

人類理解 MARS 的障礙 (The Barrier to Human Understanding of MARS)

把 MARS 模型讀作單價、價值貢獻與調整

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

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

這是「基礎」單元的補充內容，而非概念條目。其標題指出了編輯歷經多年間接觀察後所察覺的一道障礙：那些能輕鬆解讀迴歸係數的評估師，在看到 MARS 程式列出的「鉸鏈項」清單時便會停下腳步，因為該清單中的內容看起來都不像「單價」。緊接本補充文章之後的四篇統計條目（「條件期望」、遺漏變數偏差、「內生性」及「參數識別」）皆探討了迴歸模型賦予某個特徵的「率」。本期刊自身的建模採用多變量自適應迴歸樣條 (MARS; Friedman 1991)¹，該方法並非針對每個特徵報告一個率，而是針對每個範圍報告一個率；而從 MARS 程式輸出結果到估價師能為其辯護的率表之間，存在一個關鍵步驟，若在此處手動編製表格，便會出錯。本補充資料闡述了該流程、關鍵步驟，並提供一個能依此流程運作且能自行驗證結果的程式。

形式定義。在 MARS 模型中，擬合值由截距加上各項項的總和組成，每項皆為係數乘以特徵的鉸鏈函數之積。擬合值即為模型對價格的估計值。對於單一特徵，在其他特徵保持不變的情況下，擬合值是一系列直線段。這些線段的交點即為節點。該特徵在其取值範圍內的價值貢獻圖即為變量圖。其在各區間內的斜率即為：即該區間內每單位特徵量所對應的擬合值變化量；某個特徵量在特定數值下的值貢獻，即該特徵在該數值處模型各項的總和；在範例中，所有這些項在 1,200 平方英尺處皆為零，因此貢獻值從該點開始計算，且在該點以下為負值；模型未必具有此類數值，且僅有貢獻值之間的差異（即調整值）才具有意義；而從可比物件到標的物件的調整值，是指為使可比物件的貢獻值等於標的物件的貢獻值而需加上的數額，該數額等於在兩者數值範圍內，率乘以寬度的總和。如此定義的單價即為該模型的單價；唯有當該模型被接受為對銷售情況的公正描述時，此單價才被視為市場支付的單價，而本補充說明之後的條目，則探討何時該接受是合理的。

為什麼需要這篇補充

一位符合 MARS 模型的估價師，無論是在 earthUI (Craytor 2026) 中，還是直接在 R 套件 earth (Milborrow 2024a) 中，都會收到一份類似以下的內容清單：

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

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.

沒有人會報告這一點。進入銷售表格的其實是調整值：標的物所貢獻的價值減去可比物件的價值。報告中會討論每平方英尺的單價，作為闡明 MARS 模型合理性的步驟之一。從書面條款到這些數字的推導過程雖簡短，但其中有一步容易導致手動編製的表格出現錯誤；若模型產出的數字無法向審閱者加以解釋，那麼使用模型的全部意義便不復存在。

簡明版：四個區間，四個單價

若在線性迴歸模型中將居住面積作為單一線性項，則會得出適用於整個市場的統一單價；這正是傳統估價中迴歸分析的做法，也是本補充說明中「普通迴歸」的涵義。MARS 則允許單價在銷售數據顯示其發生變動之處隨之調整。以範例中的沿海市場為例，居住面積在 1,200 平方英尺以內每平方英尺可能值 \$600，1,200 至 2,000 平方英尺之間每平方英尺 \$1,000，2,000 至 2,800 之間 \$750，超過該範圍後僅為 \$100。這就是全部的差異，而這正是房仲能立刻理解的一點：市場對能將小房子轉變為家庭住宅的空間出價最高，對下一間房間的出價較低，而對於超出該社區買家需求範圍的空間，則幾乎不願出價。

考量到這些單價後，其餘的計算都是估價師早已進行過的算術運算。一處面積為 1,800 平方英尺的可比物件與一處面積為 2,400 平方英尺的標的物之間相差 600 平方英尺，而 2,000 平方英尺這個關鍵點正位於兩者之間。加到可比物件上的調整值，是標的物之擬合值減去可比物件之擬合值，即 $\$2,620,000 - \$2,120,000 = \$500,000$ ，這同時也是標的物之居住面積貢獻值減去可比物件之居住面積貢獻值，即 $\$1,100,000 - \$600,000$ 。單價可說明此點：變化在 \$1,000 處跨越 200 平方英尺，接著在 \$750 處跨越 400 平方英尺，總計為 \$500,000。圖 3 顯示，根據模型對具備這些特徵的房屋所估算的「擬合值」（即模型賦予該房屋的價格），其範圍從 \$1,280,000（800 平方英尺，數據中面積最小的房屋）到約 \$3 million（3,500 平方英尺，面積最大的房屋），其中大部分漲幅發生在 2,800 平方英尺之前。

一條通過這兩個點的直線，在該次移動中唯一可能採用的單價，就是連接這兩點的單價，\$500,000 除以 600 平方英尺，約為 \$833 每平方英尺：此為該對點的割線單價，亦即該次移動的實際（或混合）單價：即按各穿越區間寬度加權後，該模型各單價的平均值。這並非模型中的單一費率，且對於下一組組合而言，其平均值將會不同。一般的迴歸分析，若僅針對居住面積採用單一費率，就必須以該單一數值套用至每組組合，因此無法重現這些逐對計算的平均值；而單一費率所做的正是：針對模型所預測的市場變動進行平均計算。

若讀者讀到這裡便已掌握方法：先計算兩項貢獻值的差額，接著透過查閱各區間的費率、以該費率計算與標的物相類似的項目，並將結果相加，以此解釋該差額。以下內容是為那些想了解「為何單價會是這樣」、「哪裡藏著一個陷阱」、「當各項特徵相互作用時會產生什麼變化」，以及「程式如何檢查算術運算」的讀者而寫；單價表和調整方式並不會改變。

細緻版：鉸鏈函數與容易漏掉的正負號

將變數在其定義域範圍內之值貢獻的圖形稱為「變數圖」。它並非一般日常意義上那種平滑彎曲的曲線；而是一條分段線性圖，在節點之間呈直線，僅在節點處發生彎曲。MARS 透過「鉸鏈」來建構此圖；Friedman (1991) 將其表述為截斷幂函數，而「鉸鏈」則是後世文獻所採用的名稱 (Milborrow, 未註明日期)；Hastie、Tibshirani 與 Friedman (2009, §9.4) 將位於同一個節點的兩條曲線稱為「反射對」。

定義（鉸鏈函數）。鉸鏈函數是一種單變數函數 x ，具有單一斷點，稱為其結點 t ；該函數在結點一側的值為零，而在另一側的值則等於距結點的距離。在任何結點處均有兩個此類點，每側各一個，如圖 1 所示：

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

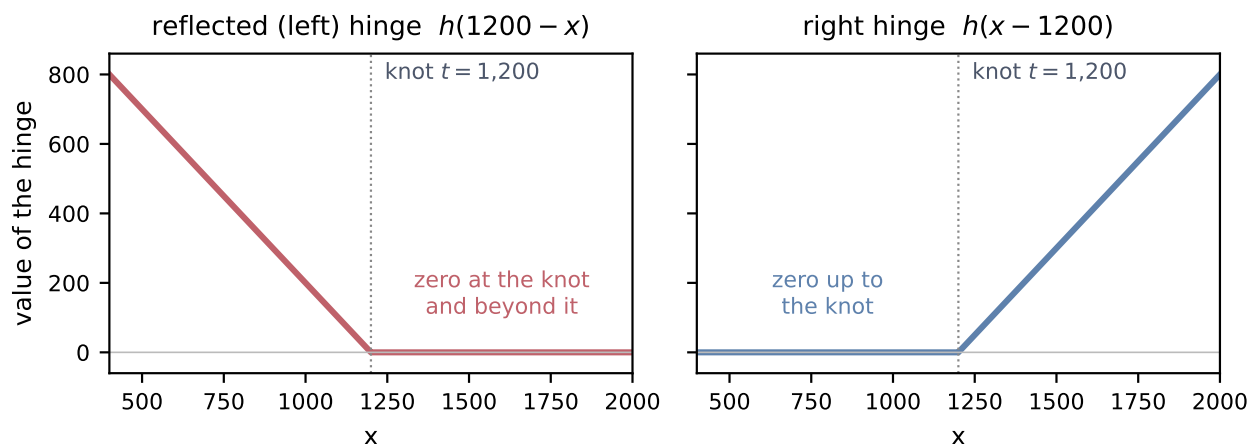


Figure 1: 這兩項鉸鏈函數的結點位於 $t = 1,200$ ：反射（左側）鉸鏈 $h(1200 - x)$ ，從左向右遞減，在結點處為零，並在結點之外仍為零；右側鉸鏈 $h(x - 1200)$ ，在結點之前為零，並在結點之後從左向右遞增。這些是基本鉸鏈函數；在模型中，每個鉸鏈函數都會乘以一個係數，該係數可能會逆轉其方向。

第一個， $h(x - t)$ ，是右鉸鏈：在結之前為零，越過結後則從左至右遞增。第二個， $h(t - x)$ ，是反射（或左）鉸鏈：在結之前從左至右遞減，並在結處及越過結後為零。兩者皆屬同一意義上的鉸鏈；它們的差異僅在於結的哪一側呈現非零值。這些方向描述的是圖1中從 x 軸繪製出的「純鉸鏈」，並無其他含義。在模型中，鉸點會乘以一個係數（該係數可能是負值），並與其他項相加；因此，基於右鉸點構建的項可能從左向右遞減，而基於反射鉸點構建的項則可能遞增，且變量圖會根據其他項所決定的高度與結相交。本補充說明中所述的所有方向均由左至右讀取。鉸點不會單獨使用。在 MARS 模型中，每個鉸點都會乘以一個係數，然後將各項相加；而模型的斜率（即單價）是根據這些係數計算得出，而非基於任何單一鉸點：以累積方式，範圍接範圍地計算，每個範圍的單價即為前一範圍單價，經由其間結點所對應的鉸鏈係數調整後所得的值。鉸鏈是 MARS 針對單一變數所擁有的唯一基本構成單元；變數圖中的每個彎曲處皆為某個鉸鏈的結點，而每個直線段則是該處為零或線性項的總和。以下所有內容需貫徹的核心要點是：鉸鏈的係數並非斜率。它是對斜率的調整，即在該鉸鏈的結點處，加至從左側傳入的斜率上的數值。由左至右閱讀時，變量圖從其第一段的單價開始，每個結點皆透過位於該處的係數來調整該單價。這就是整個解讀規則，而下文的定理 1 與 2 便是其證明。該模型由截距加上每個係數乘以其對應的鉸點組成，而任何區間內的變化率即是該和的斜率。

請參閱上方的列印條款；圖 2 將各項條款與其總和一併繪製，並在下方標示各區間的費率；而圖 3 則說明這些費率如何解釋可比對象與受測對象之間的調整。位於 1,200、2,000 及 2,800 的節點劃分出四個區間。在每個區間內，若其節點位於區間邊界或其左側，則為「右鉸鏈」；若其節點位於區間邊界或其右側，則為「反向鉸鏈」——右鉸鏈作用於其節點之後的所有區間，反向鉸鏈則作用於其節點之前的所有區間。該率是根據有效鉸點的係數計算得出：將有效右鉸點的係數相加，再減去有效反向鉸點的係數。所得的表格即為模型在各區間的單價，其餘特徵均維持在同一組數值（程式採用受試者的數值）；在沒有交互作用的模型中，此限制條件不會改變任何結果；而在下文的「案例 2」中，則會徹底改變一切：

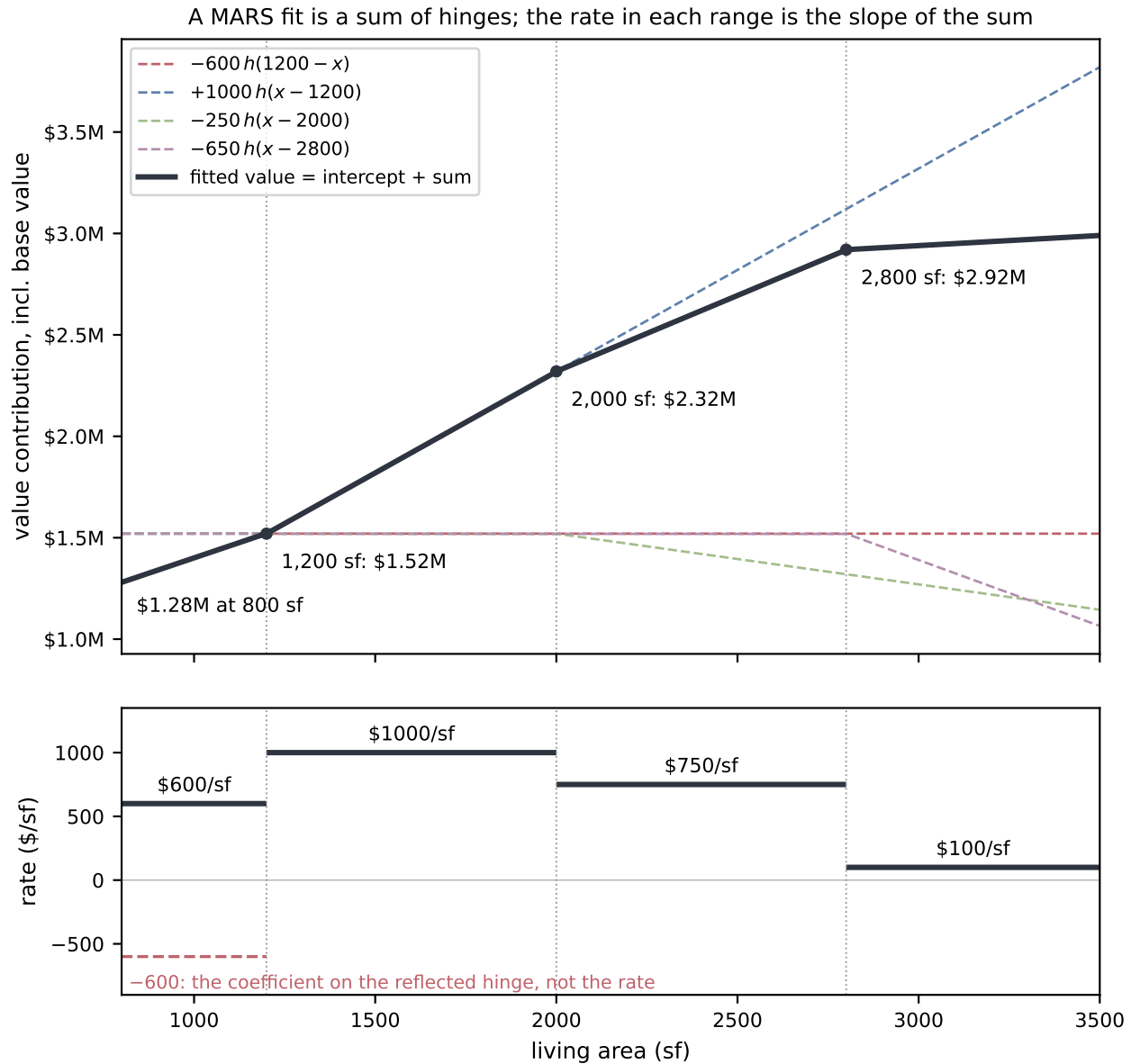
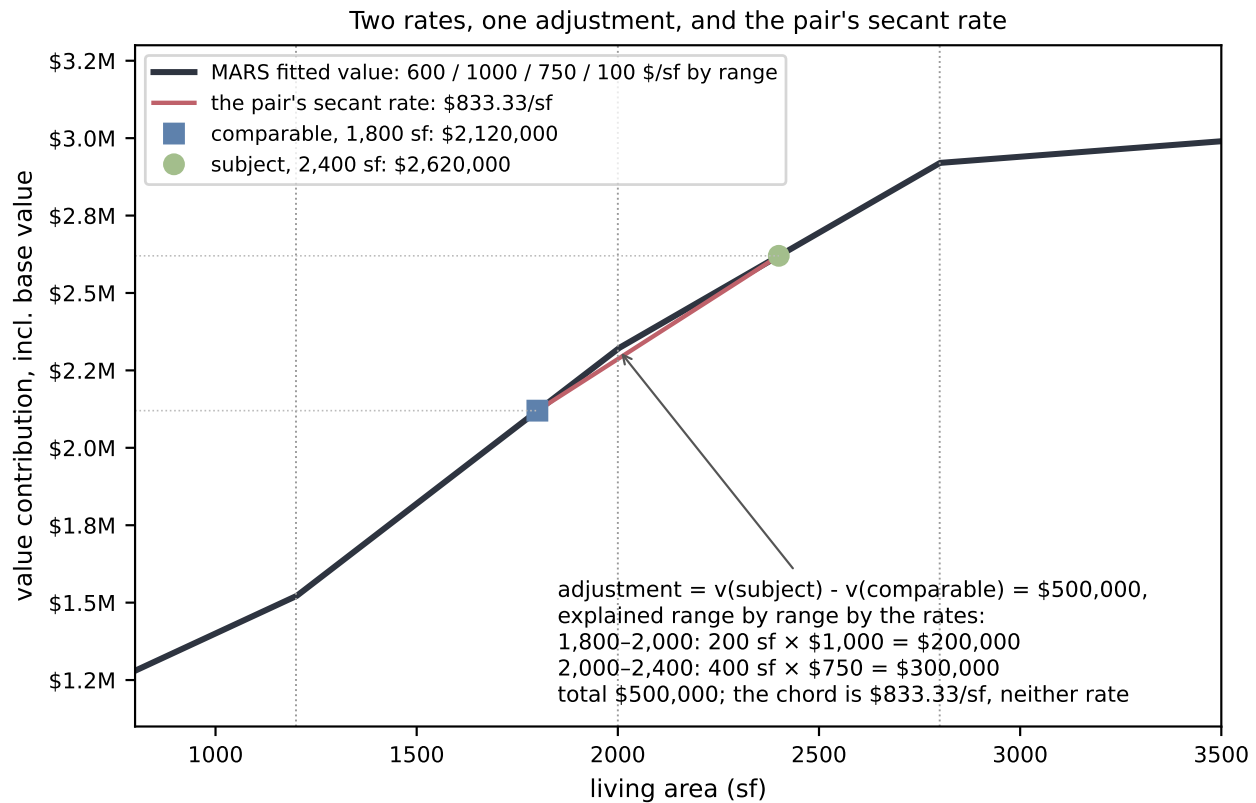


Figure 2: 範例中的四個鉸點條件、其總和，以及各區間的單價。反射鉸點 $h(1200 - gla)$ 上的係數 -600 已特別標示，以免讀者誤以為那是單價；1,200 平方英尺以下的單價為 $+600$ 。



```

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
Adjustment from the comparable (1,800 sf) to the subject (2,400 sf) =  $v(2,400) - v(1,800) = \$500,000$ ;
explained range by range by the rates:
1800-2000 sf: 200 sf  $\times$  $1000/sf = $200,000
2000-2400 sf: 400 sf  $\times$  $750/sf = $300,000
total = $500,000 secant rate for the pair:  $500,000 / 600 = \$833.33/\text{sf}$ 

```

Figure 3: 以 1,800 平方英尺的可比物件與 2,400 平方英尺的標的物進行擬合所得的數值。該調整值（即兩項貢獻值的差額）為 \$500,000，而現行費率可按各區間逐一解釋此數值；連接兩點的弦（約 \$833 每平方英尺）即為該對的割線費率，這是一條直線專屬於此對組合且不適用於其他組合的單一費率；一般迴歸分析則必須為每對組合使用一個費率。

範圍 (平方英尺)	活動鉸鏈	單價 (\$/sf)
低於 1,200	$h(1200 - \text{gla})$ 僅	$-(-600) = +\mathbf{600}$
1,200 至 2,000	$h(\text{gla} - 1200)$	1000
2,000 至 2,800	$h(\text{gla} - 1200), h(\text{gla} - 2000)$	$1000 - 250 = \mathbf{750}$
高於 2,800	所有三個右側鉸鏈	$1000 - 250 - 650 = \mathbf{100}$

第一行是常被忽略的步驟。反射鉸鏈上的係數 -600 並非每平方英尺的單價 $-\$600$ 。該鉸鏈 $h(1200 - \text{gla})$ 衡量房屋面積與 1,200 平方英尺之間的差距，而 -600 乘以該差距即為折扣額：對於 800 平方英尺的房屋，折扣額為 $\$240,000$ ，並在達到 1,200 平方英尺時歸零。每增加 1 平方英尺，便會減少 $\$600$ 的折扣額，因此折扣率為 $+\$600$ 。（若係數為正值，則同一鉸鏈因面積較小而增加價值，其率為負值。）規則概述如下：在某個範圍內，單價等於「有效右鉸鏈」係數之和減去「有效反向鉸鏈」係數之和，再加上線性項（若模型中存在此項，則以無結的直線特徵輸入）的係數。對此有一項慣例：在結點處，單價取自右側，因此右鉸鏈從其結點開始計數，而反向鉸鏈僅嚴格在其結點下方計數；因此，在 1,200 平方英尺處的單價為 $\$1,000$ ，即從該處開始的區間單價，程式亦以此方式讀取該值。

驗證表格是否正確建立的依據，是節點處的擬合值。在 1,200 平方英尺時，所有拐點係數皆為零，該數值即為截距， $\$1,520,000$ ；在 2,000 平方英尺時，該數值為 $\$1,520,000 + 1,000 \times 800 = \$2,320,000$ ；在 2,800 平方英尺時，為 $\$2,320,000 + 750 \times 800 = \$2,920,000$ ；而在 800 平方英尺時（即資料中面積最小的房屋），該數值為 $\$1,520,000 - 600 \times 400 = \$1,280,000$ 。從 800 到 1,200，數值上升 $\$240,000$ ，增幅為 $\$600$ ，這正是第一行；若將該表中的係數視為增長率，數值應從左至右遞減直至 1,200 這個節點，但實際數值顯示並非如此。圖2繪製了四個鉸點、其總和以及各區間內的增長率，並在讀者可能誤將其視為增長率的處標示了係數 -600 。

該程式還處理了另外兩個複雜情況，而手動表格必須：

- **交互作用。** earth 程式在前向演算階段（即逐項添加項的階段；Friedman 1991，§3），當某個特徵的率取決於另一個特徵時（即 交互作用），會輸出類似 $h(\text{gla} - 1500) \times \text{lot}$ 的結果。此時，「lot」的率值在居住面積高於或低於 1,500 平方英尺時會有所不同，且某個特徵的率值表總是根據其他特徵（其值維持為研究對象的數值）所給出的表格。程式會取一個點，並為每個其他特徵設定一個數值，進而計算該點處的所有率值。
- **水準與調整。** 截距與其他特徵中的項，共同設定了衡量貢獻值的基準；在每次調整中，這些項會相互抵銷。程式將貢獻值僅以該特徵中的項呈現，因此兩個貢獻值之間的差異恰好等於調整量。

兩種情況：不含與含交互作用的模型

上一節的規則可表述為兩個定理（證明如下），這些定理針對每個 MARS 模型讀者遲早都會提出的疑問，給出了明確的答案：一個評級實際上取決於多少個列印出的項？答案可分為兩種情況。

情況一：無交互作用

設定。設 $h(u) = \max(0, u)$ 為實數 u 。一個在特徵 $x, z_1, \dots, z_q$ 上不包含交互作用的 MARS 模型，其擬合值為

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

其中 β_0 為截距，每個 g_k 僅是 z_k 的函數，而 g 彙整了 x 中的所有項：

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

此處 $t_1 < t_2 < \dots < t_K$ 為 x 中鉸鏈的各異結點； a_j 為位於 t_j 的右鉸鏈係數， b_j 則為位於 t_j 的反射鉸鏈係數，若模型中無此類項，兩者皆為 0；而 c 為純線性項 $c \cdot x$ 的係數，此為以無結直線形式輸入的特徵，`earth` 僅針對 `linpreds` 中列出的特徵產生此項，或當 `Auto.linpreds` 啟用時（參見定理 1 後的備註）；若無此項（如本補充資料中所有模型的情況），則 $c = 0$ ；設 $t_0 = -\infty$ 且 $t_{K+1} = +\infty$ ，並將 $S_i = (t_i, t_{i+1})$ 定義為 $i = 0, 1, \dots, K$ ，即 x 的線段； t_i 是 S_i 的左結點， t_{i+1} 則是其右結點。在每個 S_i 上， g 的每個加項要麼恆等為零，要麼為線性，因此 g 在 S_i 上為線性；其在該處的斜率——由於 f 的其他加項均不依賴於 x ，故在 S_i 上亦為 $\partial f / \partial x$ ——即為 S_i 上的單價，並寫作 r_i 。 x 的單價表即為清單 $r_0, r_1, \dots, r_K$ 。（若 $K = 0$ 則僅有一段，即 $(-\infty, +\infty)$ ，且該表為單一單價 $r_0 = c$ 。）在結點本身處， g 沒有導數：其左導數與右導數分別為 r_{i-1} 和 r_i ，兩者相差 $a_i + b_i$ （見下文定理 2），因此單價表在 t_i 所示的值僅為慣例（此處取自右側），而非 g' 的值。

定理 1（區間上的單價）。採用本設定的符號，對於每個 $i \in \{0, 1, \dots, K\}$ ，

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

其中空和定義為 0。特別是， r_i 取決於至多 K 個 $2K$ 鉸接係數（每個結一個），以及 c 。

定理 1 的證明。固定 i 及一點 $x \in S_i$ ，使得 $t_i < x < t_{i+1}$ 。考慮 g 中的每個項，在 x 的某個鄰域內，該鄰域足夠小以致位於 S_i 之內。基項 $c \cdot x$ 的導數為 c 。對於位於 t_j 的右鉸點：若 $j \leq i$ ，則 $t_j \leq t_i < x$ ，因此該鄰域內 $h(x - t_j) = x - t_j$ ，其導數為 1；若 $j \geq i+1$ ，則 $x < t_{i+1} \leq t_j$ ，因此該處 $h(x - t_j) = 0$ ，其導數為 0。對於位於 t_j 的反射鉸點：若 $j \geq i+1$ ，則 $x < t_j$ ，因此 $h(t_j - x) = t_j - x$ ，其導數為 -1；若 $j \leq i$ ，則 $t_j < x$ ，因此 $h(t_j - x) = 0$ ，其導數為 0。在 x 處可微的函數之有限和，在該點處以導數之和的形式仍可微，這便得出了

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

所述的公式。關於計數：索引集 $\{j : j \leq i\}$ 和 $\{j : j \geq i+1\}$ 互不相交，且共同涵蓋 $\{1, \dots, K\}$ ，因此每個結 t_j 僅貢獻 a_j 或 b_j 其中之一，絕不會同時貢獻兩者；故至多有 K 個鉸點係數，且 r_i 在 S_i 的每個點上均相同。■

已達到該上限：若對於每個 j 皆滿足 $b_j = 0$ ，則最右側的區段 S_K 會用盡 a_j 的所有 K 。

備註（當 c 不為零時）。`earth` 可以透過兩種不同的方式，將一個無結點的單一直線特徵（即設定中的 $c \cdot x$ 項）表示出來，而這兩種方式的差異至關重要。`linpreds`，這是一個包含命名特徵（例如 `linpreds = c("GLA")`）的清單，會強制清單中的每個特徵以線性方式輸入，即使使用帶有結點的擬合效果會更好；這取決於分析者的決定，但可能會影響擬合效果。`Auto.linpreds`，這是一個預設為啟用的單一開關，不會強制任何行為：它僅改變 `earth` 如何呈現實際上為真正線性（即僅有一個鉤點，且鉤點位於最小觀測值處）的特徵——將其以係數和變數名稱的形式輸出，而非以鉤點形式呈現；擬合結果保持不變。無論哪種情況，`earth` 都會將該項以純變數名稱及其係數的形式輸出，在 `earth` 的摘要形式中呈現為 `221.4 * GLA`，如下例所示。定理 1 透過 c 來處理此情況，該指令會在每個區段上增加單價。例如，呼叫

```
mod_linear <- earth(
  SalePrice ~ GLA + LotSize + BathrmCount + BedrmCount,
```

```

data = sales,
degree = 1,
linpreds = c("GLA", "LotSize", "BathrmCount", "BedrmCount")
)

```

會以線性方式處理每個特徵（透過 `degree = 1`，因此線性項的乘積也不會被納入擬合），並以 `earth` 的摘要形式輸出（此為示意性輸出，非擬合模型）：

```

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

```

對於居住面積，此設定採用 $c = 221.4$ 且無鉸點， $K = 0$ ：僅有一段，且根據定理 1，可得單一係數 $r_0 = c = \$221.40$ 每平方英尺，適用於整個數據範圍；地塊面積、浴室和臥室的計算方式相同，分別為 $\$11.80$ 每平方英尺地塊面積、 $\$34,872$ 每間浴室，以及 $\$17,641$ 每間臥室。此類模型在本補充說明的意義上屬於普通迴歸，即每個特徵對應一個係數。在數據上，此類項與單一右鉸點的直線相同，其鉸點位於最小的觀測值處；而《Journal》採用 `Auto.linpreds = FALSE` 進行分析，因此在其模型中，若某個特徵的變量圖在數據上呈直線，則會以該單一鉸點呈現，例如 `h(gla-800)` 700，其鉸點位於數據中最小的房屋處，並包含 $c = 0$ 、 $K = 1$ 及 $a_1 = 700$ ：在結點下方，單價為 0，此區段無任何銷售紀錄；而在結點上方，單價為 700，因此資料上的單價表僅有一行。由此可得出一個推論；結點處的單價變化即為下文的定理 2。

推論 1（反射符號）。係數為 b_j 的反射鉸鏈，會對 t_j 左側的每個區段的單價貢獻 $-b_j$ ，對右側的每個區段貢獻 0。在範例中，位於 $t_1 = 1,200$ 處的 $b_1 = -600$ ，會對 r_0 貢獻 $+600$ ，而這正是本補充資料所圍繞的整個陷阱。

證明。參照定理 1 的第二個和式。 ■

七個節點的例子，以及讓讀者止步的假設

這個計數之所以重要，是因為幾乎所有人對列印出的 MARS 模型都抱持著一個假設：印在結點旁邊的係數，就是該結點所屬區段的斜率。但事實並非如此。鉸鏈節點的係數是用來調整從左側傳入的斜率，而非斜率本身；因此，線段的斜率等於首個線段的斜率，加上其左側節點以及更左側每個節點處所有此類調整值的累計總和。抱持此錯誤假設的讀者，若在腦中計算，或透過小型程式或試算表公式進行運算，會發現計算結果無法重現擬合值，進而放棄並斷定 MARS 難以理解。它並非難以理解；而是具有累積性。

定理 2（累計總和）。採用案例 1 的符號標記，對於每個 $i \in \{1, \dots, K\}$ ，

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

因此，對於每個 i ， $r_i = r_0 + \sum_{j=1}^i (a_j + b_j)$ ，其中 $r_0 = c - \sum_{j=1}^K b_j$ 為最左側區間的單價， $r_K = c + \sum_{j=1}^K a_j$ 為最右側區間的單價。

定理 2 的證明。固定 $i \in \{1, \dots, K\}$ 。根據定理 1，

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

兩式相減後， c 相互抵消； a 的兩個和僅在單項 a_i 上有所不同；且由於 $\sum_{j=i}^K b_j = b_i + \sum_{j=i+1}^K b_j$ ， b 的兩個和僅在單項 b_i 上有所不同，該項在 r_{i-1} 中被減去，但在 r_i 中未被減去。因此 $r_i - r_{i-1} = a_i + b_i$ 。將此恆等式 $j = 1, \dots, i$ 代入 i 後，左邊可進行縮合： $(r_1 - r_0) + (r_2 - r_1) + \dots + (r_i - r_{i-1}) = r_i - r_0$ ，由此得出求和公式。 r_0 與 r_K 的數值分別見於 $i = 0$ （其中 a 的和為空）及 $i = K$ （其中 b 的和為空）中的定理 1。■

推論 2（讀取規則）。在案例 1 的設定下，對於每個 $i \in \{1, \dots, K\}$ ，其左端結為 t_i 的區段的單價為 $r_i = r_{i-1} + a_i + b_i$ ：印在結上的係數會加到從左側傳來的單價上，而這些係數本身並非該區段的單價。這兩者相等，即 $r_i = a_i + b_i$ ，恰當且僅當 $r_{i-1} = 0$ ；特別是，對於僅含右鉸鏈且無線性項的模型，此情況發生於 $i = 1$ ，且根據定理 2，除非其前面的鉸鏈係數之和為零，否則不會發生於更後面的鉸鏈。

證明。第一個等式即是定理 2。 $r_i = a_i + b_i$ 當且僅當可由此推得 $r_{i-1} = 0$ 。若對所有 j 及 $c = 0$ 皆成立 $b_j = 0$ ，則根據定理 2 可得 $r_0 = 0$ ，因此 $r_1 = a_1$ ，且 $r_{i-1} = \sum_{j<i} a_j$ ；而僅當這些係數之和為零時，才等於 0。■

推論 3（調整作為積分）。在案例 1 的設定下， g 在每個結點處皆為連續函數，因為每個鉸鏈皆為連續函數；同樣地，對於每個 z ， $f(\cdot, z)$ 亦為連續函數；這正是為何無論從左側或右側抵達結點，其值皆相同，並使得單價能夠從一個量逐步過渡到另一個量。設 x_c 及 x_s 分別為某可比對象與某主體在 x 下的數量，且兩者的其他特徵均維持相同數值，因此 $g(x)$ ——即 f 中涉及 x 的項——即為 x 在數量 x 下的貢獻。設 $u_1 = \min(x_c, x_s)$ 、 $u_2 = \max(x_c, x_s)$ 及 $|I|$ 分別表示區間 I 的長度。接著，調整值 $g(x_s) - g(x_c)$ ——即為使可比對象的貢獻等於研究對象的貢獻而需加上的數值——為

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

。換言之：調整值即是兩量之間單價的積分；由於單價在每個區間內恆定，因此該積分等於沿途各區間中「單價乘以區間寬度」之和。

證明。函數 g 是連續的，且為具有有限個區段的分段線性函數，因此是絕對連續的；根據微積分基本定理，可得 $g(x_s) - g(x_c) = \int_{x_c}^{x_s} g'(u) du$ ，其中導數在所有點處均存在，唯獨在結點處不存在，而結點集為有限集。在每個 S_i 上，根據定理 1，被積函數等於 r_i ，因此 $\int_{u_1}^{u_2} g'(u) du$ 是 i 上 r_i 乘以 $S_i \cap [u_1, u_2]$ 長度的總和，未相交的線段貢獻長度 0；且根據 $\int_{x_c}^{x_s} = \text{sign}(x_s - x_c) \int_{u_1}^{u_2}$ 積分的取向，。■

率表為 g' ；該調整值直接計算為 $g(x_s) - g(x_c)$ ，等於該表在兩個數值之間的積分；而定理 1 與 2 闡明了 g' 的組成。實際上，調整值總是計算為各項貢獻的差值，這完全不需要使用率數；該推論闡明了為何按範圍逐一讀取的率數能重現該結果，這正是使率數成為對調整值的解釋與驗證，而非計算方法的原因。

以某資料集中的居住面積為例，其中最小的房屋為 800 平方英尺，共有七個節點，分別位於 1,200、1,600、2,000、2,400、2,800、3,200 及 3,600 平方英尺，且模型中有九個項，其中兩個項代表鉸鏈，各與同一節點上的右鉸鏈配對：分別位於 1,200 和 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

$h(\text{gla}-3200)$	-150
$h(\text{gla}-3600)$	50

七個結點構成八個區段。每個區段的單價等於有效右鉸鏈係數之和，減去有效反射鉸鏈係數之和：

區段 (sf)	有效條款	單價 (\$/平方英尺)
800 至 1,200	$h(1200 - \text{gla}), h(2400 - \text{gla})$	$500 + 200 = 700$
1,200 至 1,600	$h(\text{gla} - 1200), h(2400 - \text{gla})$	$700 + 200 = 900$
1,600 至 2,000	右側鉸鏈位於 1,200、1,600； $h(2400 - \text{gla})$	$700 - 100 + 200 = 800$
2,000 至 2,400	右鉸鏈的價格介於 1,200 至 2,000 之間； $h(2400 - \text{gla})$	$700 - 100 + 150 + 200 = 950$
2,400 至 2,800	右鉸鏈在 1,200 至 2,400 之間	$700 - 100 + 150 - 250 = 500$
2,800 至 3,200	右鉸鏈在 1,200 至 2,800 之間	$500 - 300 = 200$
3,200 至 3,600	右鉸鏈在 1,200 至 3,200 之間	$200 - 150 = 50$
超過 3,600	所有七個右側鉸鏈	$50 + 50 = 100$

有三點需要從表格中讀取。最右側區段的單價 $700 - 100 + 150 - 250 - 300 - 150 + 50 = 100$ ，結合了所有七個右鉸鏈的係數（每個結對應一個），這正是定理 1 中 $K = 7$ 的上界；其中兩個反射鉸鏈在此不發揮作用，因為反射鉸鏈在其自身結的右側是平坦的，而最右側區段位於每個結的右側。最左側的區段，從 800 起至其右側結點 1,200 為止，則是鏡像：僅計入反射鉸鏈，且其符號反轉，因為每個結點皆位於其右側結點處或其右側。而在每個攜帶一對的結點處，步長即為兩個係數之和：在 1,200 處， $-500 + 700 = +200$ ，範圍從 700 到 900；在 2,400 處，單價從 950 降至 500，變化量為 -450 ，這恰好等於該結處兩個係數的和，即 -250 與 -200 ：穿越該結會開啟右鉸鏈並關閉反射鉸鏈，而這兩項變化皆朝同一方向作用。

最右側區段的相同步速也可透過「步行」方式達成：從第一個區段的 700 開始，經由 800 至 1,200，並在經過每個節點時，加上該處係數的總和： $+200$ 、 -100 、 $+150$ 、 -450 、 -300 、 -150 、 $+50$ 。該路徑會經過所有八個區段和九個項，最終同樣抵達 100。直和（direct sum）的長度較短，也是程式所執行的操作；而這條路徑則是讀者應親手走一遍，以確認結點旁的係數代表的是「步數」而非「層級」。位於 2,400 處的 -450 包含兩個半步：穿越該結點後，會將反射鉸鏈 $h(2400 - \text{gla})$ 留在後方，連同它在結點下方所添加的 $+\$200$ ；同時引入右鉸鏈 $h(\text{gla} - 2400)$ ，其 -250 即由此處開始。 $-200 - 250 = -450$ ，且速率從 \$950 降至 \$500。

圖 4 針對五節點模型得出了相同的結論，該模型規模較小，足以憑肉眼驗證。橫桿代表印在節點上的係數；階梯圖則表示各區段的單價；階梯圖始於第一個區段的單價 \$900（由反射鉸鏈設定），而每個階梯皆為該節點處所有橫桿數值的總和。最右側段落的單價 $650 - 250 + 150 - 200 - 150 = 200$ 運用了所有五個右側鉸鏈；位於 1,200 處的反向鉸鏈（標示為 -900 ）將第一段設定為 $+900$ ，且不與其他任何部分產生接觸。

關於該圖形的形狀，有兩點值得一提，因為它們涉及的是房屋，而非 MARS。首先，坐標軸的起點是 800 平方英尺而非零，且圖表從坐標軸開始繪製：本附錄中所有擬合模型的圖表，如同建模軟體繪製的圖形一樣，皆從資料中的最小住宅延伸至最大住宅，並不會在該範圍之外繪製任何內容。每個市場區域都有最小的房屋，低於該面積便無銷售紀錄，而變量圖表在該值以下也未顯示任何內容：並非零，而是銷售數據無法支撐的數值；公式在該處仍會返回一個數值，而該數值屬於外推結果。圖表採用 800 平方英尺作為下限，但該下限是市場區域的特徵，而非恆定值：不同市場區域在 MLS 中顯示的房產最小面積各不相同，而該數值的具體位置由資料集決定，而非估價師。其次，曲線形狀因市場而異：在沿海地區的例子中，斜率最陡的區間是 1,200 至 2,000 平方英尺，而在圖 4 中，則是第一個區段。在阿瑟頓（Atherton）或希爾斯伯勒（Hillsborough）等高價區域，居住面積（GLA）曲線的陡峭區間可能位於 4,000 或 5,000 平方英尺之後，那正是買家集中的區間。此外，若某個社區曾於特定時期由建商興建單一規格的住宅且存在問題，該社區的中等

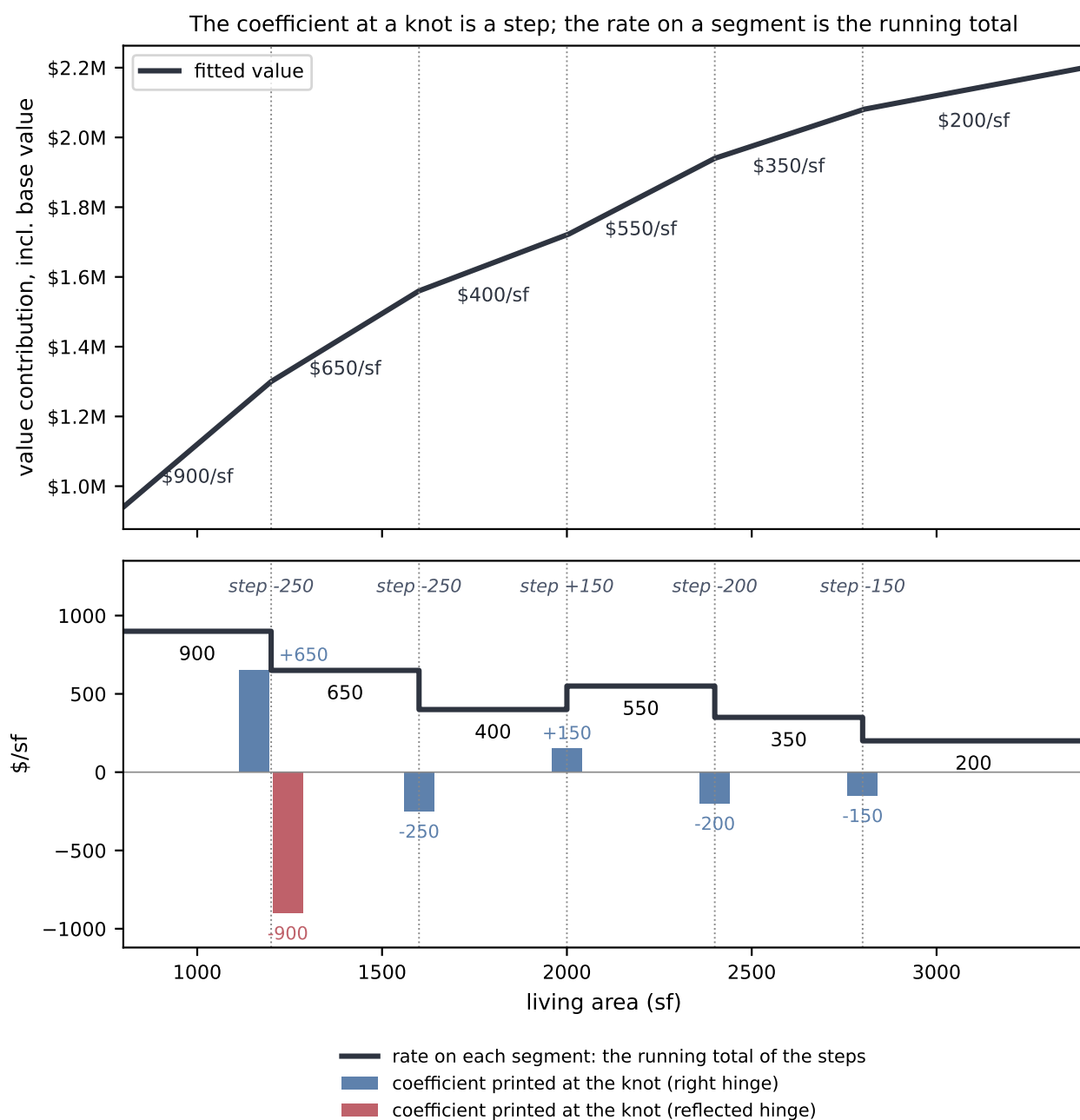


Figure 4: 基於起始面積為 800 平方英尺的數據所建立的五節點模型： $-900h(1200 - gla) + 650h(gla - 1200) - 250h(gla - 1600) + 150h(gla - 2000) - 200h(gla - 2400) - 150h(gla - 2800)$ 。上圖為擬合值，各區段均標註了單價；由於不存在面積低於 800 平方英尺的房屋，因此該區間無數據。下方圖中，各節點處的係數以條形圖呈現，各區間的單價則以階梯圖顯示：階梯的每一步皆為該節點處條形圖的總和（此處為一條，若節點包含一對數據則為兩條，定理 2），因此一段的單價等於第一段的單價加上其左側階梯的累計總和，而非其旁邊的係數：每個係數皆用於調整從左側傳入的單價。

面積區間（如圖中 1,600 至 2,000 平方英尺區間）的估值率可能會出現急遽下降。在更廣闊的市場區域內，變量曲線形狀各異，而費率表正是估價師解讀特定曲線所呈現資訊的方式。圖 5 和 6 繪製了其中兩種曲線形狀（如實印製），以便讀者能依照規則建立各費率表，並與階梯曲線進行比對。

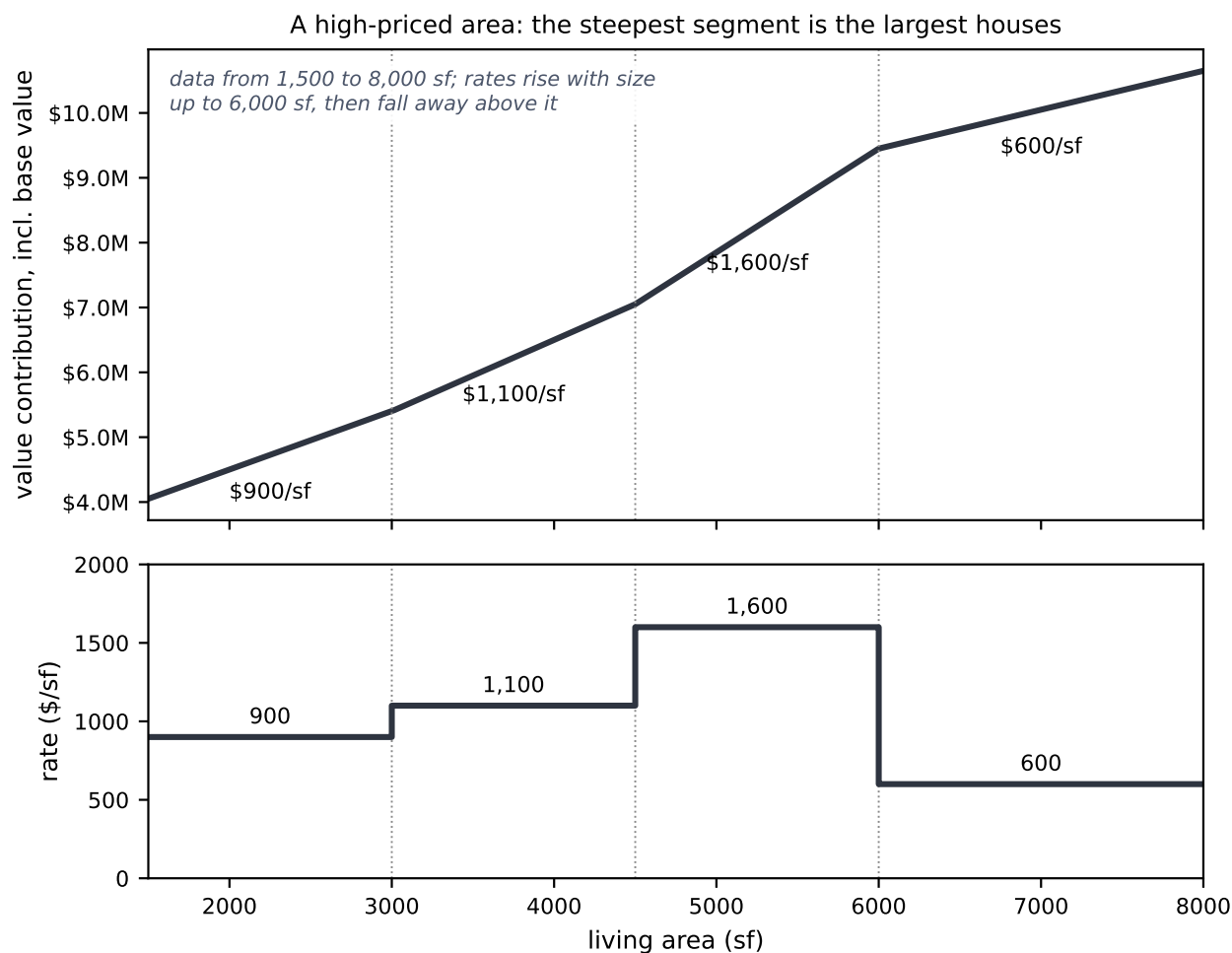


Figure 5: 根據 1,500 至 8,000 平方英尺的數據，此為高價區： $-900h(3000 - gla) + 1100h(gla - 3000) + 500h(gla - 4500) - 1000h(gla - 6000)$ ，在第一個結處與 \$5,400,000 相交。每平方英尺價格走勢 \$900、\$1,100、\$1,600 及 \$600：漲幅最陡的區間為 4,500 至 6,000 平方英尺，此處正是買家集中的地帶；首個區間漲幅較為平緩，而超過 6,000 平方英尺後，漲幅便逐漸趨緩。規則未變，唯市場已然不同。

情況二：有交互作用

交互項是指係數乘以不同變數中各因子之積，其中每個因子可為鉸點或變數本身，例如： $\gamma \cdot h(gla - 1500) \cdot lot$ 。案例 1 中的所有內容均得以保留，僅有一個詞語有所不同：區間內的單價現在是其他變數的函數，而非常數。以下是該陳述及其證明，接著說明這對估價師的算術運算有何影響。

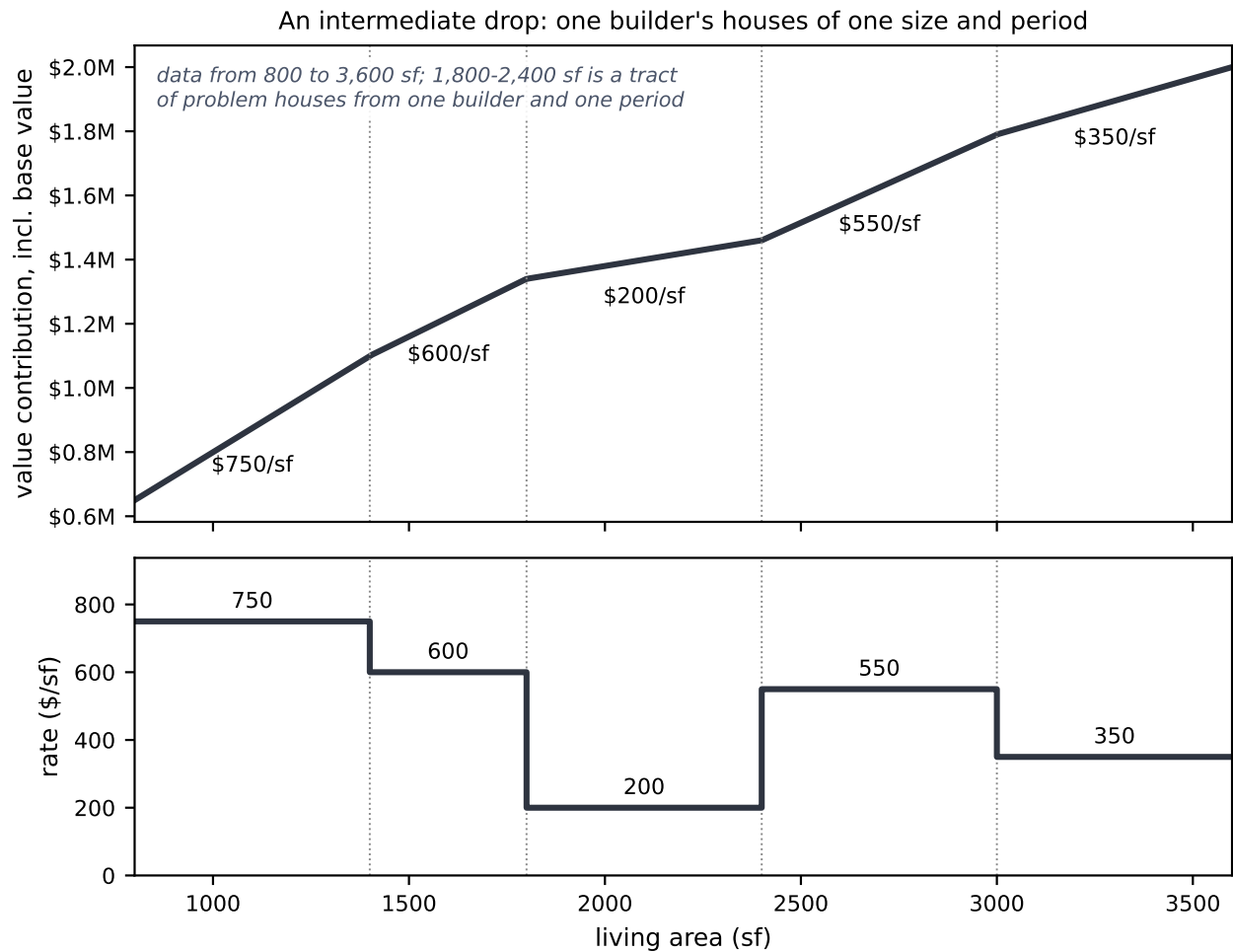


Figure 6: 根據 800 至 3,600 平方英尺的數據，出現一次中段下降： $-750h(1400 - gla) + 600h(gla - 1400) - 400h(gla - 1800) + 350h(gla - 2400) - 200h(gla - 3000)$ ，在第一個結處與 \$1,100,000 相交。每平方英尺的價格分別為 \$750、\$600、\$200、\$550 及 \$350。假設 1,800 至 2,400 平方英尺的區段是一片由同一建商於同一時期興建、且存在市場價格問題的同尺寸住宅區（此為構思示意）；該區段的價格在此處降至 \$200，並在該區段之上回升。若讀者將 -400 該係數視為 1,800 處於該區間的價格，或假設價格僅隨規模擴大而下降，便會誤解該地塊的情況。

設定。一個包含特徵間交互作用的 MARS 模型 $x, z_1, \dots, z_q$ 已擬合出以下值：

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

其中 M 為截距以外的項數， β_0 為截距，而該模型即為截距加上各項之和，其中第 m 項為 $\gamma_m \prod_{k \in F_m} B_{m,k}(z_k)$ ：其係數 γ_m 乘以該項中每個特徵對應的一個因子。此處 $z = (z_1, \dots, z_q)$ ，且為使其中一個索引遍歷所有特徵， z_0 代表 x 。出現在項 m 中的特徵構成集合 $F_m \subseteq \{0, 1, \dots, q\}$ ，單一項中不會出現重複的特徵，且每個因子 $B_{m,k}$ 皆為單一特徵 z_k 的函數，並分為三種類型：特徵本身 z_k ；右鉸鏈 $h(z_k - t)$ ；或反射鉸鏈 $h(t - z_k)$ 。設 $M_x = \{m : 0 \in F_m\}$ 為包含 x 的項。對於 $m \in M_x$ ， $B_{m,0}(x)$ 是該項在 x 中的因子，而該項的其他特徵——即那些 z_k 且 $k \in F_m, k \neq 0$ 的項——則是該項中 x 的配對因子。寫作

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

表示夥伴因子的乘積，而空乘積則為 1。換言之：一旦主語的其他特徵填入後，項的夥伴因子相乘即得一個普通數值；該數值乘以列印係數即為普通居住面積項的係數，且該項的讀法與案例 1 相同。伴侶積 $P_m(z)$ 是項 m 在從 x 中抽出其因子後所剩的部分：項 m 即為 $\gamma_m B_{m,0}(x) P_m(z)$ 。在某個特定點 z 處，它是一個數，因此在一條線段上，該項對 x 的作用如同一個單一鉸鏈，其係數為 $\gamma_m P_m(z)$ ，而該積即為伴侶進入單價的方式。沒有伴侶的項具有 $P_m(z) = 1$ ，屬於案例 1 的項。

例如，在下文範例中的 $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ 這個項中，假設居住面積為 x ，則 x 中的係數為 $h(\text{gla} - 1200)$ ，其中一個因子是地塊，且 $P_m(z) = h(\text{lot} - 6000)$ 。以 5,000 平方英尺的地塊為例 $P_m(z) = 0$ ，該項對費率無影響；以 10,000 平方英尺的地塊為例 $P_m(z) = 4,000$ ，該項在 1,200 處作為樞點，係數為 $0.03 \times 4,000 = 120$ ，使超過 1,200 平方英尺的部分，費率每平方英尺增加 \$120。此處並未限制一項可能包含多少個特徵；地球的 **degree** 參數設定了該上限，而以下定理對任何階數皆成立。一項包含 d 個特徵的項，其圖形具有 $d + 1$ 維度，一旦 d 超過二，便無人能繪製該圖；該軟體僅顯示切片，每次僅呈現一或兩個特徵，其餘特徵則保持固定；而定理 3 即是在主體自身點處進行相同的切片操作。 $P_m(z)$ 即是該切片操作發生的位置：無論項有幾個夥伴，在某一點上，它們的乘積皆為一個數值。在四項式 $\gamma h(\text{gla} - 1200) h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$ 中，當 $\gamma = 0.001$ 時，其圖形具有五維度， $P_m(z) = h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$ ；若以 10,000 平方英尺的地塊、20 年的房齡及 3 間浴室為例，則 $4,000 \times 10 \times 3 = 120,000$ ，且沿著居住面積，整個項在 1,200 處形成鉸點，係數為 $\gamma \times 120,000 = 0.001 \times 120,000 = 120$ ，每平方英尺增加 \$120：這與上述的兩項係數相同，但係數值卻小了三十倍。這就是為什麼當其組成項以大單位衡量時，高次項的係數往往會呈現為極小的數值：因為它會乘以多個組成項數值的積。此外，若任何一項參數位於其節點的零側（例如低於 6,000 平方英尺或年齡低於 10 年），則 $P_m(z) = 0$ ，該項將不產生任何貢獻，無論其他參數為何；項次越高，在特定屬性下該項被關閉的可能性就越多。本補充資料中的範例採用雙向交互作用（**degree** = 2）。

設 $t_1 < \dots < t_K$ 為 $B_{m,0}$ 中屬於鉸鏈的各條不同結，其中 $t_0 = -\infty, t_{K+1} = +\infty$ 以及線段 $S_i = (t_i, t_{i+1})$ 與案例 1 相同。對於 $m \in M_x$ 及 $i \in \{0, \dots, K\}$ ，定義

$$s_{m,i} = \begin{cases} +1 & \text{若 } B_{m,0} \text{ 為特徵 } x \text{ 本身，或為滿足 } t_j \leq t_i \text{ 的右鉸鏈 } h(x - t_j), \\ -1 & \text{若 } B_{m,0} \text{ 為滿足 } t_j \geq t_{i+1} \text{ 的反射鉸鏈 } h(t_j - x), \\ 0 & \text{其他情況。} \end{cases}$$

一個點是其他特徵的值 z 。由於每個 $B_{m,0}$ 在 S_i 上皆為線性或恆等於零，因此 f 在 S_i 上對於 x 呈線性關係，且 z 為固定值；其斜率即為在 S_i 上於點 z 處的單價，對於 $x \in S_i$ 而言，可寫作 $r_i(z) = \partial f / \partial x$ 。

若要一次查看所有符號，請參照下方範例中 Earth 列印的模型：截距 900,000， $600 h(\text{gla} - 1200)$ ， $-350 h(\text{gla} - 2400)$ ， $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ 以及 15lot 。其特徵包含居住面積 $x = z_0 = \text{gla}$ 及地塊面積 $z_1 = \text{lot}$ ，因此 $q = 1$ ；截距為 $\beta_0 = 900,000$ ，且包含 $M = 4$ 項。以一處居住面積為 2,000 平方英尺、地塊面積為 10,000 平方英尺的房產為例：

m	γ_m	F_m	$B_{m,0}(x)$	$B_{m,1}(z_1)$	$P_m(z)$	該物業處的項值
1	600	$\{0\}$	$h(\text{gla} - 1200)$	無	1	$600 \times 800 = 480,000$
2	-350	$\{0\}$	$h(\text{gla} - 2400)$	無	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	$\{1\}$	無	lot	(不在 M_x 中)	$15 \times 10,000 = 150,000$

因此， $M_x = \{1, 2, 3\}$ ：第 4 項在 x 中沒有因子，無論居住面積為何皆增加相同數值，且不進入 x 的單價中。擬合值為 $f = 900,000 + 480,000 + 0 + 96,000 + 150,000 = \$1,626,000$ 。第 1 項和第 2 項沒有配對項，因此 $P_1(z) = P_2(z) = 1$ ；第 3 項有一個配對項，且 $P_3(z) = h(10,000 - 6,000) = 4,000$ ，此即上述範例中的係數。由此得出的定理 3，將這些轉化為單價。居住面積具有節點 $t_1 = 1,200$ 和 $t_2 = 2,400$ ，且該物業位於兩者之間的線段 S_1 上，在此處第 1 項與第 3 項有效，而第 2 項無效 ($s_{1,1} = s_{3,1} = 1$, $s_{2,1} = 0$)；因此 $r_1(z) = 600 \times 1 + 0.03 \times 4,000 = 720$ ，即範例表格中 10,000 平方英尺地塊的單價。

定理 3 (含交互作用的區段單價)。採用此記號，對於任意 i 及任意點 z ，

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

定理 3 的證明。固定 i 、點 z ，以及 $x \in S_i$ 。含有 $m \notin M_x$ 的項在 x 中不含任何因子，且在 x 中為常數；其在 x 中的導數為 0。含有 $m \in M_x$ 的項等於 $\gamma_m B_{m,0}(x) P_m(z)$ ，其中 $P_m(z)$ 不依賴於 x 。根據定理 1 證明中的案例分析，應用於單一因子 $B_{m,0}$ 時， $B_{m,0}$ 在 x 處的導數，對於特徵本身為 1；對於右鉸點，若其結點位於 t_i 處或其左側，則為 1；若位於 t_{i+1} 處或其右側，則為 0；對於反射鉸點，若其結點位於 t_{i+1} 處或其右側，則為 -1；若位於 t_i 處或其左側，則為 0；也就是說，它等於 $s_{m,i}$ 。因此，該項的導數為 $\gamma_m s_{m,i} P_m(z)$ ，而 f (一個有限級數) 的導數即為這些項的總和。其值不取決於選擇哪個 $x \in S_i$ 。■

定理 1 是一個特例，其中 M_x 中的任何一項都沒有配對項，因此每個 $P_m(z) = 1$ ；將和式中的各項按其結進行分組，即可得到該處所證明之公式。三個推論說明了配對項如何變化；其中最後一個推論指出「每個結一個」的界限不成立，以及下文範例所示的「項可能為零」的情況。

推論 4 (單價僅在某一點處為數值)。在案例 2 的設定下，對於每個 i ，單價 r_i 是點 z 的函數。當每個具有 $s_{m,i} \neq 0$ 的項 $m \in M_x$ 均無配對項時，該單價在 z 內為常數。當某個具有 $s_{m,i} \neq 0$ 的項有配對項時， $r_i(z)$ 即為函數 $z \mapsto \sum_m \gamma_m s_{m,i} P_m(z)$ ，其值會隨配對項的數值變化，惟當各項完全抵銷時除外；因此，在同一段上具有相同 x 的兩個屬性，只要其配對項的乘積不同且差異無法抵銷，其 x 的單價便會不同。這正是程式為何需要輸入一個點，以及為何不存在僅針對單一特徵的獨立單價表的原因。

證明。根據定理 3， $r_i(z) = \sum_{m \in M_x} \gamma_m s_{m,i} P_m(z)$ 。若每個具有 $s_{m,i} \neq 0$ 的項均無配對項，則每個此類 P_m 皆為空積 1，且該和為常數。否則，該和至少包含一個和項 $\gamma_m s_{m,i} P_m(z)$ ，其中 $\gamma_m s_{m,i} \neq 0$ ，且 P_m 為一個或多個因子的積，而每個因子皆為某個配對項的非常數函數；因此，該和僅在 z 下為常數，且僅當這些和項完全抵消時才成立。■

推論 5 (在 x 的某一段上，單價在每個夥伴中皆為可變圖)。在案例 2 的設定下，固定 i 、某個夥伴 z_p ($p \geq 1$)，以及所有其他特徵的數值。此時， $z_p \mapsto r_i(z)$ 是 z_p 的連續分段線性函數，其分段點位於 M_x 中各項的因子 $B_{m,p}$ 的節點之間。沿著 x (當每個夥伴均被固定時)，單價僅在 x 的節

點 $t_1, \dots, t_K$ 處發生變化；而在 x 與其夥伴的聯合空間中，單價曲面同時被 x 的節點以及每個夥伴的節點所切割。

證明。在其他特性固定不變的情況下，每個 $P_m(z)$ 皆為某個常數乘以 1（若 $p \notin F_m$ 成立）或單一因子 $B_{m,p}(z_p)$ ，而該因子本身即是 z_p ，亦即 $h(z_p - t)$ 或 $h(t - z_p)$ ：這是 z_p 的連續分段線性函數，且至多僅在結點處有一個斷點。此類函數的有限線性組合——這正是定理 3 針對 $r_i(z)$ 所給出的結果——是連續且分段線性的，且僅在某些加數發生斷點之處才出現斷點。 x 上的陳述即是定理 3 本身：對於固定的 z ， $r_i(z)$ 在每個 S_i 上皆為一個數，且各段僅在 x 的結點處相交。 ■

推論 6（定理 1 的上界不成立）。在案例 2 的設定下，定理 3 中 $r_i(z)$ 公式中含有非零加項的項數，僅受 $|M_x|$ （即含有 x 的項數）所限制，且可能超過 K ：一個 x 的結，可能出現在多個含有不同夥伴的 M_x 項中，且每個夥伴皆為不同的項。反之，若某項包含 $s_{m,i} \neq 0$ ，則當其配對積為零時，該項便不會產生任何貢獻。

證明。項 $\gamma_1 h(x - 1500) \text{lot}$ 與 $\gamma_2 h(x - 1500) h(\text{age} - 10)$ 是 M_x 中具有相同結點的兩個不同項；在 1,500 之上的線段上，兩者皆為 $s_{m,i} = 1$ ，因此兩者皆會貢獻項；而 $r_i(z) = \gamma_1 \text{lot} + \gamma_2 h(\text{age} - 10)$ 包含來自同一個結點的兩個加數；若繼續加入更多配對項，便會在同一結點上無限制地產生更多項。至於逆命題，考慮一個其夥伴為鉸點 $h(\text{lot} - 6000)$ 的項，該點處的 $\text{lot} < 6000$ ：此時 $P_m(z) = 0$ 且該項歸零，儘管 $s_{m,i} \neq 0$ ；但在 $\text{lot} > 6000$ 的點處則不成立。 ■

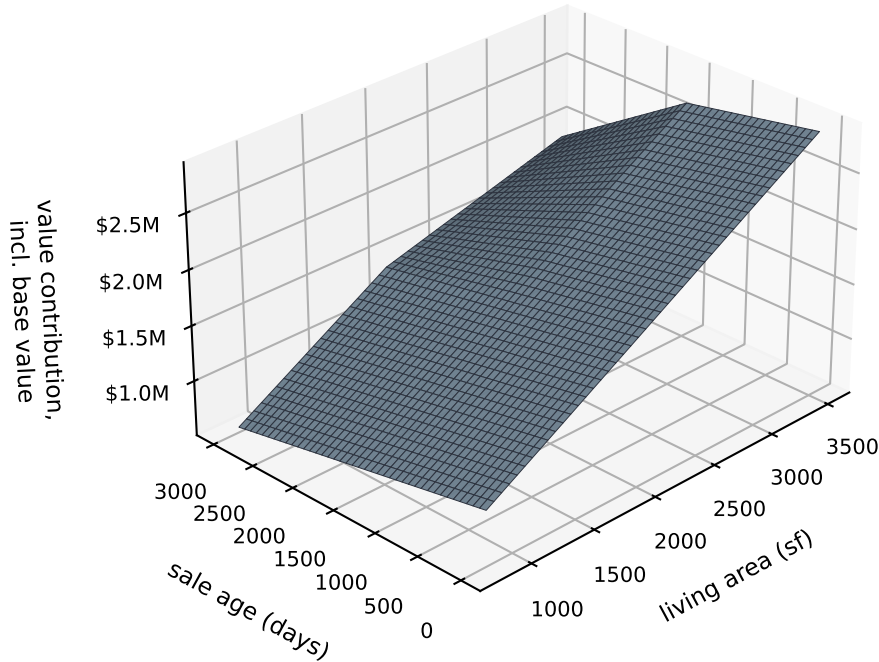
在此應提及編輯根據自身評估所得的一項實務觀察。一旦允許交互作用，軟體針對該變數單獨繪製的居住面積圖通常呈現為直線（即單一單價）；而若無交互作用，原本會出現在居住面積曲線上的彎曲部分，則會轉而顯現在交互作用項中，典型例子包括居住面積與銷售年齡（即銷售日與生效日之間的時間）的交互作用，或是居住面積與浴室數量的交互作用。定理 3 闡明了為何該直線並非實際率。earthUI 中的「每變數圖」會繪製無配對項的項；而 plotmo 的「一階圖」則會繪製整個模型的切片，其中其他變數均設為其中位數。在其他地方，交互作用項會以曲面形式呈現（例如 plotmo 的第一階與第二階圖表，以及 Milborrow 2024b 中的圖表，earthUI 會將這些圖表並列顯示）。對於任何實際房產，居住面積的費率等於直線的斜率，再加上每個有效的交互項，其係數乘以其配對因子的數值；該單價會在這些交互項的居住面積節點處發生變化，而變化幅度取決於配對因子：六年前的一筆交易與上個月的一筆交易，或是兩間浴室的房屋與四間浴室的房屋，即使居住面積圖表顯示為同一條直線，它們所對應的居住面積費率表卻各不相同。若讀者將該直線視為率值，便又遭遇了更高一級的障礙；而目標點處的率值表正是突破此障礙的途徑。圖 7 依照軟體的運作方式繪製了此類模型，呈現為居住面積與成交年份的曲面圖，並據此讀取兩個成交年份下的居住面積變數圖。

圖 8 則從圖 2 所示的市場出發，以相同概念進行反向分析。針對該市場建構了一個包含「居住面積」與「銷售年份」交互作用的模型，使得「居住面積」的主效應呈現為一條直線，\$500 每平方英尺，而四個交互作用項則產生彎曲。軟體針對居住面積自行繪製的圖表即為該直線。若以四年屋齡的交易為基準，居住面積的價格比值分別為圖 2 所示的每美元對應值：\$600、\$1,000、\$750 及 \$100；若以新屋交易為基準，則呈現直線；若以兩年屋齡為基準，則介於兩者之間。該模型是為了闡明觀點而建構的，並非根據銷售數據擬合而成；若進行真正的重新擬合，結果也不會像圖 2 那般完美吻合；但其運作機制正是真實重新擬合所採用的，這也是為何當估價師在加入交互作用項後，看到居住面積圖線趨於平直時，並未因此遺失那些彎曲部分。這些彎曲已轉移至交互作用項中，唯有在特定點位讀取時才能恢復這些彎曲。包含交互作用項的模型之主效應圖僅繪製變數的單項，這代表的是在所有交互作用項皆為零的夥伴設定下的切片（此處為一筆新交易），而非市場的整體形狀；請參閱該變數在該主題下的圖表。

最後一項影響主要涉及調整幅度而非匯率本身，而這正是影響銷售價格表的因素。

定理 4（調整的分解取決於順序）。在案例 2 的設定下，設 $y = z_p$ ($p \geq 1$) 為其他特徵之一，並以 w 代表其餘特徵，以 z_k 代表 $k \notin \{0, p\}$ ；將 $f(x, y, w)$ 寫為 f ，其參數依此分組。假設一項可比值位於 (x_c, y_c, w) ，一項主體位於 (x_s, y_s, w) ，兩者皆為相同的 w 。對於 y 的值 y^* ，定義 x 的調整

Living area by sale age: the surface the software draws



The living-area variable graph, read at two points

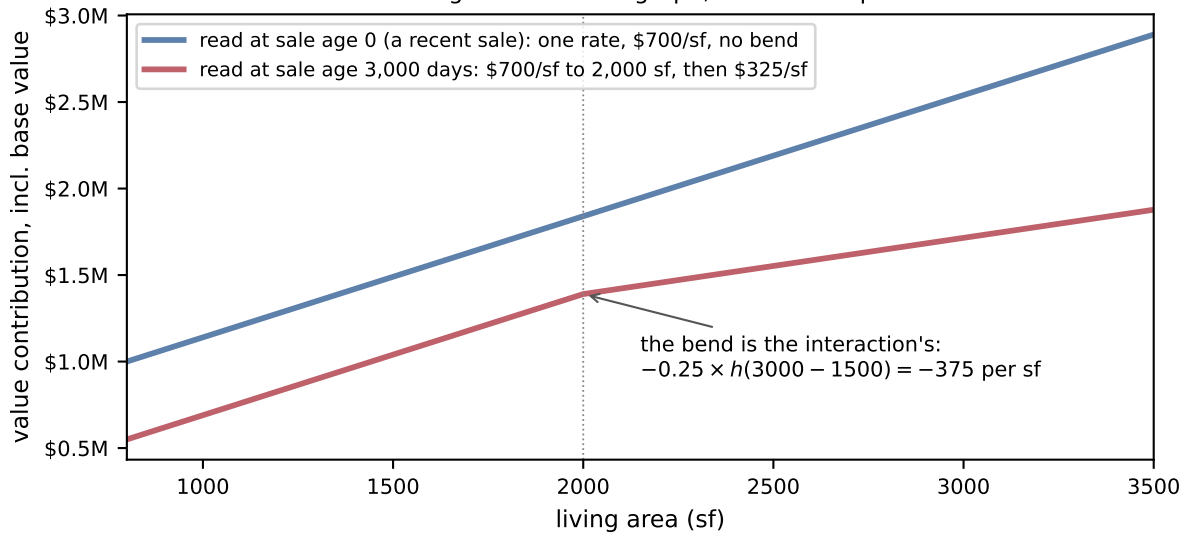


Figure 7: 居住面積與銷售天數之間的交互作用： $1,000,000 + 700 h(\text{gla} - 800) - 150 \text{saleage} - 0.25 h(\text{gla} - 2000) h(\text{saleage} - 1500)$ ，數據範圍為 800 至 3,500 平方英尺，且銷售天數不超過 3,000 天。上圖為建模軟體針對此一對變數繪製的曲面圖。下圖為居住面積變數的圖表，依據定理 3 的要求，於兩點進行讀取：對於任何銷售天數不超過 1,500 天的交易，伴侶因子 $h(\text{saleage} - 1500)$ 為零，交互作用不產生任何影響，且圖表呈現為每平方英尺 \$700 的直線；在發生於 3,000 天前的交易中，伴侶因子為 1,500，當面積超過 2,000 平方英尺時，交互作用會減去 $0.25 \times 1,500 = 375$ 每平方英尺，此時該區間的單價為 \$325。居住面積的主效應（僅包含居住面積項，無交互項）在兩種情況下均為同一條直線；彎曲部分源自交互作用，且僅在夥伴項生效時出現。

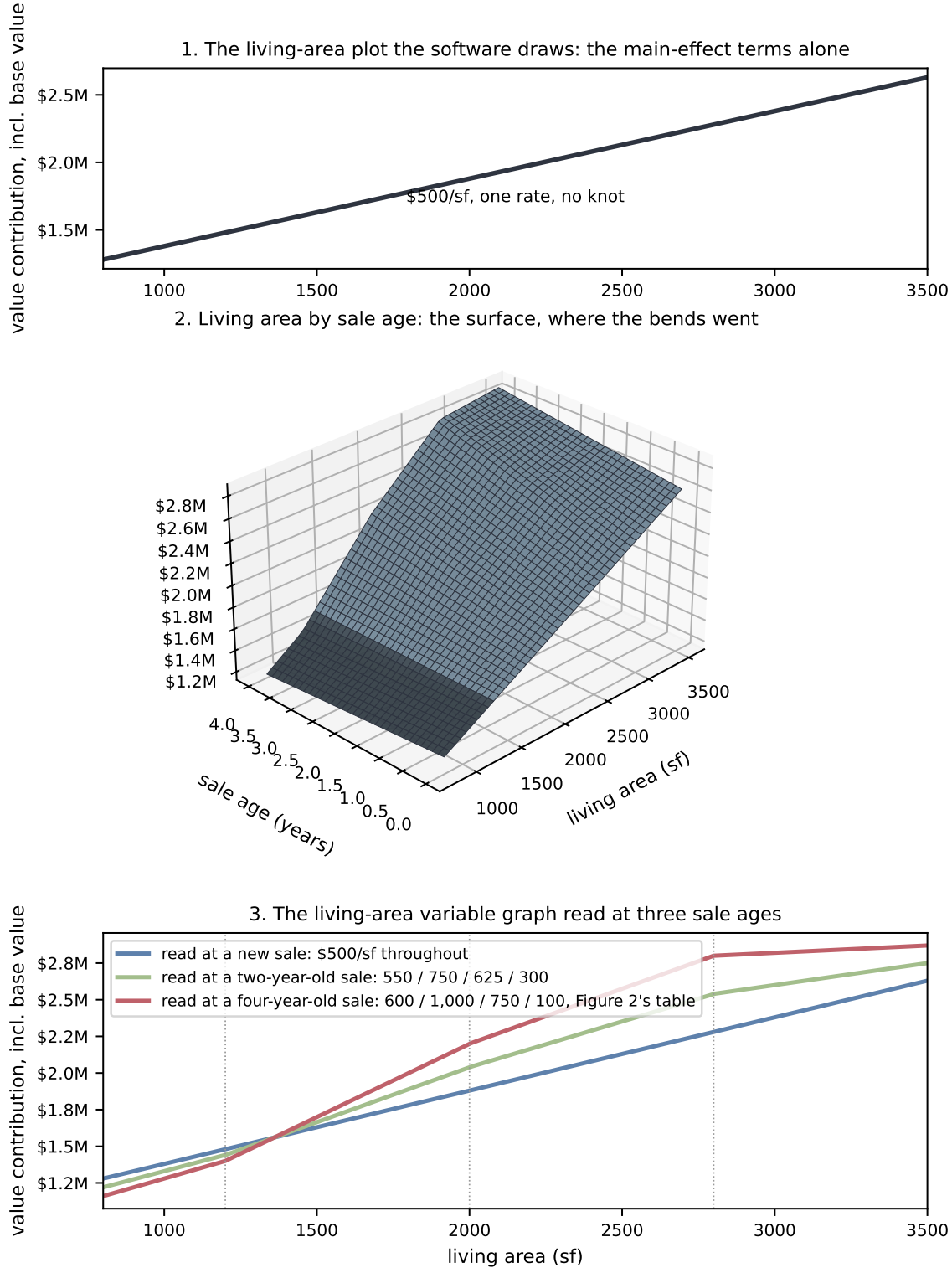


Figure 8: 圖 2 所示的市場，其中居住面積與銷售年限呈現交互作用： $1,280,000 + 500h(\text{gla} - 800) - 20,000\text{saleyrs} + [-25h(1200 - \text{gla}) + 125h(\text{gla} - 1200) - 62.5h(\text{gla} - 2000) - 162.5h(\text{gla} - 2800)] \times \text{saleyrs}$ ，銷售年限（單位：年）範圍為 0 至 4。各子圖的說明詳見其標題；在 4 年時，下方圖表的單價與圖 2、\$120,000 所示的單價完全相同，但數值較低。此圖僅為示範之用，未經擬合。

值，其中 y 維持在 y^* 為

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

設 $M_{x,y}$ 為同時包含 x 與 y 的項。每個此類項皆為其係數乘以三部分， $\gamma_m A_m(x) C_m(y) Q_m(w)$ ： $A_m(x) = B_{m,0}(x)$ 是其在 x 中的因子，該項本身即為 x 或 x 中的鉸點； $C_m(y) = B_{m,p}(y)$ 是其在 y 中的因子，該項本身即為 y 或 y 中的鉸點；而 $Q_m(w)$ 則是其在其餘特徵中的因子的乘積，每個此類特徵本身即為該項或其中的鉸點，且當該項無其他因子時，此項即為 1。在下方的範例中，其中一項術語為 $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ ，其因式為 $A_m(x) = h(\text{gla} - 1200)$ 、 $C_m(y) = h(\text{lot} - 6000)$ 及 $Q_m(w) = 1$ 。針對 x 的調整，請分別寫出 $\Delta_x(y_s)$ 和 $\Delta_x(y_c)$ ，其中 y 在標的物處取值 y_s ，在可比物處取值 y_c ，亦即 $\Delta_x(y^*)$ 在 $y^* = y_s$ 處，以及在 $y^* = y_c$ 處；在範例中， x 代表居住面積， y 代表地塊面積， $\Delta_x(y_c) = \$240,000$ 對應可比物 5,000 平方英尺的地塊， $\Delta_x(y_s) = \$288,000$ 則對應標的物 10,000 平方英尺的地塊。然後

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

而總差值 $f(x_s, y_s, w) - f(x_c, y_c, w)$ 則無需選擇 y^* 。該陳述涉及此兩特徵切片：其中一項特徵 y ，其餘所有特徵在兩者中均保持相同 w ；當兩者之間有三項或更多特徵不同時，總和在這些特徵間的分配方式，會沿著更長的一特徵變動鏈以相同方式呈現路徑依賴性，且該定理適用於鏈中的每一步，此時尚未變動的特徵保持在其當前值。特別是，當僅存在單一此類項 $\gamma A(x) C(y)$ 且 w 中無因子時，針對 x 的兩項調整值相差 $\gamma [A(x_s) - A(x_c)] [C(y_s) - C(y_c)]$ ；當 x 中的變動改變了 A ，且兩項性質在 C 上有所差異時，此差值並非 0。對於 $A = h(x - t)$ 與 $C = y$ ，此差值為 $\gamma [h(x_s - t) - h(x_c - t)] (y_s - y_c)$ ；對於 $A = h(x - t)$ 與 $C = h(y - s)$ （即下方演算範例的形式），則為 $\gamma [h(x_s - t) - h(x_c - t)] [h(y_s - s) - h(y_c - s)]$ 。

定理 4 的證明。將 f 中的各項依其所含的 x 、 y 進行分組。一個既不包含 x 也不包含 y 的項，在 (x_s, y^*, w) 和 (x_c, y^*, w) 處取相同值，並在 $\Delta_x(y^*)$ 處相互抵銷。一個包含 y 但不包含 x 的項，同樣會相互抵銷。一個包含 x 但不包含 y 的項，對 $\Delta_x(y^*)$ 的貢獻量不依賴於 y^* ，因此會在 $\Delta_x(y_s) - \Delta_x(y_c)$ 處相互抵銷。一項 $m \in M_{x,y}$ 對 $\Delta_x(y^*)$ 貢獻了 $\gamma_m [A_m(x_s) - A_m(x_c)] C_m(y^*) Q_m(w)$ ；設定 $y^* = y_s$ 與 $y^* = y_c$ 並進行減法運算，即可得到所述的和。總差值是同一函數 f 的兩個值之間的差，且未提及 y^* 。 ■

因此，交互作用的數值屬於這對特徵，而非其中任何一項。一種是將各特徵依序調整至可比物件數值的網格，另一種則是依標的物數值進行調整的網格；兩者在總數上達成一致，但在分配比例上則存在分歧。將其他特徵維持在標的物數值上，是一種可辯護的慣例，也是本程式所採用的方式；報告必須說明採用了哪種慣例。標準的住宅評估網格（例如《統一住宅評估報告》（URAR）所採用的），並未設有交互作用的欄位。若逐項調整，這兩種操作必須將配對項目置於相對兩端——先將居住面積調整至標的物地塊的數值，再將地塊調整至可比物件居住面積的數值，或反之——如此一來，交互作用僅計入一次，且各行數值相加即等於模型的總差異。若直接計算各特徵的貢獻值，評價網格亦可將交互作用作為獨立一行，列出標的物與可比物之間交互作用項的差值，並置於各特徵行旁；此時每行即代表一組術語，且無需選擇特定順序。在以下實例中，三行分別為居住面積 \$240,000、地塊面積 \$75,000，以及按地塊計算的居住面積 \$120,000，其總和為相同的 \$435,000。在存在交互作用的情況下，該對的割線單價會隨著該對的變化而產生兩種變化：一種是透過交叉的範圍，另一種則是透過該對在夥伴中的位置。

含交互作用的算例

又是一個沿海市場，這次包含兩個特徵——居住面積和地塊大小——以及這兩者之間的一項互動關係。earth prints：

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

請參照定理 3 所述。生活區有兩個結點，分別為 1,200 和 2,400，且在 M_{gla} 中包含三個項：兩個普通鉸鏈以及一個交互作用，其配對點為 $h(\text{lot} - 6000)$ 。在 1,200 至 2,400 的區間內，位於 1,200 的普通鉸鏈處於活躍狀態，而互動項位於 1,200 的鉸鏈亦然，因此該區間的單價為 $600 + 0.03 h(\text{lot} - 6000)$ ；此為地塊的變動圖，針對 6,000 平方英尺以下的地塊，單價固定為 600；超過此面積後，每增加一平方英尺，單價即增加 3 美分。超過 2,400 時，2,400 處的鉸點生效，單價為 $250 + 0.03 h(\text{lot} - 6000)$ 。地塊本身的單價在 6,000 平方英尺以上為 $15 + 0.03 h(\text{gla} - 1200)$ ，6,000 平方英尺以下為 15，因此每個特徵的單價取決於另一特徵的數量。

兩點、兩張表格。該程式針對 800 至 3,500 平方英尺居住面積的單價表，分別以 5,000 平方英尺地塊和 10,000 平方英尺地塊為例進行計算。每欄皆為該模型在居住面積下的斜率，且僅以該特定地塊面積值為基準，不包含其他值：

居住面積（平方英尺）	第 5,000 號地塊的單價（每平方英尺 \$）	第 10,000 號地塊的單價（每平方英尺 \$）
800 至 1,200	0	0
1,200 至 2,400	600	720
2,400 至 3,500	250	370

在較小的區間內，夥伴因子 $h(5000 - 6000)$ 為零，因此交互作用不產生任何貢獻，而該表格與不含交互作用的模型所給出的結果相同；這即是推論 6 中「項可消失」的情況。在較大的地塊中，夥伴因子為 4,000，交互作用會為每個高於 1,200 的區段增加 $0.03 \times 4,000 = 120$ ；且同一個結點 1,200 同時饋入兩個活躍項，這正是「每個結點僅一個項」限制未能成立的例證。

現在來進行調整。該可比物件佔地 5,000 平方英尺，建築面積為 1,800 平方英尺，估算價值為 \$1,335,000；本案物業佔地 10,000 平方英尺，建築面積為 2,200 平方英尺，估算價值為 \$1,770,000。需解釋的總差額為 \$435,000。首先根據可比物件的地塊面積調整目標物件的居住面積，接著再根據目標物件的居住面積調整地塊面積。每次調整皆為其所比較之兩項數值之間的差額；若調整範圍僅限於單一區段內，該差額即為寬度乘以單價，這正是表格中的計算方式：

	先看居住空間，再看地塊	先看地塊，再看居住空間
居住面積調整	$400 \times 600 = 240,000$	$400 \times 720 = 288,000$
地塊調整	195,000	147,000
總計	435,000	435,000

地塊調整同樣是貢獻額的差異，而費率則解釋了這些差異；該地塊的專屬費率在 6,000 平方英尺處出現分段：以標的物 2,200 平方英尺為例，地塊費率在 \$15 以下，在 \$45 以上，因此從 5,000 到 10,000 的變化幅度為 $1,000 \times 15 + 4,000 \times 45 = 195,000$ ，而非 $5,000 \times 45$ ；以可比物件的 1,800 平方英尺為例，其計算結果為 $1,000 \times 15 + 4,000 \times 33 = 147,000$ 。總和相符，正如定理 4 所言；而各部分的差異為 \$48,000，這恰好是該定理的公式： $0.03 \times [h(2200 - 1200) - h(1800 - 1200)] \times [h(10000 - 6000) - h(5000 - 6000)] = 0.03 \times 400 \times 4,000$ 。該 \$48,000 即為額外 400 平方英尺房屋面積與額外 4,000 平方英尺地塊面積所帶來的總價值；此價值不屬於任一特徵，而無論哪一項特徵被調整在後，皆可獲得此價值。根據該程式的慣例，其他項目均維持在被測物件的估值，因此右欄顯示的居住面積為 \$288,000，在 \$720 處為每平方英尺，且其報告中會註明讀數的取值點：

```

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.

```

這兩項貢獻值均以該物業的 10,000 平方英尺地塊面積為基準計算；若以該可比物業自身 5,000 平方英尺的地塊面積為基準，其居住面積貢獻值則為 \$360,000。

一位評測者在對比機的測試場地重新執行此測試時，得到的結果是 \$240,000，這並無不妥；評測者與程式都在同一個表面上的兩個點進行讀取，而報告必須說明具體是哪兩個點。

簡而言之：若無交互作用項，MARS 模型即為按特徵劃分的單價表，而針對一般迴歸所提出的任何單價論點，皆可逐個區段套用。若包含交互作用項，單價則形成一個曲面，定理 3 精確闡明了其彎曲方式，而估價師必須指明是在哪一點讀取該曲面的。

程式

`mars_rates.pl`（完整內容收錄於附錄中）會讀取「earth」的列印表格，或以 Prolog 項形式撰寫的模型，並產出單價表、各節點的數值、兩項貢獻及其差值，調整值（附各區間的說明）、該對數的割線率，並驗證各區間數值的總和是否等於調整值。根據資料範圍（即銷售紀錄中面積最小與最大的房屋），若稅率超出此範圍，系統將不報告該稅率：鉸鏈的代數運算在單側會趨於無窮大，但銷售紀錄並非如此，因此超出範圍的稅率即屬外推。以 800 至 3,500 平方英尺的資料為例：

```

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.

該區塊中包含兩類數值。這些貢獻值僅包含居住面積項，並省略了截距項及其他所有特徵項：\$600,000 與 \$1,100,000 之間的差異恰好等於調整量。節點處的數值為擬合值，其中包含截距，而在包含交互作用的模型中，該點的伴侶項也包含在內；在此，1,200 平方英尺處的 \$1,520,000 即為截距，因為該處所有鉸點的值皆為零。在可比單位與研究對象處的擬合值，分別為 \$2,120,000 及 \$2,620,000，同樣包含截距。報告將節點值標示為擬合值，以免讀者誤將其視為貢獻項。

那段內容正是應列入報告附錄的部分：單價、結數，以及兩者之間的計算過程，且須以審閱者能親手重現的形式呈現。該程式透過 69 項測試，驗證了符號規則、交互作用規則、經調整後區段的一致性，以及對地球輸出資料的解析器，並在《期刊》的 Prolog 環境 Trealla、SWI-Prolog 及 Scyer Prolog 下均通過測試。對於使用 R 語言的讀者，無需離開 R 環境即可閱讀相同內容：隨程式一併發布的 `mars_rates.R` 檔案，可接收 `earth` 物件或其列印的表格，並輸出相同的區塊，其中單價值會被限制在資料範圍內；Prolog 程式仍是經核驗的參考版本，而 R 檔案則針對本增補資料中的範例進行自身的自我檢查。該程式、其測試程式碼及圖表腳本已隨本期期刊發布，作為 PDF 檔旁的下載檔案，並依據 GNU Affero 通用公共授權條款（版本 3 或更新版）發佈。

為什麼選 MARS，以及本文留給其他條目的問題

本期刊採用 MARS 進行建模有三個原因，而這正是本增刊得以存在的理由。客觀性：結點的放置位置取決於銷售數據的分布，遵循一套明定的程序與一套明定的修剪規則（即在後期階段移除不應存在於該位置的項），除上述設定（互動的最大度數、修剪懲罰值、結點的最小間距）之外，變量圖的形狀在事前並未被預先決定。可發現性：前向遍歷會搜尋每個特徵與每個候選結點（包含不超過指定程度的交互作用），並回報其發現結果；這點是任何從預先選定形狀出發的方法都無法做到的。可解釋性：在數據所選範圍內保持恆定單價，以及基於相同鉸點構建的交互作用，正是讓模型能以單價表形式呈現，並將調整值解釋為「單價乘以寬度」的關鍵。另有兩種方法同樣會產生隨特徵範圍變化而變化的單價，且它們彼此之間存在差異。第一種，分段（「斷棒」）迴歸，取一個變數及分析師選定的多個分段點，並估算各分段點的位置，每個分段點皆附有標準誤差（用以衡量該估計值在其他銷售數據下可能偏移的程度）；分段點之間，單價保持恆定，如同 MARS 模型。第二種方法是「形狀受限的加法模型」，它不使用直線區段，而是擬合一條平滑曲線，並強制符合分析師所設定的形狀——例如，該數值在居住面積增加時絕不會下降，或是每增加一平方英尺所帶來的增幅都小於前一次。本文中未採用這兩種方法。這兩種方法各自解答了本補充資料未提出的問題，且各自犧牲了上述三項特性中的至少一項：分段迴歸犧牲了可發現性，因為變數與分段數是預先選定的；形狀受限模型則犧牲了可解釋性，因為曲線沒有恆定的單價可供製表；同時也犧牲了客觀性，因為其形狀是被強加的。本補充資料解讀 MARS 產出的模型；至於方法的選擇，以及 MARS 從一組銷售數據到下一組所找到的節點與單價的穩定性，則將另行撰文探討。

本文未解決的問題

這些率值源自模型，而以下四項統計要點同樣適用於它們且無需調整：率值對條件期望的近似程度，僅限於模型形式所允許的範圍內（條件期望）；它可能包含被省略特徵的值（省略變數偏差）；當某個特徵隨結構性誤差波動時，該率可能會被扭曲（內生性）；且當兩個特徵總是同步變動時，無論銷售數據有多少，該率都可能完全無法由銷售數據確定（參數識別）。MARS 僅能在銷售數據中包含能顯示該變動的對比時，才找出費率變動之處；一個「結點」是對市場的論斷，其可靠性取決於該結點兩側銷售數據的品質。本補充說明所增添的內容則更為精確：在給定模型的前提下，此處闡述了如何詮釋模型所傳達的訊息。

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 Scyer 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], I0, I, Terms) :-
    strip(L, S),
    ( S == [] -> parse_lines(Ls, I0, 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, I0, 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).

```