese Steigung durch die dort sitzenden Koeffizienten an. Das ist die gesamte Leseregel, und die folgenden Sätze 1 und 2 sind ihr Beweis. Das Modell besteht aus dem Achsenabschnitt plus jedem Koeffizienten multipliziert mit seinem Scharnier, und die Steigung in jedem Bereich ist die Steigung dieser Summe.



```

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: Die vier Schwellenwerte aus dem Beispiel, ihre Summe und der Einheitspreis in jedem Bereich. Der Koeffizient -600 für den reflektierten Schwellenwert $h(1200 - \text{gla})$ ist so gekennzeichnet, dass der Leser ihn möglicherweise mit dem Einheitspreis verwechseln könnte; der Einheitspreis unterhalb von 1.200 Quadratfuß beträgt $+600$.

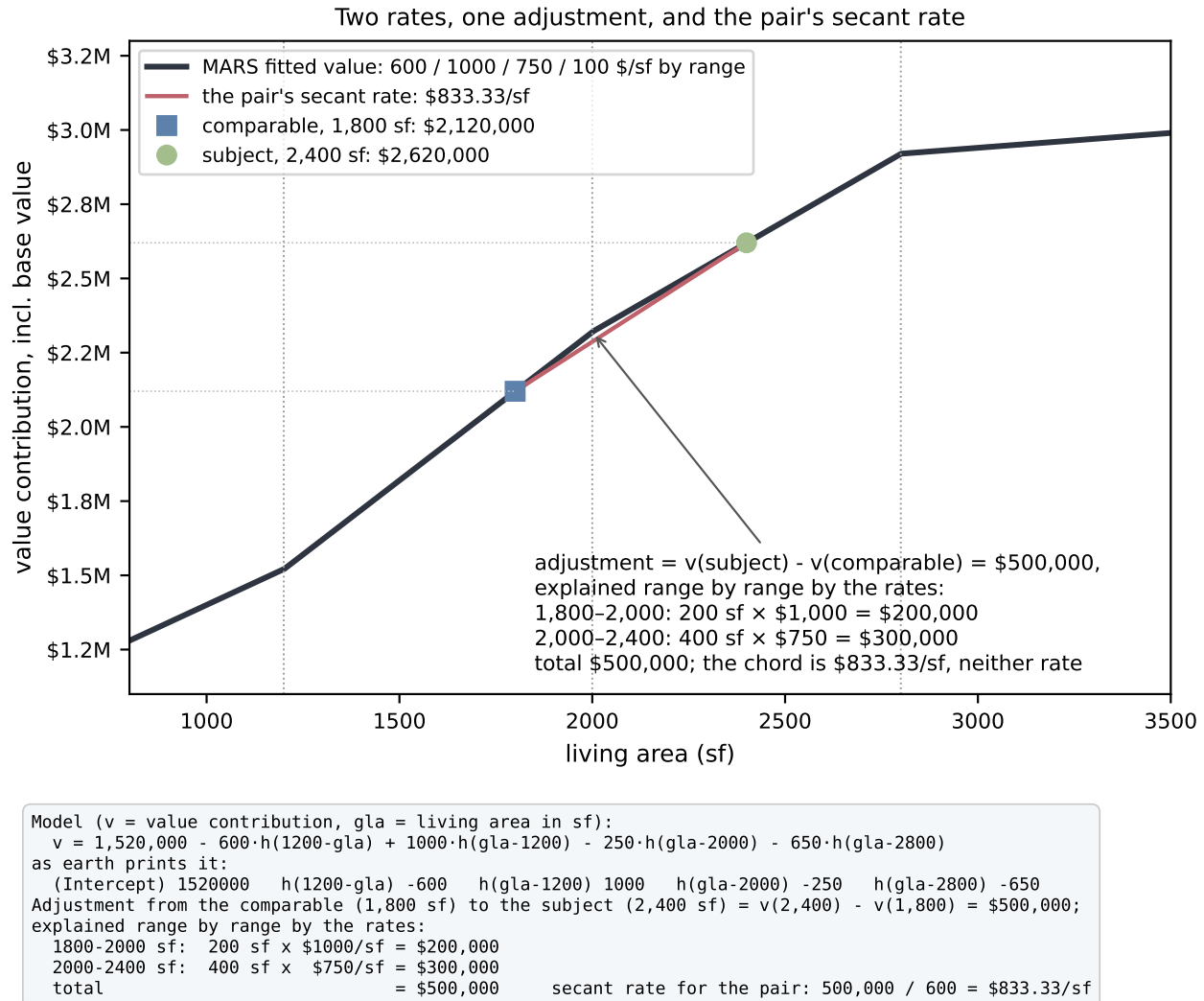


Figure 3: Die Anpassungen des Werts, wobei die Vergleichsimmobilie 1.800 Quadratfuß und das Bewertungsobjekt 2.400 Quadratfuß groß ist. Die Anpassung, also die Differenz der beiden Beiträge, beträgt \$500,000, und die geltenden Einheitspreise erklären dies Bereich für Bereich; die Sehne, die die beiden Punkte verbindet – etwa \$833 pro Quadratfuß – ist der Sekantenrate des Paares, der einzelne Einheitspreis, den eine Gerade für dieses Paar und kein anderes benötigen würde; eine gewöhnliche Regression muss für jedes Paar einen eigenen Einheitspreis verwenden.

Betrachten Sie die oben aufgeführten Werte; Abbildung 2 stellt jeden dieser Werte zusammen mit seiner Summe dar, darunter jeweils der Einheitspreis für den jeweiligen Bereich, und Abbildung 3 veranschaulicht, wie diese Einheitspreise die Anpassung zwischen einem Vergleichswert und einem Bewertungsobjekt erklären. Knoten bei 1.200, 2.000 und 2.800 bilden vier Bereiche. In jedem Bereich ist ein rechtes Scharnier *aktiv*, wenn sein Knoten auf oder links vom Bereich liegt, und ein gespiegeltes Scharnier, wenn sein Knoten auf oder rechts davon liegt – ein rechtes Scharnier wirkt in jedem Bereich jenseits seines Knotens, ein gespiegeltes Scharnier in jedem Bereich davor. Der Einheitspreis wird aus den Koeffizienten der aktiven Scharniere berechnet: Man addiert die Koeffizienten der aktiven rechten Scharniere und subtrahiert die Koeffizienten der aktiven gespiegelten Scharniere. Die daraus resultierende Tabelle gibt der Einheitspreis des Modells für jeden Bereich an, wobei die anderen Merkmale auf einen festen Satz von Werten festgelegt sind (das Programm verwendet die Werte des Bewertungsobjekts); in einem Modell ohne Wechselwirkungen ändert diese Einschränkung nichts, und im folgenden Fall 2 ändert sie alles:

Fläche (sf)	aktive Scharniere	Preis (\$/sf)
unter 1.200	$h(1200 - \text{gla})$ nur	$-(-600) = \mathbf{+600}$
1.200 bis 2.000	$h(\text{gla} - 1200)$	1000
2.000 bis 2.800	$h(\text{gla} - 1200), h(\text{gla} - 2000)$	$1000 - 250 = \mathbf{750}$
über 2.800	alle drei rechten Scharniere	$1000 - 250 - 650 = \mathbf{100}$

Die erste Zeile ist der Schritt, der übersehen wird. Ein Koeffizient von -600 bei einem reflektierten Scharnier entspricht nicht einem Einheitspreis von $-\$600$ pro Quadratfuß. Der Scharnierwert $h(1200 - \text{gla})$ gibt an, um wie viel die Wohnfläche eines Hauses unter 1.200 Quadratfuß liegt, und -600 mal dieser Fehlbetrag ergibt einen Preisnachlass: $\$240,000$ für ein 800-Quadratfuß-Haus, der bei 1.200 auf null sinkt. Jeder zusätzliche Quadratfuß reduziert $\$600$ des Preisnachlasses, sodass der Einheitspreis $+\$600$ beträgt. (Bei einem positiven Koeffizienten sorgt dasselbe Scharnier aufgrund seiner geringeren Größe für einen Mehrwert, und sein Satz ist negativ.) Die Regel, einmal formuliert: In einem Bereich ist der Einheitspreis die Summe der Koeffizienten der aktiven rechten Scharniere *minus* der Summe der Koeffizienten der aktiven gespiegelten Scharniere, zuzüglich des Koeffizienten eines linearen Terms (ein Merkmal, das als gerade Linie ohne Knoten eingegeben wird), falls das Modell einen solchen enthält. Damit geht eine Konvention einher: Genau an einem Knoten wird der Einheitspreis von rechts her genommen, sodass ein rechtes Scharnier ab seinem Knoten zählt und ein gespiegeltes Scharnier nur streng unterhalb seines Knotens; der Einheitspreis bei 1.200 Quadratfuß beträgt daher $\$1,000$, der Einheitspreis des dort beginnenden Bereichs, und das Programm liest sie auf dieselbe Weise aus.

Die Überprüfung, ob eine Tabelle korrekt erstellt wurde, erfolgt anhand der Anpassungen an den Knotenpunkten. Bei 1.200 Quadratfuß ist jeder Scharnier gleich Null und der Wert entspricht dem Achsenabschnitt: $\$1,520,000$; bei 2.000 beträgt er $\$1,520,000 + 1,000 \times 800 = \$2,320,000$; bei 2.800 $\$2,320,000 + 750 \times 800 = \$2,920,000$; und bei 800 Quadratfuß, dem kleinsten Haus in den Daten, beträgt er $\$1,520,000 - 600 \times 400 = \$1,280,000$. Von 800 bis 1.200 steigt der Wert von $\$240,000$, was einem Einheitspreis von $\$600$ entspricht – dies ist die erste Zeile; eine Tabelle, in der der Koeffizient als Einheitspreis gelesen würde, hätte den Wert von links nach rechts bis zum Knotenpunkt bei 1.200 abnehmend, doch die Werte zeigen, dass dies nicht der Fall ist. Abbildung2 stellt die vier Scharniere, deren Summe und der Einheitspreis in jedem Bereich dar, wobei der Koeffizient -600 dort markiert ist, wo ein Leser ihn möglicherweise mit dem Einheitspreis verwechseln könnte.

Zwei weitere Komplikationen, die das Programm bewältigen muss, und eine Handtabelle:

- **Wechselwirkungen.** „earth“ gibt in seinem Vorlauf (der Phase, in der Terme nacheinander

hinzugefügt werden; Friedman 1991, §3) ein Produkt wie „ $h(\text{gla} - 1500) \times \text{lot}$ “ aus, wenn der Einheitspreis für ein Merkmal von einem anderen abhängt: eine *Wechselwirkung*. Der Einheitspreis für „lot“ unterscheidet sich dann oberhalb und unterhalb von 1.500 Quadratfuß Wohnfläche, und die Einheitspreistabelle für ein Merkmal ist immer die Tabelle, *die sich aus* den anderen *ergibt*, wobei deren Werte beibehalten werden. Das Programm nimmt einen Punkt, einen Wert für jedes andere Merkmal, und berechnet jeden Einheitspreis an diesem Punkt.

- **Niveau versus Anpassung.** Der Achsenabschnitt und die Terme in anderen Merkmalen legen das Niveau fest, von dem aus der Beitrag gemessen wird; sie heben sich bei jeder Anpassung gegenseitig auf. Das Programm gibt Beiträge als die Terme des jeweiligen Merkmals allein an, sodass sich zwei Beiträge genau um den Anpassungswert unterscheiden.

Zwei Fälle: Modelle ohne und mit Wechselwirkungen

Die Regel aus dem vorangegangenen Abschnitt lässt sich in Form von zwei Sätzen formulieren, die im Folgenden bewiesen werden. Sie liefern eine klare Antwort auf eine Frage, die sich jeder Leser eines MARS-Modells früher oder später stellt: Von wie vielen der aufgeführten Terme hängt ein Maß tatsächlich ab? Die Antwort lässt sich in zwei Fälle unterteilen.

Fall 1: keine Wechselwirkungen

Vorgabe. $h(u) = \max(0, u)$, wobei u reelle Zahlen sind. Ein MARS-Modell ohne Wechselwirkungen auf den Merkmalen $x, z_1, \dots, z_q$ hat Anpassungen beim Wert

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

wobei β_0 der Achsenabschnitt ist, jedes g_k eine Funktion allein von z_k ist und g alle Terme in x zusammenfasst:

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

Hier sind $t_1 < t_2 < \dots < t_K$ die verschiedenen Knoten der Scharniere in x ; a_j ist der Koeffizient des rechten Scharniers bei t_j und b_j der des gespiegelten Scharniers bei t_j , wobei beide 0 sind, wenn das Modell keinen solchen Term enthält; und c ist der Koeffizient eines reinen linearen Terms $c \cdot x$, eines Merkmals, das als Gerade ohne Knoten eingegeben wird und das earth nur für Merkmale erzeugt, die in `linpreds` genannt sind, oder für jedes Merkmal, wenn `Auto.linpreds` aktiviert ist (siehe die Anmerkung nach Satz 1); $c = 0$, wenn kein solcher Term vorhanden ist, wie in jedem Modell dieses Anhangs. Setzen wir $t_0 = -\infty$ und $t_{K+1} = +\infty$ und bezeichnen wir $S_i = (t_i, t_{i+1})$ für $i = 0, 1, \dots, K$ als die *Segmente* von x ; t_i ist der linke Knoten von S_i und t_{i+1} sein rechter Knoten. Auf jedem S_i ist jeder Summand von g entweder identisch Null oder linear, sodass g auf S_i linear ist; seine Steigung dort, die auch $\partial f / \partial x$ auf S_i ist, da kein anderer Summand von f von x abhängt, ist die *Einheitspreis* auf S_i und wird als r_i geschrieben. Die Einheitspreistabelle von x ist die Liste $r_0, r_1, \dots, r_K$. (Wenn $K = 0$ ein Segment vorliegt, ist $(-\infty, +\infty)$, und die Tabelle ist der einzelne Einheitspreis $r_0 = c$.) An einem Knoten selbst hat g keine Ableitung: Seine linke und rechte Ableitung sind r_{i-1} und r_i , die sich um $a_i + b_i$ unterscheiden (Satz 2 unten), sodass der Wert, den eine Einheitspreistabelle unter t_i anzeigt, eine Konvention ist, die hier von rechts übernommen wurde, und kein Wert von g' .

Satz 1 (der Einheitspreis auf einem Segment). Unter Verwendung der in diesem Rahmen verwendeten Notation gilt für jedes $i \in \{0, 1, \dots, K\}$:

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

, wobei eine leere Summe als 0 definiert ist. Insbesondere hängt r_i von höchstens K der $2K$ Scharnierkoeffizienten ab – einem pro Knoten – sowie von c .

Beweis von Satz 1. Man wähle i und einen Punkt $x \in S_i$, sodass $t_i < x < t_{i+1}$ gilt. Betrachten wir jeden Summanden von g in einer Umgebung von x , die klein genug ist, um innerhalb von S_i zu liegen. Der nackte Term $c \cdot x$ hat die Ableitung c . Für das rechte Scharnier bei t_j gilt: Wenn $j \leq i$, dann ist $t_j \leq t_i < x$, also gilt $h(x - t_j) = x - t_j$ in der Umgebung und seine Ableitung ist 1; wenn $j \geq i + 1$, dann ist $x < t_{i+1} \leq t_j$, also gilt $h(x - t_j) = 0$ dort und seine Ableitung ist 0. Für das gespiegelte Scharnier bei t_j gilt: Wenn $j \geq i + 1$, dann ist $x < t_j$, also gilt $h(t_j - x) = t_j - x$ und seine Ableitung ist -1 ; wenn $j \leq i$, dann ist $t_j < x$, also gilt $h(t_j - x) = 0$ und seine Ableitung ist 0. Eine endliche Summe von Funktionen, die in x differenzierbar sind, ist dort mit der Summe der Ableitungen differenzierbar, was

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

die angegebene Formel ergibt. Zur Zählung: Die Indexmengen $\{j : j \leq i\}$ und $\{j : j \geq i + 1\}$ sind disjunkt und erschöpfen gemeinsam $\{1, \dots, K\}$, sodass jeder Knoten t_j genau einen der Werte a_j und b_j beiträgt und niemals beide; daher treten höchstens K Scharnierkoeffizienten auf, und r_i ist an jedem Punkt von S_i gleich. ■

Die Grenze ist erreicht: Wenn für jedes j gilt, dass $b_j = 0$, dann nutzt das Segment ganz rechts S_K alle K des a_j .

Anmerkung (wenn „c“ ungleich Null ist). earth kann ein Merkmal als einzelne gerade Linie ohne Knoten darstellen – der Begriff „ $c \cdot x$ “ in der Einstellung – und zwar auf zwei verschiedene Arten, wobei der Unterschied von Bedeutung ist. „`linpreds`“, eine Liste benannter Merkmale wie „`linpreds = c("GLA")`“, erzwingt für jedes Merkmal in der Liste eine lineare Darstellung, selbst wenn eine Anpassung mit Knoten besser wäre; dies ist die Entscheidung des Analysten und kann zu Lasten der Anpassungsgüte gehen. „`Auto.linpreds`“, ein einzelner Schalter, der standardmäßig aktiviert ist, erzwingt nichts: Er ändert lediglich, wie „earth“ ein Merkmal darstellt, dessen Anpassungen sich als wirklich linear herausgestellt haben – ein einzelnes Scharnier mit seinem Knoten am kleinsten beobachteten Wert –, indem er es als Koeffizienten und Variablennamen anstelle eines Scharniers ausgibt; die Anpassungen bleiben dieselben. In beiden Fällen gibt „earth“ den Term als bloßen Variablennamen mit seinem Koeffizienten aus – „`221.4 * GLA`“ in der Zusammenfassungsfom von „earth“, wie im folgenden Beispiel. „Satz 1“ behandelt dies über „c“, was der Einheitspreis auf jedem Segment erhöht. Beispielsweise wird bei dem Aufruf „

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

“ jede Eigenschaft linear einbezogen (mit „ `degree = 1`“, sodass auch keine Produkte der linearen Terme angepasst werden) und in der Zusammenfassungsform von „earth“ (veranschaulichende Ausgabe, kein angepasstes Modell) Folgendes ausgegeben: „

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

“ Für die Wohnfläche ist dies die Einstellung mit „ $c = 221.4$ “ und ohne Scharniere, „ $K = 0$ “: Es gibt ein Segment, und Satz 1 liefert den einheitlichen Wert $r_0 = c = \$221.40$ pro Quadratfuß über den gesamten Datenbereich; Grundstücksgröße, Badezimmer und Schlafzimmer werden auf die gleiche Weise ausgewertet: \$11.80 pro Quadratfuß Grundstück, \$34,872 pro Badezimmer und \$17,641 pro Schlafzimmer. Ein solches Modell ist im Sinne dieser Ergänzung eine gewöhnliche Regression: ein Wert pro Merkmal. Über den Daten entspricht ein solcher Term einer Geraden mit einem einzigen rechten Scharnier, dessen Knoten am kleinsten beobachteten Wert liegt, und das Journal arbeitet mit `Auto.linpreds = FALSE`, sodass in seinen Modellen ein Merkmal, dessen Variablenverlauf über den Daten geradlinig verläuft, als dieser einzelne Scharnier erscheint, `h(gla-800) 700` zum Beispiel, der Knoten am kleinsten Haus in den Daten, mit $c = 0$, $K = 1$ und $a_1 = 700$: Der Einheitspreis beträgt 0 unterhalb des Knotens, auf einem Abschnitt, auf dem keine Verkäufe stattfinden, und 700 oberhalb davon, sodass die Einheitspreistabelle über den Daten eine Zeile aufweist. Daraus folgt eine Konsequenz; die Änderung des Einheitspreises an einem Knoten ist Satz 2 unten.

Korollar 1 (das gespiegelte Vorzeichen). Ein gespiegeltes Scharnier mit dem Koeffizienten b_j trägt auf jedem Segment links von t_j mit $-b_j$ zum Einheitspreis bei und auf jedem Segment rechts davon mit 0. Im Beispiel $b_1 = -600$ bei $t_1 = 1,200$ trägt $+600$ zu r_0 bei, was den Kern der Falle bildet, um die sich diese Ergänzung dreht.

Beweis. Man entnehme die zweite Summe aus Satz 1. ■

Ein Beispiel mit sieben Knoten, und die Annahme, an der Leser scheitern

Die Zählweise ist von Bedeutung, da fast jeder bei einem gedruckten MARS-Modell von einer bestimmten Annahme ausgeht: dass der neben einem Knoten aufgedruckte Koeffizient die Steigung des Segments ist, dessen linker Knoten dieser ist. Das ist jedoch nicht der Fall. **Der Koeffizient eines Scharniers ist eine Anpassung der von links kommenden Steigung, nicht die Steigung selbst**, und die Steigung eines Segments ist daher die Steigung des ersten Segments zuzüglich der laufenden Summe aller derartigen Anpassungen an seinem linken Knoten und an jedem weiter links liegenden Knoten. Ein Leser, der von dieser falschen Annahme ausgeht, versucht die Berechnung im Kopf, in einem kleinen Programm oder mit einer Tabellenkalkulationsformel durchzuführen, stellt fest, dass die Zahlen nicht mit den Werten nach den Anpassungen übereinstimmen, gibt auf und kommt zu dem Schluss, dass MARS undurchsichtig ist. Es ist nicht undurchsichtig; es ist kumulativ.

Satz 2 (die laufende Summe). Unter Verwendung der Notation aus Fall 1 gilt für jedes $i \in \{1, \dots, K\}$:

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

Folglich gilt $r_i = r_0 + \sum_{j=1}^i (a_j + b_j)$ für jedes i , wobei $r_0 = c - \sum_{j=1}^K b_j$ der Einheitspreis auf dem Segment ganz links und $r_K = c + \sum_{j=1}^K a_j$ der Einheitspreis auf dem Segment ganz rechts ist.

Beweis von Satz 2. Sei $i \in \{1, \dots, K\}$ festgelegt. Nach Satz 1 gilt:

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

Subtrahiert man, so hebt sich der Term c auf; die beiden Summen der a unterscheiden sich um den einzelnen Term a_i ; und da $\sum_{j=i}^K b_j = b_i + \sum_{j=i+1}^K b_j$ gilt, unterscheiden sich die beiden Summen der b um den einzelnen Term b_i , der in r_{i-1} subtrahiert wird, in r_i jedoch nicht. Daher gilt $r_i - r_{i-1} = a_i + b_i$. Fügt man diese Identität für $j = 1, \dots, i$ anstelle von i hinzu, so lassen sich die linken Seiten zusammenfassen: $(r_1 - r_0) + (r_2 - r_1) + \dots + (r_i - r_{i-1}) = r_i - r_0$, was die Summenformel ergibt. Die Werte von r_0 und r_K sind Satz 1 unter $i = 0$, wo die Summe der a leer ist, und unter $i = K$, wo die Summe der b leer ist. ■

Folgerung 2 (die Leseregel). Im Rahmen von Fall 1 gilt für jedes $i \in \{1, \dots, K\}$, dass der Einheitspreis auf dem Segment, dessen linker Knoten t_i ist, $r_i = r_{i-1} + a_i + b_i$ beträgt: Die an einem Knoten angegebenen Koeffizienten werden zur von links kommenden Einheitspreis addiert und stellen nicht selbst der Einheitspreis des dort beginnenden Segments dar. Die beiden stimmen genau dann überein, $r_i = a_i + b_i$, wenn $r_{i-1} = 0$ gilt; insbesondere tritt dies bei einem Modell, dessen einzige Scharniere rechte Scharniere sind und das keinen linearen Term aufweist, bei $i = 1$ auf und gemäß Satz 2 an keinem späteren Knoten, es sei denn, die Koeffizienten an den davor liegenden Knoten summieren sich zu Null.

Beweis. Die erste Gleichheit ist Satz 2. Daraus folgt, dass „ $r_i = a_i + b_i$ “ genau dann gilt, wenn „ $r_{i-1} = 0$ “ gilt. Wenn „ $b_j = 0$ “ für alle „ j “ und „ $c = 0$ “ gilt, ergibt sich aus Satz 2, dass „ $r_0 = 0$ “ gilt, also „ $r_1 = a_1$ “ und „ $r_{i-1} = \sum_{j<i} a_j$ “, was nur dann „0“ ist, wenn die Summe dieser Koeffizienten Null ist. „ ■

Korollar 3 (die Anpassung als Integral). Im Rahmen von Fall 1 ist g an jedem Knotenpunkt stetig, da jeder Scharnierpunkt stetig ist; ebenso gilt dies für $f(\cdot, z)$ für jedes z ; genau das sorgt dafür, dass der Wert an einem Knotenpunkt derselbe ist, unabhängig davon, ob er von links oder von rechts erreicht wird, und ermöglicht es, die Einheitspreise von einer Größe zur anderen zu überführen. Seien x_c und x_s die Mengen von x für ein Vergleichsobjekt und ein Subjekt, wobei die anderen Merkmale für beide auf denselben Werten gehalten werden, sodass $g(x)$ gilt; die Terme von f , die x beinhalten, sind der Beitrag von x bei der Menge x . Schreiben wir $u_1 = \min(x_c, x_s)$ und $u_2 = \max(x_c, x_s)$ sowie $|I|$ für die Länge eines Intervalls I . Dann ist die Anpassung $g(x_s) - g(x_c)$, der Betrag, um den der Wertbeitrag des Vergleichsobjekts erhöht wird, damit er dem des Bewertungsobjekts entspricht, gleich

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

In Worten: Die Anpassung ist das Integral des Einheitspreises über die beiden Größen, und da der Einheitspreis auf jedem Segment konstant ist, ist das Integral die Summe aus Einheitspreis mal Breite über die auf dem Weg durchlaufenen Segmente.

Beweis. g ist stetig und stückweise linear mit endlich vielen Stücken, also absolut stetig; daher folgt aus dem Fundamentalsatz der Analysis, dass $g(x_s) - g(x_c) = \int_{x_c}^{x_s} g'(u) du$ gilt, wobei die Ableitung überall außer an den Knoten, einer endlichen Menge, existiert. Auf jedem S_i ist der Integrand gemäß Satz 1 gleich r_i , sodass $\int_{u_1}^{u_2} g'(u) du$ die Summe über i von r_i mal der Länge von $S_i \cap [u_1, u_2]$ ist, wobei sich nicht überschneidende Segmente die Länge 0 beitragen; und $\int_{x_c}^{x_s} = \text{sign}(x_s - x_c) \int_{u_1}^{u_2}$ aufgrund der Orientierung des Integrals. ■

Die Einheitspreistabelle lautet g' ; die Anpassung, die direkt als $g(x_s) - g(x_c)$ berechnet wird, entspricht ihrem Integral zwischen den beiden Größen; und die Sätze 1 und 2 besagen, woraus sich g' zusammensetzt. In der Praxis wird die Anpassung stets als diese Differenz der Beiträge berechnet, wofür überhaupt keine Einheitspreise benötigt werden; die Folgerung erklärt, warum die Einheitspreise, Bereich für Bereich gelesen, diese Differenz wiedergeben, was sie zu einer Erklärung der Anpassung und zu einer Überprüfung derselben macht, nicht jedoch zu einer Methode, sie zu berechnen.

Nehmen wir die Wohnfläche in einem Datensatz, dessen kleinstes Haus 800 Quadratfuß groß ist und sieben Knoten aufweist, bei 1.200, 1.600, 2.000, 2.400, 2.800, 3.200 und 3.600 Quadratfuß, sowie neun Terme im Modell, von denen zwei Scharniere widerspiegeln, die jeweils mit einem rechten Scharnier am selben Knoten gepaart sind: bei 1.200 und bei 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

Sieben Knoten ergeben acht Segmente. Der Einheitspreis jedes Segments ist die Summe der Koeffizienten der aktiven rechten Scharniere abzüglich der Summe der Koeffizienten der aktiven gespiegelten Scharniere:

Segment (sf)	aktive Begriffe	Preis (\$/sf)
800 bis 1.200	$h(1200 - gla)$, $h(2400 - gla)$	$500 + 200 = 700$
1.200 bis 1.600	$h(gla - 1200)$, $h(2400 - gla)$	$700 + 200 = 900$
1.600 bis 2.000	Rechtsanschlüsse bei 1.200, 1.600; $h(2400 - gla)$	$700 - 100 + 200 = 800$
2.000 bis 2.400	Rechtsanschlag bei 1.200 bis 2.000; $h(2400 - gla)$	$700 - 100 + 150 + 200 = 950$
2.400 bis 2.800	rechts angeschlagen bei 1.200 bis 2.400	$700 - 100 + 150 - 250 = 500$
2.800 bis 3.200	Rechtsanschlag bei 1.200 bis 2.800	$500 - 300 = 200$
3.200 bis 3.600	Rechtsanschlag bei 1.200 bis 3.200	$200 - 150 = 50$
über 3.600	alle sieben rechten Scharniere	$50 + 50 = 100$

Drei Werte, die aus der Tabelle abzulesen sind. Der Einheitspreis des Segments ganz rechts, $700 - 100 + 150 - 250 - 300 - 150 + 50 = 100$, kombiniert die Koeffizienten aller sieben rechten Scharniere, eines pro Knoten, was die Grenze von Satz 1 mit $K = 7$ darstellt; die beiden gespiegelten Scharniere spielen dabei keine Rolle, da ein gespiegeltes Scharnier rechts von seinem eigenen Knoten flach ist und das Segment ganz rechts rechts von jedem Knoten liegt. Das Segment ganz links, von 800 bis zu seinem rechten Knoten bei 1.200, ist das Spiegelbild: Es zählen nur die gespiegelten Scharniere mit umgekehrten Vorzeichen, da jeder Knoten an seinem rechten Knoten oder rechts davon liegt. Und an jedem Knoten, der ein Paar trägt, ist der Schritt die Summe der beiden Koeffizienten: bei 1.200 beträgt er $-500 + 700 = +200$ von 700 bis 900; bei 2.400 sinkt der Einheitspreis von 950 auf 500, eine Änderung von -450 , was genau der Summe der beiden Koeffizienten an diesem Knoten entspricht, -250 und -200 : Das Überqueren des Knotens schaltet das

rechte Scharnier ein und das gespiegelte Scharnier aus, und beide Änderungen wirken in dieselbe Richtung.

Derselbe Wert auf dem Segment ganz rechts lässt sich auch durch „Gehen“ erreichen: Man beginnt bei 700 auf dem ersten Segment, geht von 800 bis 1.200 und addiert an jedem Knotenpunkt die Summe der dort befindlichen Koeffizienten: +200, -100, +150, -450, -300, -150, +50. Diese Route berührt alle acht Segmente und alle neun Terme und landet auf demselben Wert 100. Die direkte Summe ist kürzer und wird vom Programm ausgeführt; den Durchlauf sollte der Leser einmal manuell durchführen, um zu erkennen, dass der Koeffizient neben einem Knoten einen Schritt und keine Ebene darstellt. Der „-450“ bei 2.400 besteht aus zwei Halbsprüngen: Beim Überqueren des Knotens bleibt das gespiegelte Scharnier $h(2400 - gla)$ zurück, und mit ihm das +\$200, das es unterhalb des Knotens hinzugefügt hat; und es führt das rechte Scharnier $h(gla - 2400)$ ein, dessen -250 dort beginnt. $-200 - 250 = -450$, und der Einheitspreis fällt von \$950 auf \$500.

Abbildung 4 veranschaulicht dasselbe Prinzip anhand eines Modells mit fünf Knoten, das klein genug ist, um es mit bloßem Auge zu überprüfen. Die Balken stellen die an den Knoten angegebenen Koeffizienten dar; die Treppe zeigt die Geschwindigkeit auf jedem Segment an; die Treppe beginnt bei der Geschwindigkeit des ersten Segments, \$900, die durch das reflektierte Scharnier festgelegt wird, und jede Stufe entspricht der Summe der Balken an diesem Knoten. Der Einheitspreis auf dem Segment ganz rechts, $650 - 250 + 150 - 200 - 150 = 200$, nutzt alle fünf rechten Scharniere; das reflektierte Scharnier bei 1.200, das als -900 angegeben ist, setzt das erste Segment auf +900 und berührt nichts anderes.

Zwei Aspekte zur Form dieser Grafik sind erwähnenswert, da sie sich eher auf Häuser als auf MARS beziehen. Erstens beginnt die Achse bei 800 Quadratfuß und nicht bei Null, und die Kurve beginnt an der Achse: Jede Kurve eines Modells mit Anpassungen in dieser Beilage verläuft – genau wie die von der Modellierungssoftware erstellten Diagramme – vom kleinsten Haus in den Daten bis zum größten und zeigt außerhalb dieses Bereichs nichts an. Jedes Marktgebiet hat ein kleinstes Haus, unterhalb dessen keine Verkäufe stattfinden, und die variable Grafik zeigt nichts darunter: nicht Null, sondern keinen Wert, den die Verkäufe stützen; die Formel würde dort dennoch eine Zahl liefern, und diese Zahl ist eine Extrapolation. Die Abbildungen verwenden eine Untergrenze von 800 Quadratfuß, doch diese Untergrenze ist ein Merkmal des Marktgebiets und keine Konstante: Verschiedene Marktgebiete weisen unterschiedliche Mindestgrößen für Immobilien auf, die im MLS erscheinen, und der Datensatz – nicht der Gutachter – gibt an, wo diese liegt. Zweitens variiert die Form von Markt zu Markt: Im Beispiel an der Küste liegt der steilste Bereich zwischen 1.200 und 2.000 Quadratfuß, und in Abbildung 4 ist dies das erste Segment. In einer hochpreisigen Gegend wie Atherton oder Hillsborough kann das steilste Segment der Kurve der Wohnfläche (GLA) erst jenseits von 4.000 oder 5.000 Quadratfuß liegen, wo sich die Käufer befinden. Und ein Stadtteil, in dem ein Bauträger in einem bestimmten Zeitraum Häuser einer bestimmten Größe mit Mängeln errichtet hat, könnte einen starken Rückgang des Einheitspreises in einem mittleren Bereich aufweisen, wie es die erstellte Grafik zwischen 1.600 und 2.000 zeigt. Über ein größeres Marktgebiet hinweg sind die variablen Kurven unterschiedlich, und die Sättigungstabelle gibt wieder, wie der Gutachter die jeweilige Kurve interpretiert. Die Abbildungen 5 und 6 zeigen zwei dieser Kurven so, wie sie in der Praxis dargestellt würden, sodass der Leser jede Sättigungstabelle nach der Regel erstellen und anhand der Treppenkurve überprüfen kann.

Fall 2: Wechselwirkungen

Ein Interaktionsterm ist ein Koeffizient, multipliziert mit einem Produkt aus Faktoren verschiedener Variablen, wobei jeder Faktor entweder ein Scharnier oder die Variable selbst ist, zum Beispiel $\gamma \cdot h(gla - 1500) \cdot lot$. Alles aus Fall 1 bleibt erhalten, bis auf ein Wort: Der Einheitspreis eines

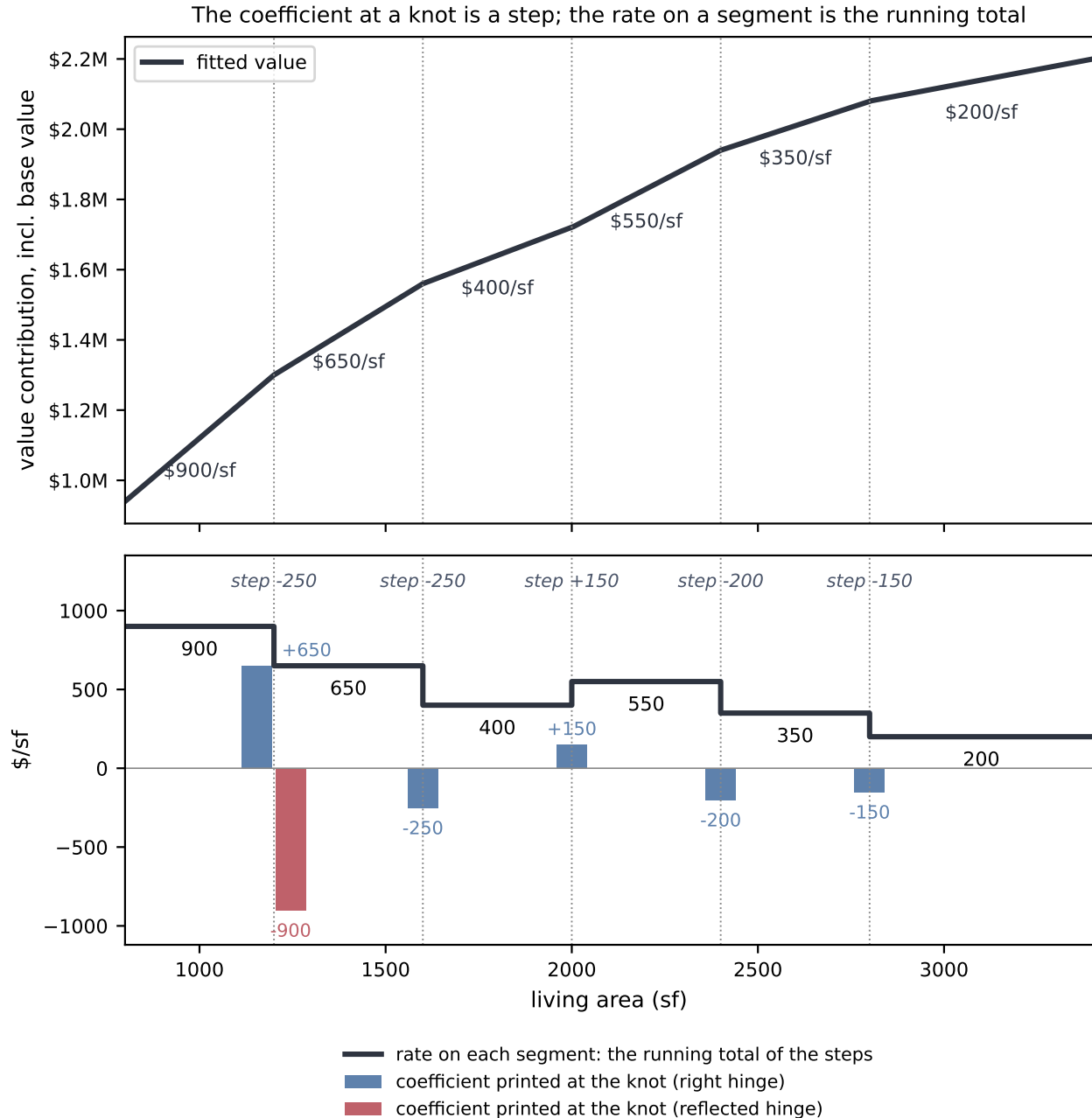


Figure 4: Ein Fünf-Knoten-Modell auf der Grundlage von Daten, die bei 800 Quadratfuß beginnen: $-900 h(1200 - \text{gla}) + 650 h(\text{gla} - 1200) - 250 h(\text{gla} - 1600) + 150 h(\text{gla} - 2000) - 200 h(\text{gla} - 2400) - 150 h(\text{gla} - 2800)$. Oben ist der Wert mit den Anpassungen und der auf jedem Segment angegebenen Einheitspreis dargestellt; unterhalb von 800 Quadratfuß gibt es keinen Wert, da es keine solchen Häuser gibt. Unten ist der Koeffizient an jedem Knoten als Balken und der Einheitspreis auf jedem Segment als Treppe dargestellt: Jede Stufe der Treppe entspricht der Summe der Balken an diesem Knoten (ein Balken hier, zwei dort, wo ein Knoten ein Paar enthält, Satz 2), daher ist der Einheitspreis auf einem Segment der Einheitspreis des ersten Segments plus die laufende Summe der Stufen links davon, nicht der Koeffizient daneben: Jeder Koeffizient passt die von links kommenden Einheitspreise an.

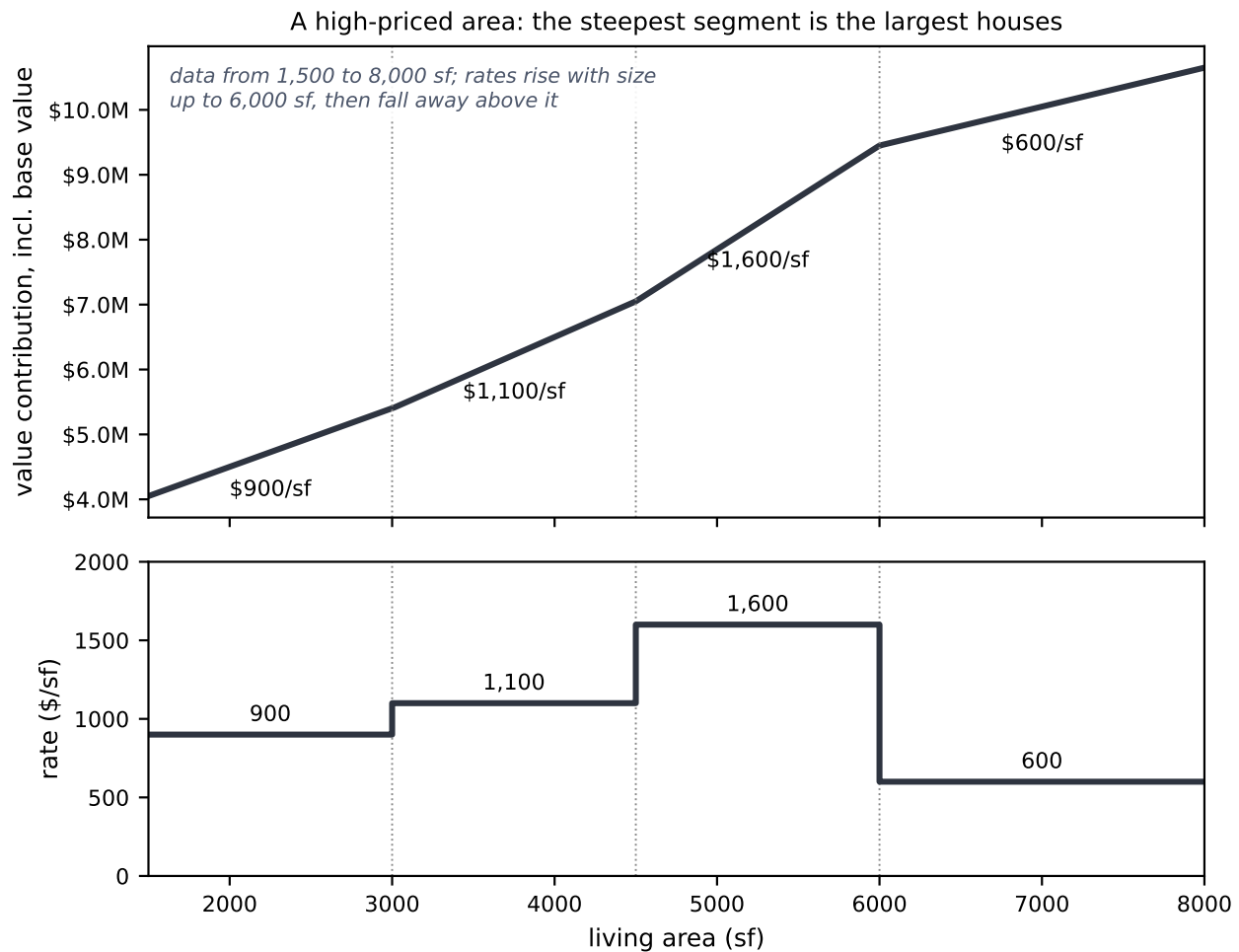


Figure 5: Eine hochpreisige Gegend, mit Flächen von 1.500 bis 8.000 Quadratfuß: $-900 h(3000 - \text{gla}) + 1100 h(\text{gla} - 3000) + 500 h(\text{gla} - 4500) - 1000 h(\text{gla} - 6000)$, Schnittpunkt mit \$5,400,000 am ersten Knotenpunkt. Preise \$900 , \$1,100 , \$1,600 und \$600 pro Quadratfuß: Das steilste Segment liegt bei 4.500 bis 6.000 Quadratfuß, wo sich die Käufer befinden; das erste Segment ist flacher, und ab 6.000 fällt der Preis ab. Die Regel bleibt unverändert; nur der Markt hat sich verändert.

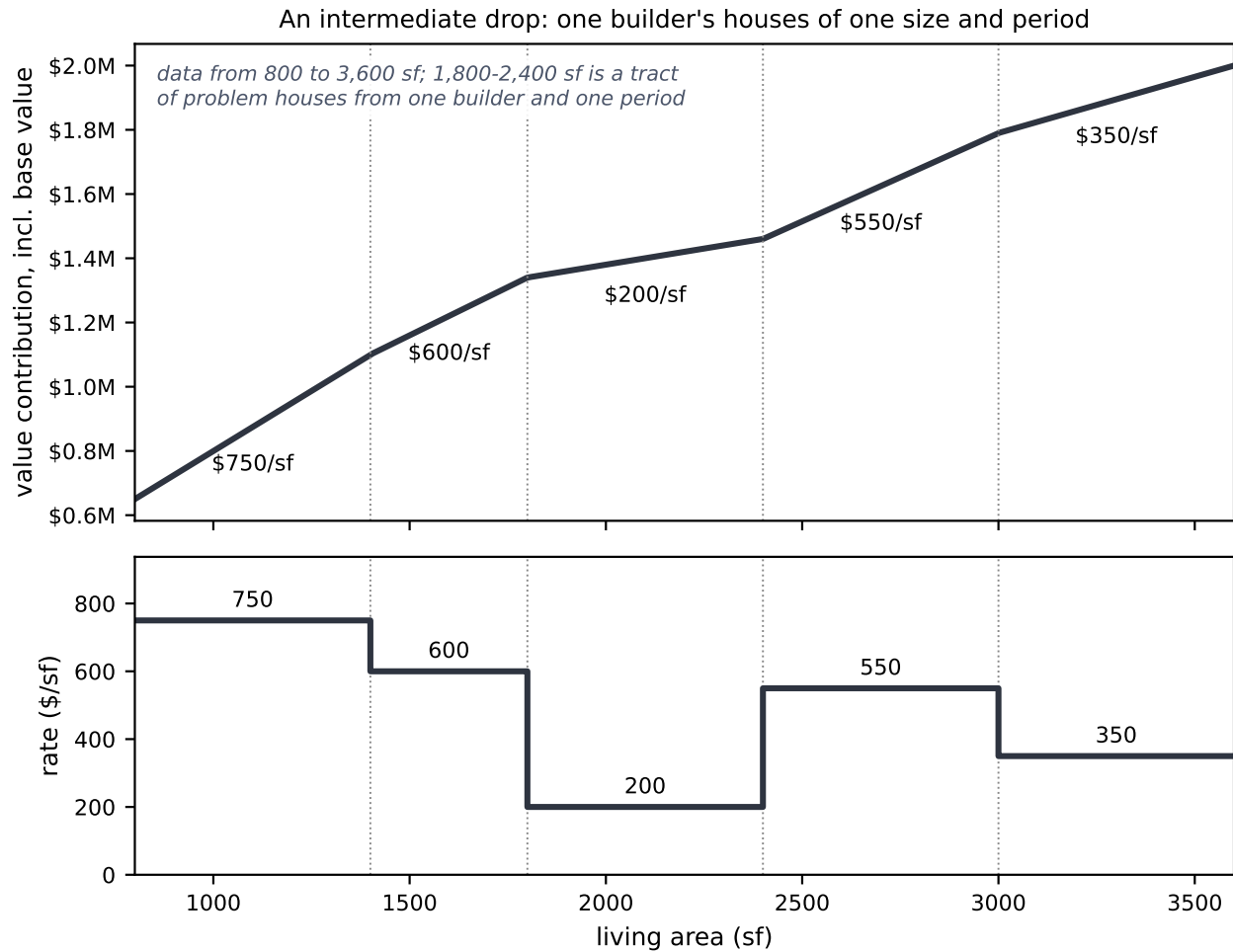


Figure 6: Ein Zwischenabfall, basierend auf Daten von 800 bis 3.600 Quadratfuß: $-750 h(1400 - \text{gla}) + 600 h(\text{gla} - 1400) - 400 h(\text{gla} - 1800) + 350 h(\text{gla} - 2400) - 200 h(\text{gla} - 3000)$, Schnittpunkt mit \$1,100,000 am ersten Knotenpunkt. Preise pro Quadratfuß: \$750, \$600, \$200, \$550 und \$350. Angenommen, das Segment von 1.800 bis 2.400 umfasst eine Reihe von Häusern dieser Größe, die von einem Bauunternehmer in einem bestimmten Zeitraum errichtet wurden und deren Marktpreise Probleme aufweisen (ein konstruiertes Beispiel); dort sinkt der Preis auf \$200 und erholt sich oberhalb dieser Marke wieder. Ein Leser, der den Koeffizienten -400 bei 1.800 als Kurs dieses Segments ansetzt oder davon ausgeht, dass die Kurse nur mit der Größe sinken, liegt bei der Bewertung dieses Gebiets falsch.

Segments ist nun eine Funktion der anderen Variablen und keine Konstante mehr. Es folgen die Aussage und ihr Beweis sowie die Auswirkungen auf die Berechnungen eines Gutachters.

Einstellung. Ein MARS-Modell mit Wechselwirkungen zwischen den Merkmalen $x, z_1, \dots, z_q$ hat den folgenden angepassten Wert:

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

wobei M die Anzahl der Terme außer dem Schnittpunkt ist, β_0 der Schnittpunkt ist und das Modell aus dem Schnittpunkt plus der Summe seiner Terme besteht, wobei der Term m gleich $\gamma_m \prod_{k \in F_m} B_{m,k}(z_k)$ ist: sein Koeffizient γ_m mal einem Faktor für jedes Merkmal in diesem Term. Hier gilt $z = (z_1, \dots, z_q)$, und damit ein Index über jedes Merkmal läuft, steht z_0 für x . Die in dem Term m vorkommenden Merkmale bilden die Menge $F_m \subseteq \{0, 1, \dots, q\}$; kein Merkmal kommt zweimal in einem Term vor, und jeder Faktor $B_{m,k}$ ist eine Funktion des einen Merkmals z_k , die zu einer von drei Arten gehört: das Merkmal selbst, z_k ; ein rechtes Scharnier $h(z_k - t)$; oder ein gespiegeltes Scharnier $h(t - z_k)$. Seien $M_x = \{m : 0 \in F_m\}$ die Terme, die x enthalten. Für $m \in M_x$ ist $B_{m,0}(x)$ der Faktor des Terms in x , und die anderen Merkmale des Terms, jene z_k mit $k \in F_m$, $k \neq 0$, sind die *Partner* von x in diesem Term. Schreiben Sie

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

für das Produkt der Partnerfaktoren, wobei ein leeres Produkt 1 ist. Mit anderen Worten: Sobald die anderen Merkmale des Bewertungsobjekts ausgefüllt sind, ergeben die Partner eines Terms im Ausmultiplizieren eine gewöhnliche Zahl; diese Zahl multipliziert mit dem gedruckten Koeffizienten ergibt den Koeffizienten eines einfachen Terms für die Wohnfläche, und der Term lautet wie in Fall 1. Das Partnerprodukt $P_m(z)$ ist das, was vom Term m übrig bleibt, wenn sein Faktor in x herausgezogen wird: Der Term m lautet $\gamma_m B_{m,0}(x) P_m(z)$. An einem bestimmten Punkt z ist es eine Zahl, sodass der Term auf einem Segment auf x wie ein einzelnes Scharnier mit dem Koeffizienten $\gamma_m P_m(z)$ wirkt, und dieses Produkt gibt an, wie die Partner in der Einheitspreis einfließen. Ein Term ohne Partner hat $P_m(z) = 1$ und ist ein Term des Falls 1.

Beispielsweise lautet im Ausdruck „ $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ “ des folgenden Beispiels, bei dem „ x “ die Wohnfläche ist, der Faktor in „ x “ „ $h(\text{gla} - 1200)$ “, der eine Partner ist das Grundstück und „ $P_m(z) = h(\text{lot} - 6000)$ “. Bei einem 5.000 Quadratfuß großen Grundstück $P_m(z) = 0$ trägt der Term nichts zum Preis bei; bei einem 10.000 Quadratfuß großen Grundstück $P_m(z) = 4,000$ wirkt der Term als Scharnier bei 1.200 mit dem Koeffizienten $0.03 \times 4,000 = 120$ und erhöht den Preis oberhalb von 1.200 Quadratfuß um \$120 pro Quadratfuß. Nichts hier schränkt ein, wie viele Merkmale ein Term enthalten darf; das Argument „**degree**“ von „earth“ legt dieses Maximum fest, und die folgenden Sätze gelten für jeden Grad. Ein Term mit d Merkmalen hat einen Graphen in $d + 1$ Dimensionen, den niemand zeichnen kann, sobald d den Wert zwei überschreitet; die Software zeigt nur Schnitte an, jeweils ein oder zwei Merkmale, während die anderen festgehalten werden, und Satz 3 beschreibt denselben Schnitt, der am jeweiligen Punkt des Bewertungsobjekts vorgenommen wird. $P_m(z)$ ist der Ort, an dem dieser Schnitt stattfindet: Egal wie viele Partner ein Term hat, an einem Punkt ist ihr Produkt eine Zahl. Im Term mit vier Merkmalen „ $\gamma h(\text{gla} - 1200) h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$ “, mit „ $\gamma = 0.001$ “, dessen Grafik fünf Dimensionen hat, gilt „ $P_m(z) = h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$ “; bei einem Grundstück von 10.000 Quadratfuß, einem Alter von 20 Jahren und 3 Bädern ergibt sich „ $4,000 \times 10 \times 3 = 120,000$ “, und entlang der Wohnfläche ist der gesamte Term ein Drehpunkt bei 1.200 mit dem Koeffizienten „ $\gamma \times 120,000 = 0.001 \times 120,000 =$

120“, der „\$120 pro Quadratfuß“ hinzufügt: genau wie der obige Term mit zwei Variablen, jedoch bei einem dreißigmal kleineren Koeffizienten. Deshalb erscheint der Koeffizient eines Terms mit hohem Grad oft als sehr kleine Zahl, wenn seine Variablen in großen Einheiten gemessen werden: Er multipliziert das Produkt mehrerer Variablenwerte. Und wenn sich ein Partnerwert auf der Nullseite seines Knotens befindet – also deutlich unter 6.000 Quadratfuß oder ein Alter unter 10 Jahren –, dann gilt „ $P_m(z) = 0$ “, und der Term trägt nichts bei, unabhängig davon, wie die anderen Partnerwerte lauten; je höher der Grad, desto mehr Möglichkeiten gibt es, einen Term bei einer bestimmten Eigenschaft auszuschalten. Die Beispiele in dieser Beilage verwenden wechselseitige Wechselwirkungen (**degree** = 2).

Seien $t_1 < \dots < t_K$ die verschiedenen Knoten jener $B_{m,0}$, die Scharniere sind, wobei $t_0 = -\infty$, $t_{K+1} = +\infty$ und die Segmente $S_i = (t_i, t_{i+1})$ wie in Fall 1 gelten. Für $m \in M_x$ und $i \in \{0, \dots, K\}$ definiere man

$$s_{m,i} = \begin{cases} +1 & \text{falls } B_{m,0} \text{ das Merkmal } x \text{ selbst oder ein rechtes Scharnier } h(x - t_j) \text{ mit } t_j \leq t_i \text{ ist,} \\ -1 & \text{falls } B_{m,0} \text{ ein gespiegeltes Scharnier } h(t_j - x) \text{ mit } t_j \geq t_{i+1} \text{ ist,} \\ 0 & \text{sonst.} \end{cases}$$

Ein *Punkt* ist ein Wert z der anderen Merkmale. Da jede $B_{m,0}$ auf S_i linear ist oder genau Null ist, ist f auf S_i für festgelegte z linear in x ; ihre Steigung dort ist die *Einheitspreis* auf S_i am Punkt z , geschrieben als $r_i(z) = \partial f / \partial x$ für $x \in S_i$.

Um alle Symbole auf einen Blick zu sehen, nehmen Sie das Modell des untenstehenden Beispiels so, wie es von „Earth“ ausgegeben wird: y-Achsen-Schnittpunkt 900.000, $600 h(\text{gla} - 1200)$, $-350 h(\text{gla} - 2400)$, $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ und 15lot . Die Merkmale sind Wohnfläche, $x = z_0 = \text{gla}$, und Grundstücksgröße, $z_1 = \text{lot}$, also $q = 1$; der y-Achsen-Schnittpunkt ist $\beta_0 = 900,000$ und es gibt $M = 4$ Terme. Bei einer Immobilie mit einer Wohnfläche von 2.000 Quadratfuß auf einem 10.000 Quadratfuß großen Grundstück:

m	γ_m	F_m	$B_{m,0}(x)$	$B_{m,1}(z_1)$	$P_m(z)$	Term beim Objekt
1	600	$\{0\}$	$h(\text{gla} - 1200)$	keiner	1	$600 \times 800 = 480,000$
2	-350	$\{0\}$	$h(\text{gla} - 2400)$	keiner	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	$\{1\}$	keiner	lot	(nicht in M_x)	$15 \times 10,000 = 150,000$

$M_x = \{1, 2, 3\}$: Der Term 4 enthält keinen Faktor in x , addiert unabhängig von der Wohnfläche immer denselben Betrag und geht nicht in der Einheitspreis von x ein. Der angepasste Wert lautet $f = 900,000 + 480,000 + 0 + 96,000 + 150,000 = \$1,626,000$. Die Terme 1 und 2 haben keinen Partner, daher gilt $P_1(z) = P_2(z) = 1$; Term 3 hat einen, und $P_3(z) = h(10,000 - 6,000) = 4,000$, den Multiplikator aus dem obigen Beispiel. Der sich daraus ergebende Satz 3 wandelt diese in der Einheitspreis um. Die Wohnfläche weist die Knoten $t_1 = 1,200$ und $t_2 = 2,400$ auf, und die Immobilie befindet sich auf dem Segment S_1 zwischen diesen beiden Knoten, wo die Terme 1 und 3 aktiv sind und Term 2 nicht ($s_{1,1} = s_{3,1} = 1$, $s_{2,1} = 0$); daher gilt $r_1(z) = 600 \times 1 + 0.03 \times 4,000 = 720$, was dem Preis für ein 10.000 Quadratfuß großes Grundstück in der Tabelle des Beispiels entspricht.

Satz 3 (der Einheitspreis auf einem Segment, unter Berücksichtigung von Wechselwirkungen). Mit dieser Notation gilt für jedes i und jeden Punkt z :

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

Beweis von Satz 3. Seien i , ein Punkt z und $x \in S_i$ festgelegt. Ein Term mit $m \notin M_x$ enthält keinen Faktor in x und ist konstant in x ; seine Ableitung in x ist 0. Ein Term mit $m \in M_x$ ist gleich $\gamma_m B_{m,0}(x) P_m(z)$, wobei $P_m(z)$ nicht von x abhängt. Nach der Fallanalyse im Beweis von Satz 1, angewandt auf den Einzelfaktor $B_{m,0}$, ist die Ableitung von $B_{m,0}$ bei x gleich 1 für das Merkmal selbst, 1 oder 0 für ein rechtes Scharnier, je nachdem, ob sich sein Knoten bei oder links von t_i oder bei oder rechts von t_{i+1} befindet, und -1 oder 0 für ein gespiegeltes Scharnier, je nachdem, ob sich sein Knoten bei oder rechts von t_{i+1} oder bei oder links von t_i befindet; das heißt, es entspricht $s_{m,i}$. Die Ableitung des Terms lautet also $\gamma_m s_{m,i} P_m(z)$, und die Ableitung von f , einer endlichen Summe, ist die Summe dieser Werte. Der Wert hängt nicht davon ab, welches $x \in S_i$ gewählt wurde. ■

Satz 1 ist der Sonderfall, in dem kein Term in M_x einen Partner hat, sodass für jeden $P_m(z) = 1$ gilt; gruppiert man die Terme der Summe nach ihren Knoten, erhält man die dort bewiesene Formel. Drei Korollaren beschreiben, wie sich die Partner ändern; die letzte davon ist das Scheitern der „Einer-pro-Knoten“-Schranke sowie der Fall, in dem ein Term verschwinden kann, wie das nachstehende Beispiel zeigt.

Folgerung 4 (der Einheitspreis ist nur an einem Punkt eine Zahl). Im Rahmen von Fall 2 ist für jedes i der Einheitspreis r_i eine Funktion des Punktes z . Sie ist in z konstant, wenn jeder Term $m \in M_x$ mit $s_{m,i} \neq 0$ keinen Partner hat. Wenn ein Term mit $s_{m,i} \neq 0$ einen Partner hat, ist $r_i(z)$ die Funktion $z \mapsto \sum_m \gamma_m s_{m,i} P_m(z)$, die sich mit den Werten der Partner ändert, außer wenn sich die Terme exakt aufheben; zwei Eigenschaften mit demselben x auf demselben Segment haben dann unterschiedliche Einheitspreise für x , wann immer sich ihre Partnerprodukte unterscheiden und sich die Unterschiede nicht aufheben. Aus diesem Grund nimmt das Programm einen Punkt als Eingabe, und aus diesem Grund existiert keine eigenständige Einheitspreistabelle für eine einzelne Eigenschaft.

Beweis. Nach Satz 3 gilt: $r_i(z) = \sum_{m \in M_x} \gamma_m s_{m,i} P_m(z)$. Wenn jeder Term mit $s_{m,i} \neq 0$ keinen Partner hat, ist jeder solche P_m das leere Produkt 1 und die Summe ist eine Konstante. Andernfalls enthält die Summe mindestens einen Summanden $\gamma_m s_{m,i} P_m(z)$ mit $\gamma_m s_{m,i} \neq 0$ und P_m , einem Produkt aus einem oder mehreren Faktoren, von denen jeder eine nicht-konstante Funktion eines Partners ist, sodass die Summe nur dann in z konstant ist, wenn sich diese Summanden identisch aufheben. ■

Korollar 5 (auf einem Segment von x ist der Einheitspreis in jedem Partner ein variabler Graph). Im Rahmen von Fall 2 werden i , ein Partner z_p ($p \geq 1$) sowie die Werte aller anderen Merkmale festgelegt. Dann ist $z_p \mapsto r_i(z)$ eine stetige, stückweise lineare Funktion von z_p , deren Scharniere zwischen den Knoten der Faktoren $B_{m,p}$ der Terme in M_x liegen. Entlang x , bei denen jeder Partner festgehalten wird, ändert sich der Einheitspreis nur an den Knoten $t_1, \dots, t_K$ von x ; im gemeinsamen Raum von x und seinen Partnern wird die Einheitspreisfläche sowohl von den Knoten von x als auch von den Knoten jedes Partners geschnitten.

Beweis. Wenn die anderen Eigenschaften festgelegt sind, ist jede Funktion $P_m(z)$ ein konstantes Vielfaches entweder von 1 (falls $p \notin F_m$) oder des einzelnen Faktors $B_{m,p}(z_p)$, der selbst entweder z_p , $h(z_p - t)$ oder $h(t - z_p)$ ist: eine stetige, stückweise lineare Funktion von z_p mit höchstens einer Bruchstelle an ihrem Knoten. Eine endliche Linearkombination solcher Funktionen – was Satz 3 für $r_i(z)$ angibt – ist stetig und stückweise linear, wobei Unterbrechungen nur dort auftreten, wo ein Summand unterbrochen ist. Die Aussage unter x ist Satz 3 selbst: Für ein festes z ist $r_i(z)$ auf jedem S_i eine Zahl, und die Segmente treffen sich nur an den Knoten von x . ■

Korollar 6 (die Schranke aus Satz 1 gilt nicht). Im Rahmen von Fall 2 ist die Anzahl der Terme mit einem Summanden ungleich Null in der Formel aus Satz 3 für $r_i(z)$ nur durch $|M_x|$

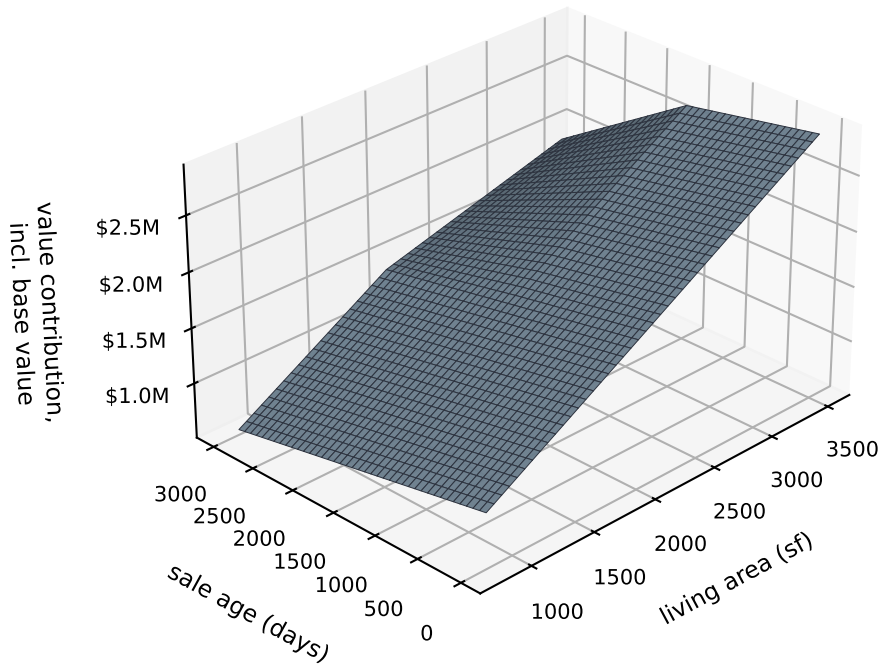
begrenzt, die Anzahl der Terme, die x enthalten, und kann K überschreiten: Ein Knoten von x kann in mehreren Termen von M_x mit unterschiedlichen Partnern auftreten, wobei jeder Term ein eigenständiger Term ist. Umgekehrt trägt ein Term mit „ $s_{m,i} \neq 0$ “ an einer Stelle, an der das Produkt seines Partners null ist, nichts bei.

Beweis. Die Terme $\gamma_1 h(x - 1500) \text{lot}$ und $\gamma_2 h(x - 1500) h(\text{age} - 10)$ sind verschiedene Terme von M_x mit demselben Knoten, und auf dem Segment oberhalb von 1,500 haben beide $s_{m,i} = 1$, sodass beide einen Beitrag leisten, und $r_i(z) = \gamma_1 \text{lot} + \gamma_2 h(\text{age} - 10)$ enthält zwei Summanden aus einem Knoten; das Hinzufügen weiterer Partner fügt weitere Terme am selben Knoten ohne Begrenzung hinzu. Für den umgekehrten Fall betrachte man einen Term, dessen Partner ein Scharnier ist: $h(\text{lot} - 6000)$ an einem Punkt mit $\text{lot} < 6000$: Dann gilt $P_m(z) = 0$ und der Summand verschwindet, obwohl $s_{m,i} \neq 0$; an einem Punkt mit $\text{lot} > 6000$ ist dies nicht der Fall. ■

An dieser Stelle sei eine praktische Beobachtung aus den eigenen Auswertungen des Herausgebers angeführt. Sobald Wechselwirkungen zugelassen werden, verläuft die von der Software für die Variable allein erstellte Wohnflächekurve oft geradlinig und zeigt einen einheitlichen Wert, während die Krümmungen, die ein Modell ohne Wechselwirkungen in die Wohnfläche eingebracht hätte, stattdessen in den Wechselwirkungstermen erscheinen, typischerweise bei der Wohnfläche in Abhängigkeit vom Verkaufsalter (der Zeit zwischen dem Verkauf und dem Stichtag) oder bei der Wohnfläche in Abhängigkeit von der Anzahl der Badezimmer. Satz 3 erklärt, warum die Gerade nicht der Einheitspreis darstellt. Die variablenbezogene Darstellung in earthUI zeichnet die Terme ohne Partner; die Darstellung ersten Grades von plotmo zeigt hingegen einen Ausschnitt des gesamten Modells, wobei die anderen Variablen auf ihren Medianwerten liegen. Die Interaktionsterme werden an anderer Stelle als Flächen dargestellt (die Darstellungen ersten und zweiten Grades von plotmo, Milborrow 2024b, die earthUI nebeneinander zeigt). Bei jeder konkreten Immobilie entspricht der Preis pro Wohnfläche der Steigung der Geraden zuzüglich – für jeden aktiven Interaktionsterm – dessen Koeffizienten multipliziert mit dem Wert seines Partnerfaktors; sie ändert sich an den Knoten der Wohnflächen dieser Terme, und um wie viel, hängt von den Partnern ab: Ein Verkauf vor sechs Jahren und ein Verkauf im letzten Monat oder ein Haus mit zwei Bädern und eines mit vier Bädern unterliegen unterschiedlichen Preistabellen für die Wohnfläche, auch wenn das Wohnflächendiagramm eine einzige Linie zeigt. Ein Leser, der die Linie für den Wert hält, ist erneut auf die Barriere gestoßen, nur eine Ebene höher; die Wertetabelle am Punkt des Bewertungsobjekts ist der Weg, um sie zu überwinden. Abbildung 7 stellt ein solches Modell so dar, wie es die Software tut, nämlich als Oberfläche über Wohnfläche und Verkaufsalter, und liest dann die grafische Darstellung der Wohnflächenvariable an zwei Verkaufsaltern ab.

Abbildung 8 verfolgt denselben Ansatz in umgekehrter Richtung, ausgehend vom Markt in Abbildung 2. Für diesen Markt wird ein Modell erstellt, bei dem die Wohnfläche mit dem Verkaufsalter interagiert, sodass der Haupteffekt der Wohnfläche eine gerade Linie darstellt, \$500 pro Quadratfuß, und die vier Interaktionsterme die Krümmungen bewirken. Die Kurve, die die Software für die Wohnfläche allein zeichnet, ist die Gerade. Bei einem vier Jahre alten Verkauf liegen die Preise für Wohnflächen bei Abbildung 2 pro Dollar, \$600, \$1,000, \$750 und \$100; bei einem neuen Verkauf ist es die Gerade; bei einem zwei Jahre alten Verkauf liegt der Wert dazwischen. Das Modell wurde konstruiert, um den Punkt zu verdeutlichen, und nicht an die Verkaufszahlen angepasst; eine echte Anpassung würde nicht so genau auf Abbildung 2 passen; aber der Mechanismus ist derselbe, den eine echte Anpassung verwendet, und deshalb hat ein Gutachter, der sieht, wie sich die Kurve der Wohnfläche nach Hinzufügen der Wechselwirkungen begradigt, die Kurven nicht verloren. Die Krümmungen haben sich in die Wechselwirkungsterme verlagert, wo sie nur durch einen Messwert an einem bestimmten Punkt wiederhergestellt werden können. Das Haupteffektdiagramm eines Modells mit Wechselwirkungen stellt nur die Terme der Variablen dar; dies ist der Ausschnitt bei

Living area by sale age: the surface the software draws



The living-area variable graph, read at two points

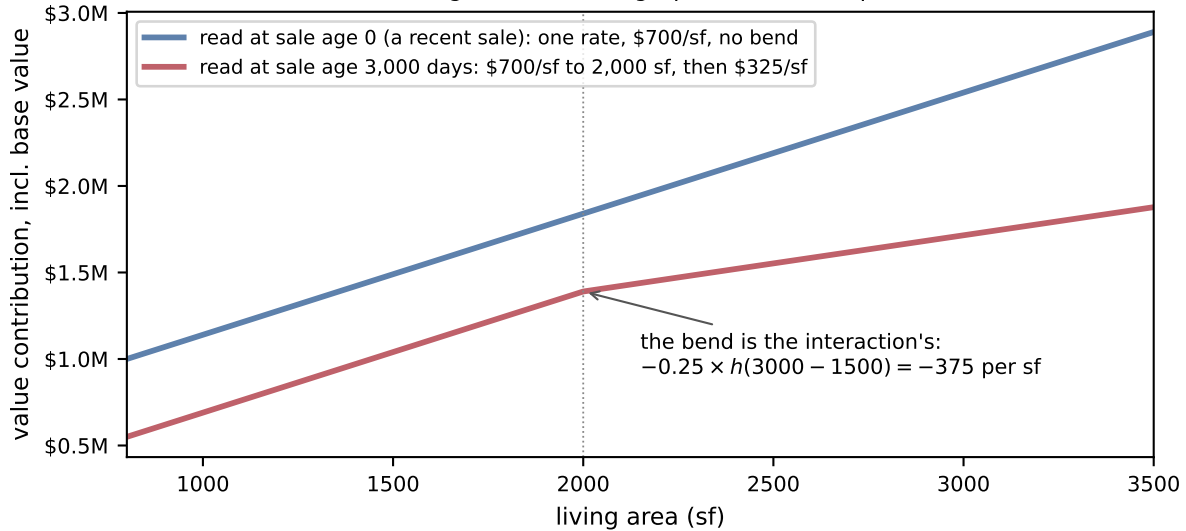


Figure 7: Eine Wechselwirkung zwischen Wohnfläche und Verkaufsalter: $1,000,000 + 700 h(\text{gla} - 800) - 150 \text{saleage} - 0.25 h(\text{gla} - 2000) h(\text{saleage} - 1500)$, basierend auf Daten von 800 bis 3.500 Quadratfuß und Verkäufen, die bis zu 3.000 Tage zurückliegen. Oben ist die Fläche dargestellt, die die Modellierungssoftware für dieses Paar zeichnet. Unten ist die Kurve der Variable „Wohnfläche“ an zwei Punkten dargestellt, wie es Satz 3 vorschreibt: Bei jedem Verkauf, der bis zu 1.500 Tage zurückliegt, beträgt der Partnerfaktor $h(\text{saleage} - 1500)$ null, die Wechselwirkung trägt nichts bei und die Kurve verläuft als gerade Linie bei \$700 pro Quadratfuß; bei einem Verkauf, der 3.000 Tage zurückliegt, beträgt der Partnerfaktor 1.500, die Wechselwirkung subtrahiert $0.25 \times 1,500 = 375$ pro Quadratfuß oberhalb von 2.000 Quadratfuß, und der Preis beträgt dort \$325. Der Haupteffekt der Wohnfläche (die Terme, die nur die Wohnfläche betreffen, ohne Partner) ist in beiden Fällen dieselbe Gerade; die Krümmung ist auf die Wechselwirkung zurückzuführen und tritt nur dort auf, wo der Partner aktiv ist.

einer Einstellung der Partner, bei der jeder Wechselwirkungsterm Null ist (hier ein neuer Verkauf), nicht die Form des Marktes; lesen Sie das Variablendiagramm am Bewertungsobjekt.

Die letzte Konsequenz betrifft eher die Anpassung als den Satz, und sie ist es, die sich auf das Verkaufsnetz auswirkt.

Satz 4 (Die Aufteilung einer Anpassung ist ordnungsabhängig). Im Rahmen von Fall 2 sei $y = z_p$ ($p \geq 1$) eines der anderen Merkmale und w die übrigen, z_k für $k \notin \{0, p\}$; $f(x, y, w)$ steht für f mit entsprechend gruppierten Argumenten. Sei ein Vergleichswert unter (x_c, y_c, w) und ein Bewertungsobjekt unter (x_s, y_s, w) , wobei für beide w gilt. Für einen Wert y^* von y definiere die *Anpassung für x , wobei y unter y^* beibehalten wird*, wie folgt:

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

Seien $M_{x,y}$ die Terme, die sowohl x als auch y enthalten. Jeder solche Term ist gleich seinem Koeffizienten multipliziert mit drei Teilen: $\gamma_m A_m(x) C_m(y) Q_m(w)$: $A_m(x) = B_{m,0}(x)$ ist sein Faktor in x , der entweder x selbst oder ein Scharnier in x ist; $C_m(y) = B_{m,p}(y)$ ist sein Faktor in y , der entweder y selbst oder ein Scharnier in y ist; und $Q_m(w)$ ist das Produkt seiner Faktoren in den verbleibenden Merkmalen, wobei jedes dieser Merkmale entweder das Merkmal selbst oder ein Scharnier darin ist, und ist gleich 1, wenn der Term keinen anderen Faktor hat. Im folgenden Beispiel ist der eine solche Term $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$, mit $A_m(x) = h(\text{gla} - 1200)$, $C_m(y) = h(\text{lot} - 6000)$ und $Q_m(w) = 1$. Schreiben Sie „ $\Delta_x(y_s)$ “ und „ $\Delta_x(y_c)$ “ für die Anpassung für „ x “, wobei „ y “ auf den Wert des Objekts „ y_s “ und auf den Wert des Vergleichsobjekts „ y_c “ festgelegt wird, d. h. „ $\Delta_x(y^*)$ “ bei „ $y^* = y_s$ “ und bei „ $y^* = y_c$ “; im Beispiel mit „ x “ als Wohnfläche und „ y “ als Grundstücksfläche, „ $\Delta_x(y_c) = \$240,000$ “ bei der 5.000-Quadratfuß-Grundstücksfläche des Vergleichsobjekts und „ $\Delta_x(y_s) = \$288,000$ “ bei den 10.000 des Objekts. Dann

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

während die Gesamtdifferenz $f(x_s, y_s, w) - f(x_c, y_c, w)$ keine Wahl von y^* beinhaltet. Die Aussage bezieht sich auf diesen Ausschnitt mit zwei Merkmalen, wobei ein Partner y bei beiden Eigenschaften auf demselben w gehalten wird; wenn sich drei oder mehr Merkmale zwischen den beiden unterscheiden, ist die Aufteilung der Gesamtsumme unter ihnen auf dieselbe Weise entlang einer längeren Kette von Ein-Merkmal-Verschiebungen pfadabhängig, und der Satz gilt für jeden Schritt der Kette, wobei die noch nicht verschobenen Merkmale auf ihren aktuellen Werten gehalten werden. Insbesondere bei einem einzigen solchen Term $\gamma A(x) C(y)$ und ohne Faktoren in w unterscheiden sich die beiden Anpassungen für x um $\gamma [A(x_s) - A(x_c)] [C(y_s) - C(y_c)]$, was nicht 0 ist, wann immer die Verschiebung in x A verändert und sich die beiden Eigenschaften in C unterscheiden. Für $A = h(x - t)$ und $C = y$ ist dies $\gamma [h(x_s - t) - h(x_c - t)] (y_s - y_c)$; für $A = h(x - t)$ und $C = h(y - s)$ – die Form des unten stehenden Beispiels – ist es $\gamma [h(x_s - t) - h(x_c - t)] [h(y_s - s) - h(y_c - s)]$.

Beweis von Satz 4. Gruppiere die Terme von f danach, welche der Terme x und y sie enthalten. Ein Term, der weder x noch y enthält, nimmt bei (x_s, y^*, w) und (x_c, y^*, w) denselben Wert an und hebt sich bei $\Delta_x(y^*)$ auf. Ein Term, der y , aber nicht x enthält, hebt sich ebenfalls auf. Ein Term, der x , aber nicht y enthält, trägt zu $\Delta_x(y^*)$ mit einem Betrag bei, der nicht von y^* abhängt, sodass er sich bei $\Delta_x(y_s) - \Delta_x(y_c)$ aufhebt. Ein Term $m \in M_{x,y}$ trägt $\gamma_m [A_m(x_s) - A_m(x_c)] C_m(y^*) Q_m(w)$ zu $\Delta_x(y^*)$ bei; setzt man $y^* = y_s$ und $y^* = y_c$ und subtrahiert, erhält man die angegebene Summe. Die Gesamtdifferenz ist eine Differenz zweier Werte der einen Funktion f und erwähnt kein y^* . ■

Der Wert der Wechselwirkung bezieht sich also auf das Merkmalspaar und nicht auf eines der beiden Merkmale. Ein Raster, das jeweils ein Merkmal an die Werte des Vergleichsobjekts anpasst, und

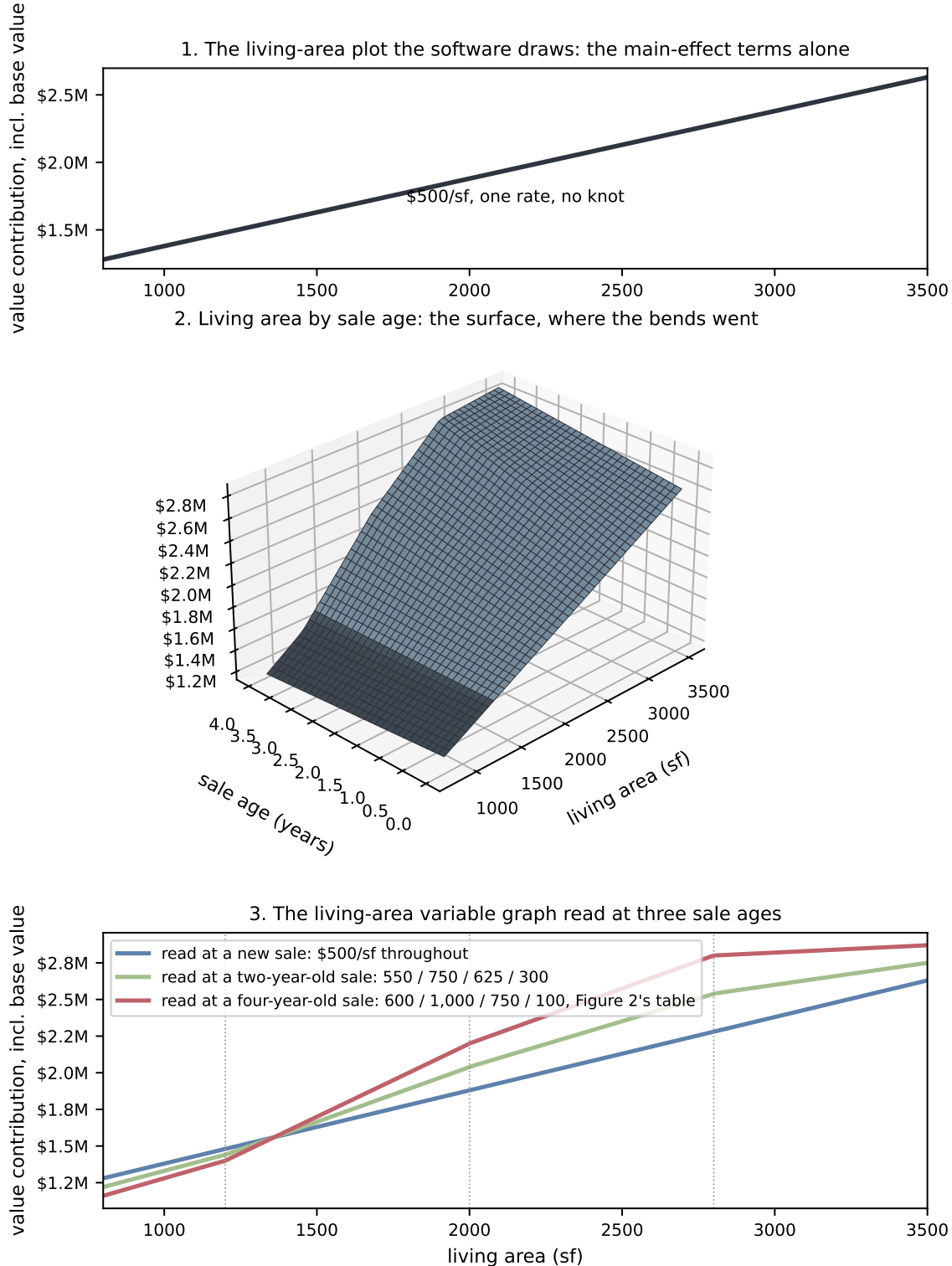


Figure 8: Der Markt aus Abbildung 2, mit einer Wechselwirkung zwischen Wohnfläche und Verkaufsalter: $1,280,000 + 500 h(\text{gla} - 800) - 20,000 \text{ saleyrs} + [-25 h(1200 - \text{gla}) + 125 h(\text{gla} - 1200) - 62.5 h(\text{gla} - 2000) - 162.5 h(\text{gla} - 2800)] \times \text{saleyrs}$, Verkaufsalter in Jahren von 0 bis 4. Die Diagramme sind in ihren Titeln beschrieben; bei vier Jahren weist das untere Diagramm genau die Einheitspreise aus Abbildung 2 auf, \$120,000, jedoch auf einem niedrigeren Niveau. Zur Veranschaulichung erstellt, keine Anpassungen vorgenommen.

eines, das es an die Werte des Bewertungsobjekts anpasst, stimmen hinsichtlich der Gesamtsumme überein, unterscheiden sich jedoch bei der Aufteilung. Die anderen Merkmale auf den Werten des Bewertungsobjekts beizubehalten, ist eine vertretbare Konvention und wird vom Programm verwendet; in einem Bericht muss angegeben werden, welche Konvention angewendet wurde. Ein Standardraster für Wohnimmobilien, wie beispielsweise das des „Uniform Residential Appraisal Report“ (URAR), enthält keine Zeile für eine Wechselwirkung. Wenn jeweils ein Merkmal nach dem anderen angepasst wird, müssen die beiden Schritte den Partner an entgegengesetzten Enden halten – die Wohnfläche wird an die Grundstücksgröße des Bewertungsobjekts angepasst, dann das Grundstück an die Wohnfläche des Vergleichsobjekts oder umgekehrt –, sodass die Wechselwirkung einmal gezählt wird und sich die Zeilen zur Gesamtdifferenz des Modells summieren. Bei direkt berechneten Beiträgen kann eine Tabelle die Wechselwirkung auch als eigene Zeile führen, nämlich als Differenz der Wechselwirkungsterme zwischen dem Bewertungsobjekt und dem Vergleichsobjekt, neben den Zeilen für die einzelnen Merkmale; jede Zeile stellt dann eine Gruppe von Termen dar, und es muss keine Reihenfolge gewählt werden. Im folgenden Beispiel sind die drei Zeilen „\$240,000“, „Grundstück\$75,000“ und „Wohnfläche nach Grundstück\$120,000“, die sich auf denselben Gesamtwert von \$435,000 summieren. Bei Wechselwirkungen ändert sich die Sekantenrate des Paares in zweierlei Hinsicht: zum einen durch die durchquerten Bereiche und zum anderen durch die Position des Paares innerhalb der Partner.

Ein durchgerechnetes Beispiel mit einer Wechselwirkung

Wieder ein Markt an der Küste, diesmal mit zwei Merkmalen – Wohnfläche und Grundstücksgröße – sowie einer Wechselwirkung zwischen diesen beiden. earth prints:

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

Wie in Satz 3 beschrieben. Die Wohnfläche enthält zwei Knoten, 1.200 und 2.400, sowie drei Terme in „ M_{gla} “: die beiden einfachen Scharniere und die Wechselwirkung, deren Partner „ $h(\text{lot} - 6000)$ “ ist. Auf dem Segment von 1.200 bis 2.400 ist das einfache Scharnier bei 1.200 aktiv, ebenso wie das Scharnier der Interaktion bei 1.200, sodass der Einheitspreis dort „ $600 + 0.03 h(\text{lot} - 6000)$ “ lautet: ein variabler Einheitspreis für das Grundstück, der bei Grundstücken bis zu 6.000 Quadratfuß konstant bei 600 liegt und sich für jeden weiteren Quadratfuß des Grundstücks um 3 Cent erhöht. Oberhalb von 2.400 greift der Schwellenwert bei 2.400, und der Satz beträgt $250 + 0.03 h(\text{lot} - 6000)$. Der grundstücksspezifische Satz beträgt $15 + 0.03 h(\text{gla} - 1200)$ bei Grundstücksgrößen über 6.000 Quadratfuß und 15 darunter, sodass der Satz jedes Merkmals von der Menge des anderen abhängt.

Zwei Punkte, zwei Tabellen. Der Einheitspreis-Tabelle des Programms für die Wohnfläche auf der Grundlage der Daten von 800 bis 3.500 Quadratfuß, einmal bei einem Grundstück von 5.000 Quadratfuß und einmal bei einem Grundstück von 10.000 Quadratfuß. Jede Spalte gibt die Steigung des Modells in der Wohnfläche an, wobei das Grundstück auf genau diesen einen Wert festgelegt ist und auf keinen anderen:

Wohnfläche (sf)	Preis bei einer Menge von 5.000 (\$pro Quadratfuß)	Preis bei einer Losgröße von 10.000 (\$pro
800 bis 1.200	0	
1.200 bis 2.400	600	
2.400 bis 3.500	250	

Bei der kleineren Losgröße ist der Partnerfaktor $h(5000 - 6000)$ gleich Null, sodass die Wechselwirkung keinen Beitrag leistet und die Tabelle derjenigen entspricht, die ein Modell ohne die Wechselwirkung liefern würde; das ist der Fall „der Term kann verschwinden“ aus Korollar 6. Bei der größeren Parzelle beträgt der Partnerfaktor 4,000, die Wechselwirkung fügt jedem Segment oberhalb von 1.200 den Wert $0.03 \times 4,000 = 120$ hinzu, und derselbe Knoten, 1.200, speist zwei aktive Terme, was das Scheitern der „Einer-pro-Knoten“-Grenze darstellt.

Nun zur Anpassung. Das Vergleichsobjekt hat eine Wohnfläche von 1.800 Quadratfuß auf einem 5.000 Quadratfuß großen Grundstück mit Anpassungen im Wert von \$1,335,000; das Bewertungsobjekt verfügt über 2.200 Quadratfuß auf einem 10.000 Quadratfuß großen Grundstück mit einem Schätzwert von \$1,770,000. Der zu erklärende Gesamtbetrag beträgt \$435,000. Passen Sie zunächst die Wohnfläche an das Grundstück des Vergleichsobjekts an, dann das Grundstück an die Wohnfläche der Immobilie. Jede Anpassung entspricht der Differenz der beiden verglichenen Werte; wenn die Anpassung innerhalb eines Segments bleibt, entspricht diese Differenz der Breite multipliziert mit dem Einheitspreis, wie es in der Tabelle angegeben ist:

	Zuerst die Wohnfläche, dann das Grundstück	Zuerst das Grundstück, dann die W
Anpassung der Wohnfläche	$400 \times 600 = 240,000$	$400 \times 720 =$
Grundstücksanpassung	195,000	
insgesamt	435,000	

Die Grundstücksanpassungen sind ebenfalls Unterschiede bei den Abgaben, und die Einheitspreise erklären diese; der grundstücksspezifische Einheitspreis wechselt bei 6.000 Quadratfuß: Bei den 2.200 Quadratfuß des Bewertungsobjekts liegt der Grundstückssatz bei \$15 unterhalb der Schwelle und bei \$45 oberhalb davon, sodass der Sprung von 5.000 auf 10.000 $1,000 \times 15 + 4,000 \times 45 = 195,000$ beträgt und nicht $5,000 \times 45$; bei den 1.800 Quadratfuß des Vergleichsobjekts beträgt er $1,000 \times 15 + 4,000 \times 33 = 147,000$. Die Summen stimmen überein, wie es Satz 4 vorschreibt, und die Aufteilungen weichen um \$48,000 voneinander ab, was genau der Formel des Satzes entspricht: $0.03 \times [h(2200 - 1200) - h(1800 - 1200)] \times [h(10000 - 6000) - h(5000 - 6000)] = 0.03 \times 400 \times 4,000$. Diese \$48,000 ist der Wert, den die zusätzlichen 400 Quadratfuß Wohnfläche *und* die zusätzlichen 4.000 Quadratfuß Grundstücksfläche zusammen ausmachen; er gehört zu keinem der beiden Merkmale, und wer auch immer als Zweites angepasst wird, erhält ihn. Gemäß der Konvention des Programms, bei der die anderen Merkmale auf den Werten des Bewertungsobjekts beibehalten werden, ergibt sich in der rechten Spalte für die Wohnfläche ein Wert von \$288,000 bei einem Preis von \$720 pro Quadratfuß, und der Bericht gibt an, an welcher Stelle der Preis abgelesen wurde:

```
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.

```

Beide Werte beziehen sich auf das 10.000 Quadratfuß große Grundstück des Bewertungsobjekts; auf dem eigenen 5.000 Quadratfuß großen Grundstück beträgt der Anteil der Wohnfläche des Vergleichsobjekts \$360.000.

Ein Prüfer, der dies auf dem Gelände des Vergleichsfahrzeugs erneut durchführt, erhält den Wert „\$240,000“ und liegt damit nicht falsch; der Prüfer und das Programm messen an zwei Stellen dieselbe Oberfläche, und im Bericht muss angegeben werden, um welche es sich handelt.

Kurz gesagt: Ohne Wechselwirkungen ist ein MARS-Modell eine Einheitspreistabelle pro Merkmal, und jede Aussage über einen Einheitspreis, der im Rahmen einer gewöhnlichen Regression getroffen wurde, lässt sich Segment für Segment übertragen. Mit Wechselwirkungen ist ein Einheitspreis eine Fläche; Satz 3 besagt genau, wie sie sich wölbt, und der Gutachter muss angeben, an welchem Punkt sie abgelesen wurde.

Das Programm

`mars_rates.pl`, das im Anhang vollständig abgedruckt ist, liest die gedruckte Tabelle von `earth` oder ein als Prolog-Term geschriebenes Modell ein und gibt die Einheitspreise-Tabelle, die Werte an den Knoten, die beiden Beiträge und deren Differenz aus, die Anpassung mit ihrer Erklärung für jeden Bereich, den Sekantenwert des Paares sowie eine Überprüfung, ob die Bereiche zusammen die Anpassung ergeben. Bei Angabe des Datenbereichs – dem kleinsten und größten Haus in den Verkaufszahlen – gibt es keinen Satz außerhalb dieses Bereichs an: Die Algebra eines Scharniers verläuft auf einer Seite bis ins Unendliche, die Verkaufszahlen jedoch nicht, und ein Satz jenseits dieser Werte wäre eine Extrapolation. Für das Beispiel mit Daten von 800 bis 3.500 Quadratfuß:

```

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.

In diesem Block stehen zwei Arten von Zahlen. Die Beiträge bestehen ausschließlich aus den Termen für die Wohnfläche, wobei der Achsenabschnitt und alle anderen Terme weggelassen wurden: \$600,000 und \$1,100,000 unterscheiden sich genau um den Anpassungswert. Die Werte an den Knotenpunkten sind Ergebnisse von Anpassungen, einschließlich des Achsenabschnitts, und in einem Modell mit Wechselwirkungen auch die Terme der Partner an diesem Punkt; hier ist \$1,520,000 bei 1.200 Quadratfuß der Achsenabschnitt, da dort jeder Wechselwirkungsterm Null ist. Die geschätzten Werte am Vergleichspunkt und am Untersuchungspunkt, \$2,120,000 und \$2,620,000, enthalten den Achsenabschnitt auf dieselbe Weise. Der Bericht kennzeichnet die Knotenwerte als geschätzte Werte, damit der Leser sie nicht als Beiträge versteht.

Dieser Abschnitt gehört in den Anhang eines Berichts: die Einheitspreise, die Knoten und die Berechnungen dazwischen – in einer Form, die ein Prüfer von Hand nachvollziehen kann. Die 69 Tests des Programms überprüfen die Vorzeichenregel, die Interaktionsregel, die Konsistenz der Segmente mit Anpassungen sowie den Parser für die earth-Ausgabe und bestehen unter Trealla, dem Prolog des Journals, sowie unter SWI-Prolog und Scyer Prolog. Für Leser, die mit R arbeiten, ist derselbe Inhalt verfügbar, ohne dass sie die Umgebung verlassen müssen: Die zusammen mit dem Programm veröffentlichte Datei „mars_rates.R“ nimmt ein „earth“-Objekt oder dessen ausgedruckte Tabelle entgegen und gibt denselben Block aus, wobei der Einheitspreis auf den Datenbereich begrenzt werden; der Prolog bleibt die geprüfte Referenz, und die R-Datei führt ihre eigenen Selbstprüfungen anhand der Beispiele dieser Beilage durch. Das Programm, seine Tests und das Skript für die Abbildung werden mit dieser Ausgabe als Download neben der PDF-Datei unter der GNU Affero General Public License, Version 3 oder höher, veröffentlicht.

Warum MARS, und was diese Ergänzung anderen Einträgen überlässt

Die Zeitschrift verwendet MARS aus drei Gründen, und genau diese Gründe ermöglichen die Existenz dieser Beilage. Objektivität: Die Knoten werden dort platziert, wo die Verkaufszahlen sie platzieren, und zwar nach einem festgelegten Verfahren mit einer festgelegten Ausdünnungsregel (dem späteren Schritt, der Terme entfernt, die ihren Platz nicht verdienen), und über diese festgelegten Einstellungen hinaus (maximaler Interaktionsgrad, Strafterm, Mindestabstand zwischen den Knoten) wird nichts über die Form eines variablen Graphen im Voraus festgelegt. Entdeckbarkeit: Der Vorwärtsdurchlauf durchsucht jedes Merkmal und jeden Kandidatenknoten – einschließlich Interaktionen bis zum festgelegten Grad – und berichtet, was er gefunden hat; dies kann keine Methode leisten, die von einer gewählten Form ausgeht. Interpretierbarkeit: Konstante Einheitspreise über von den Daten gewählte Bereiche und Interaktionen, die aus denselben Scharnierpunkten aufgebaut sind, ermöglichen es, das Modell als Einheitspreistabelle darzustellen und eine Anpassung als „Einheitspreis mal Breite“ zu erklären. Zwei weitere Methoden erzeugen ebenfalls Einheitspreise, die sich über den Bereich eines Merkmals hinweg ändern, und sie unterscheiden sich voneinander. Die erste, die segmentierte („Broken-Stick“-Regression), nimmt eine Variable und eine vom Analysten gewählte Anzahl von Bruchstellen als Eingabe und schätzt, wo diese Bruchstellen liegen, jeweils mit einem Standardfehler (ein Maß dafür, wie stark sich die Schätzung bei anderen Umsätzen verschieben könnte); zwischen den Bruchstellen ist der Einheitspreis konstant, wie bei MARS. Die zweite Methode, ein formgebundenes additives Modell, führt

statt gerader Segmente eine glatte Kurve ein, die auf eine vom Analysten vorgegebene Form gezwungen wird – zum Beispiel, dass der Wert mit zunehmender Wohnfläche niemals sinkt oder dass jeder zusätzliche Quadratfuß weniger beiträgt als der vorherige. Keine der beiden Methoden wird hier verwendet. Beide beantworten eine Frage, die in dieser Ergänzung nicht gestellt wird, und beide verzichten auf mindestens eine der drei Eigenschaften: Die segmentierte Regression verzichtet auf die Erkennbarkeit, da die Variable und die Anzahl der Brüche im Voraus festgelegt werden; das formgebundene Modell verzichtet auf Interpretierbarkeit, da eine Kurve keine konstanten Einheitspreise aufweist, die tabellarisch dargestellt werden könnten, sowie auf Objektivität, da seine Form vorgegeben ist. Diese Beilage befasst sich mit dem von MARS erstellten Modell; die Wahl der Methode sowie die Stabilität der Knotenpunkte und Einheitspreise, die MARS von einem Verkaufssatz zum nächsten ermittelt, sind Bewertungsobjekte für eigene Beiträge.

Was damit nicht entschieden ist

Die Einheitspreise stammen aus dem Modell, und die vier statistischen Aussagen gelten unverändert für sie: Eine Einheitspreis approximiert den bedingten Erwartungswert nur so gut, wie es die Form des Modells zulässt (*bedingter Erwartungswert*); sie kann den Wert einer ausgelassenen Variablen enthalten (*Verzerrung durch ausgelassene Variablen*); sie kann verzerrt sein, wenn sich ein Merkmal mit dem strukturellen Fehler bewegt (*Endogenität*); und sie lässt sich möglicherweise überhaupt nicht anhand der Verkaufszahlen bestimmen, egal wie viele es sind, wenn sich zwei Merkmale immer gemeinsam bewegen (*Parameteridentifikation*). MARS ermittelt eine Veränderung eines Einheitspreises nur dort, wo die Verkaufszahlen den entsprechenden Kontrast aufweisen; ein Knotenpunkt ist eine Aussage über den Markt und nur so aussagekräftig wie die Verkaufszahlen auf beiden Seiten davon. Der Beitrag dieser Ergänzung ist enger gefasst: Ausgehend vom Modell wird hier erläutert, wie man dessen Aussagen interpretiert.

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).

```