

L'obstacle à la compréhension humaine de MARS (The Barrier to Human Understanding of MARS)

Lire un modèle MARS en taux unitaires, contributions à la valeur et ajustements

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

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

Il s'agit d'un supplément destiné à la rubrique « Fondements », et non d'une entrée thématique. Son titre fait référence à un obstacle que l'éditeur a fini par identifier au fil d'années d'observation indirecte : les évaluateurs qui interprètent sans difficulté un coefficient de régression s'arrêtent à la liste des termes pivots générée par un programme MARS, car aucun élément de cette liste ne ressemble à un taux. Les quatre articles statistiques qui suivent ce supplément (« *Espérance conditionnelle* », « *Biais de variable omise* », « *Endogénéité* » et *Identification des paramètres*) traitent du taux attribué à une caractéristique par une régression. La modélisation propre à la revue utilise des splines de régression adaptatives multivariées (MARS ; Friedman 1991) ¹, qui ne fournissent pas un taux par caractéristique mais un par plage, et le chemin menant de ce qu'affiche un programme MARS au tableau de taux qu'un expert immobilier peut justifier comporte une étape à laquelle les tableaux créés manuellement se trompent. Ce supplément présente ce cheminement, cette étape, ainsi qu'un programme qui le suit et vérifie son propre travail.

Définition formelle. Dans un modèle MARS, la valeur d'ajustement correspond à une ordonnée à l'origine plus une somme de termes, chacun étant égal à un coefficient multiplié par le produit des fonctions charnières des caractéristiques. La valeur d'ajustement représente l'estimation du prix fournie par le modèle. Pour une caractéristique donnée, les autres étant maintenues fixes, il s'agit d'une série de segments droits. Les points où les segments se rejoignent sont les nœuds. Le graphique représentant la contribution de la valeur de la caractéristique sur son intervalle est le *graphique de la variable*. Sa pente dans chaque intervalle correspond au *taux d'* : la variation de la valeur ajustée par unité de la caractéristique dans cet intervalle ; la *contribution de valeur* d'une quantité de la caractéristique est la somme des termes du modèle relatifs à cette caractéristique pour cette quantité ; dans l'exemple, tous ces termes sont nuls à 1 200 pieds carrés, de sorte que la contribution est mesurée à partir de ce point et est négative en dessous ; un modèle n'a pas nécessairement besoin d'une telle quantité, et seules les différences entre les contributions, qui constituent des ajustements, ont une signification ; et l' *ajustement* d'un bien comparable à un bien à évaluer correspond au montant ajouté à la contribution du bien comparable pour la rendre égale à celle du bien à évaluer, ce qui correspond à la somme, sur les plages comprises entre les deux quantités, du taux multiplié par la largeur. Le taux ainsi défini est celui du modèle ; il n'est interprété comme le taux payé par le marché qu'une fois que le modèle a été accepté comme une représentation fidèle des ventes, et les passages qui suivent ce complément traitent des cas où cette acceptation est justifiée.

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.

Pourquoi ce supplément existe

Un évaluateur qui effectue un ajustement avec un modèle MARS, dans earthUI (Craytor 2026) ou directement dans le package R earth (Milborrow 2024a), se voit proposer une liste de termes tels que ceux-ci :

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

Personne ne mentionne cela. Ce qui figure dans le tableau des ventes, c'est l'ajustement : la valeur ajoutée du bien à évaluer moins celle des biens comparables. Les taux au pied carré sont abordés dans le rapport, dans le cadre de la justification du modèle MARS. Le chemin qui mène des conditions imprimées à ces chiffres est court, mais il comporte une étape où les tableaux établis manuellement peuvent présenter des erreurs, et l'intérêt même d'utiliser un modèle est perdu si les chiffres qu'il produit ne peuvent pas être expliqués à un examinateur.

La version simple : quatre plages, quatre taux

Une régression linéaire qui intègre la surface habitable sous la forme d'un terme linéaire donne un taux unique pour celle-ci sur l'ensemble du marché ; c'est ce que font les régressions utilisées dans l'évaluation traditionnelle, et c'est ce que l'on entend par « régression ordinaire » dans le présent supplément. MARS permet au taux de varier là où les ventes indiquent qu'il varie. Sur le marché côtier de l'exemple, la surface habitable peut valoir \$600 par pied carré jusqu'à 1 200 pieds carrés, \$1,000 par pied carré entre 1 200 et 2 000, \$750 entre 2 000 et 2 800, et seulement \$100 au-delà. C'est là toute la différence, et c'est une différence qu'un agent comprend d'emblée : le marché paie le plus cher pour l'espace qui transforme une petite maison en maison familiale, moins cher pour la pièce suivante, et presque rien pour l'espace au-delà de ce que recherchent les acheteurs du quartier.

Compte tenu des taux, tout le reste relève de l'arithmétique qu'un expert effectue déjà. Un bien comparable de 1 800 pieds carrés et un bien à évaluer de 2 400 pieds carrés présentent un écart de 600 pieds carrés, et le nœud, situé à 2 000 pieds carrés, se trouve entre les deux. L'ajustement ajouté au bien de référence correspond à la valeur ajustée du bien à évaluer moins celle du bien de référence, soit $\$2,620,000 - \$2,120,000 = \$500,000$, ce qui correspond également à la contribution de la surface habitable du bien à évaluer moins celle du bien de référence, soit $\$1,100,000 - \$600,000$. Les taux l'expliquent : le changement franchit le seuil des 200 pieds carrés à \$1,000, puis celui des 400 pieds carrés à \$750, soit \$500,000 au total. La figure 3 l'illustre : sur une valeur d'ajustement (le prix que le modèle attribue à une maison présentant ces caractéristiques) allant de \$1,280,000 pour 800 pieds carrés, la plus petite maison du jeu de données, à environ \$3 million pour 3 500 pieds carrés, la plus grande, la majeure partie de la hausse se produisant avant 2 800 pieds carrés.

Le seul taux qu'une droite passant par ces deux points aurait pu utiliser pour ce déplacement est celui qui les relie, soit \$500,000 sur 600 pieds carrés, soit environ \$833 par pied carré : le *taux sécant* du couple, le taux effectif, ou mixte, pour ce déplacement-là : la moyenne des taux du modèle pondérée par la largeur de chaque plage traversée. Ce n'est pas l'un des taux du modèle, et cette moyenne serait différente pour la paire suivante. Une régression ordinaire, avec un seul taux pour la surface habitable, doit s'appliquer à chaque paire avec ce même chiffre ; elle ne peut donc pas

reproduire ces moyennes paire par paire ; c'est ce que fait un taux unique : il calcule la moyenne d'une variation que le modèle attribue au marché.

Un lecteur qui s'arrête ici a compris la méthode : il suffit de calculer la différence entre les deux contributions, puis de l'expliquer en lisant le taux correspondant à chaque tranche, en appliquant ce taux au bien à évaluer, et en additionnant les résultats. Ce qui suit s'adresse au lecteur qui souhaite comprendre pourquoi les taux sont ce qu'ils sont, où se trouve le seul piège, ce qui change lorsque les caractéristiques interagissent, et comment un programme vérifie les calculs ; le tableau des taux et l'ajustement ne changent pas.

La version affinée : les fonctions charnières et le signe que l'on oublie

On appelle « graphique de *variable* » le graphique représentant la contribution de la valeur d'une variable sur son intervalle de définition. Il ne s'agit pas d'une courbe au sens courant du terme, c'est-à-dire d'une ligne qui s'incurve en douceur ; il s'agit d'un graphique linéaire segmenté, rectiligne entre les nœuds et ne s'incurvant qu'au niveau de ceux-ci. MARS le construit à partir de *charnières* ; Friedman (1991) les a exprimées sous forme de fonctions de puissance tronquées, et le terme « charnière » est celui qui s'est imposé dans la littérature ultérieure (Milborrow, s.d.) ; Hastie, Tibshirani et Friedman (2009, §9.4) appellent les deux charnières situées au niveau d'un même nœud une « paire réfléchie ».

Définition (charnière). Une charnière est une fonction à une variable x présentant un seul point de rupture, appelé « *nœud* » t , dont la valeur est nulle d'un côté du nœud et, de l'autre côté, égale à la distance par rapport au nœud. Il y en a deux à chaque nœud, un de chaque côté, comme le montre la figure1 :

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

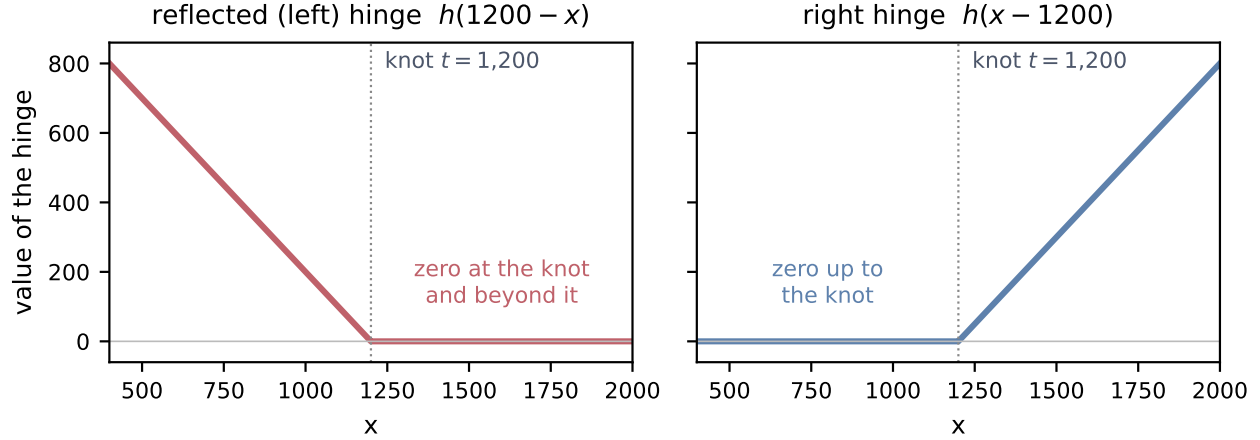
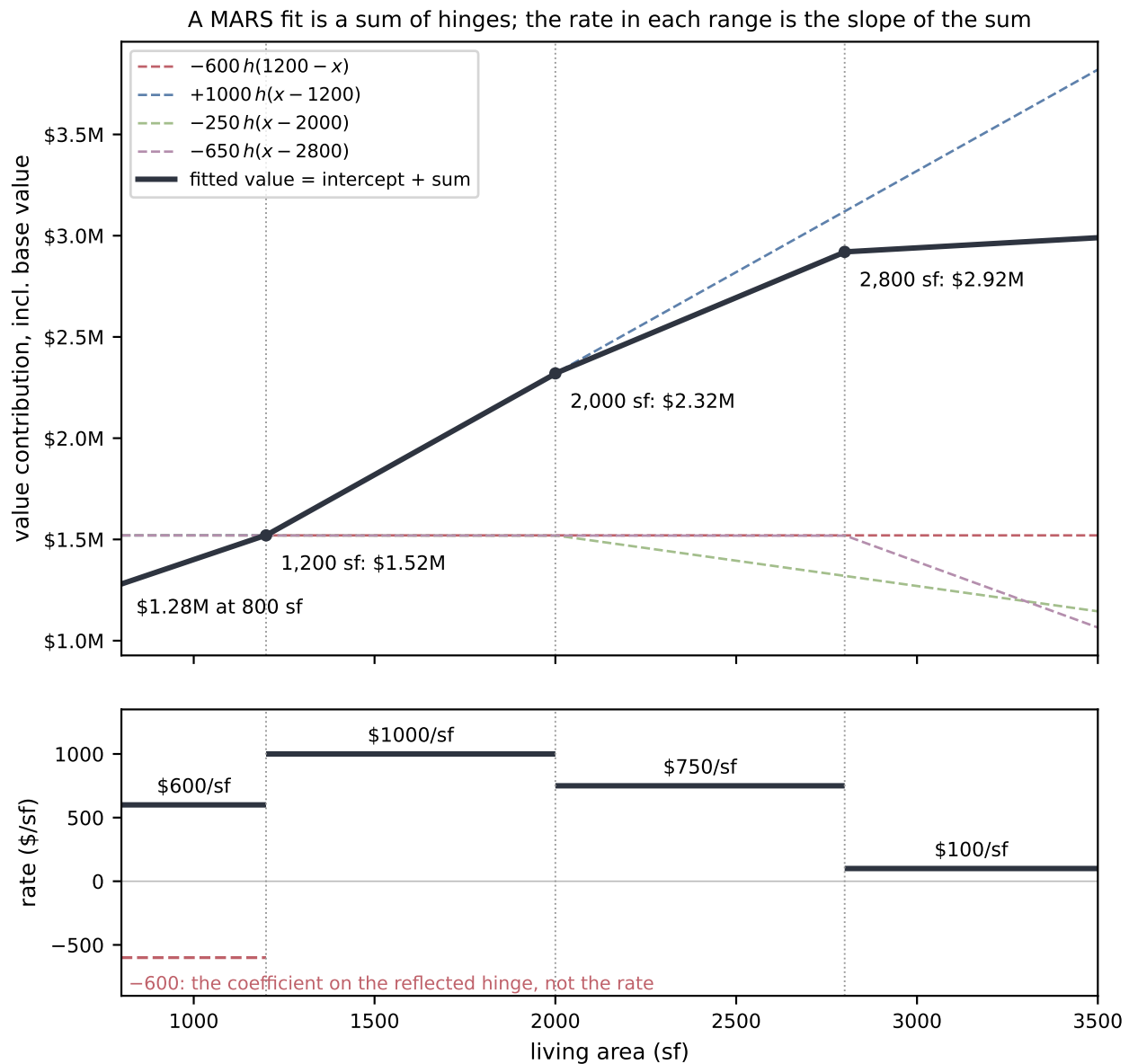


Figure 1: Les deux fonctions charnières avec le nœud à $t = 1,200$: la charnière réfléchie (à gauche) $h(1200 - x)$, décroissante de gauche à droite jusqu'à zéro au niveau du nœud et nulle au-delà ; la charnière de droite $h(x - 1200)$, nulle jusqu'au nœud et croissante de gauche à droite au-delà. Il s'agit là des charnières « nues » ; dans un modèle, chacune est multipliée par un coefficient, qui peut inverser le sens.

La première, $h(x - t)$, est la charnière *droite* : nulle jusqu'au nœud et, au-delà, croissante de gauche à droite. La seconde, $h(t - x)$, est la charnière *réfléchie* (ou gauche) : décroissante de gauche à droite jusqu'au nœud, et nulle au niveau du nœud et au-delà. Les deux sont des charnières au même sens ; elles ne diffèrent que par le côté du nœud où elles sont non nulles. Ces directions décrivent les charnières nues, tracées à partir de l'axe x dans la figure 1, et rien de plus. Dans un modèle, une articulation est multipliée par un coefficient, qui peut être négatif, puis additionnée à d'autres termes ; ainsi, un terme construit à partir d'une articulation de droite peut diminuer de gauche à droite et un terme construit à partir d'une articulation réfléchie peut augmenter, et le graphe des variables rencontre un nœud à la hauteur que lui confèrent les autres termes. Toutes les directions indiquées dans ce supplément se lisent de gauche à droite. Les charnières ne sont pas utilisées seules. Dans un modèle MARS, chaque charnière est multipliée par un coefficient et les termes sont additionnés ; ce sont à partir de ces coefficients, et non d'une charnière isolée, que sont calculées les pentes du modèle, c'est-à-dire les taux : de manière cumulative, plage par plage, le taux de chaque plage étant le taux de la plage précédente ajusté par les coefficients des charnières dont le nœud se situe entre elles. La charnière est le seul élément constitutif dont dispose MARS pour une variable unique ; chaque courbe d'un graphique de variable est un nœud d'une charnière, et chaque segment rectiligne est une somme de termes qui y sont soit nuls, soit linéaires. **Le point essentiel à retenir pour la suite : le coefficient d'une charnière n'est pas une pente. Il s'agit d'un ajustement de la pente, de la valeur ajoutée, au niveau du nœud de cette charnière, à la pente qui arrive de la gauche.** Lu de gauche à droite, un graphique de variable commence par le taux de son premier segment et chaque nœud ajuste ce taux à l'aide des coefficients qui s'y trouvent. C'est là toute la règle d'interprétation, et les théorèmes 1 et 2 ci-dessous en constituent la preuve. Le modèle est constitué de l'ordonnée à l'origine plus chaque coefficient multiplié par son point d'inflexion, et la pente dans n'importe quel intervalle correspond à la pente de cette somme.

Prenons les termes imprimés ci-dessus ; la figure 2 représente chacun d'entre eux avec leur somme et, en dessous, le taux correspondant à chaque intervalle, tandis que la figure 3 montre comment ces



```

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: Les quatre valeurs de charnière de l'exemple, leur somme et le taux dans chaque tranche. Le coefficient -600 sur la charnière réfléchi $h(1200 - gla)$ est mis en évidence là où un lecteur pourrait le confondre avec le taux ; le taux en dessous de 1 200 pieds carrés est $+600$.

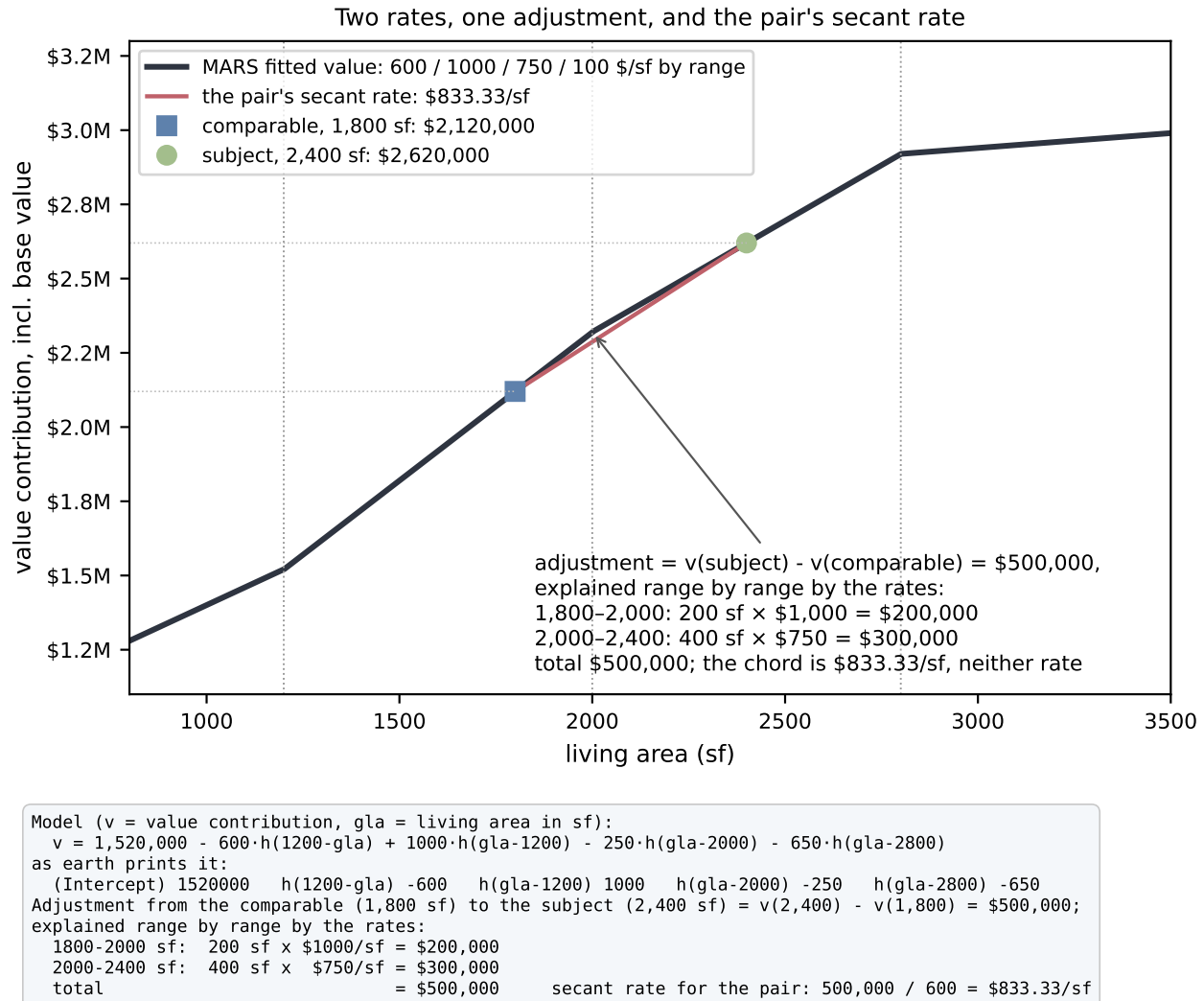


Figure 3: La valeur estimée, le bien comparable ayant une superficie de 1 800 pieds carrés et le bien à évaluer de 2 400. L'ajustement, c'est-à-dire la différence entre les deux contributions, est de \$500,000, et les taux en vigueur l'expliquent plage par plage; la corde reliant les deux points, d'environ \$833 par pied carré, est le taux sécant de la paire, c'est-à-dire le taux unique dont une ligne droite aurait besoin pour cette paire et aucune autre; une régression ordinaire doit utiliser un taux pour chaque paire.

taux expliquent l'ajustement entre un bien comparable et un bien à évaluer. Les nœuds situés à 1 200, 2 000 et 2 800 définissent quatre intervalles. Dans chaque plage, une charnière droite est *active* si son nœud se trouve sur la plage ou à gauche de celle-ci, et une charnière réfléchie si son nœud se trouve sur la plage ou à droite de celle-ci — une charnière droite agit dans toutes les plages situées au-delà de son nœud, une charnière réfléchie dans toutes les plages situées avant celui-ci. Le taux est calculé à partir des coefficients des charnières actives : on additionne les coefficients des charnières de droite actives et on soustrait les coefficients des charnières réfléchies actives. Le tableau qui en résulte correspond au taux du modèle pour chaque intervalle, les autres caractéristiques étant maintenues à un ensemble de valeurs (le programme utilise celles du sujet) ; dans un modèle sans interactions, cette précision ne change rien, tandis que dans le cas n° 2 ci-dessous, elle change tout :

surface (pieds carrés)	charnières actives	taux (\$/sf)
inférieur à 1 200	$h(1200 - gla)$ seulement	$-(-600) = +\mathbf{600}$
1 200 à 2 000	$h(gla - 1200)$	1000
2 000 à 2 800	$h(gla - 1200)$, $h(gla - 2000)$	$1000 - 250 = \mathbf{750}$
plus de 2 800	les trois charnières de droite	$1000 - 250 - 650 = \mathbf{100}$

La première ligne correspond à l'étape qui a été omise. Le coefficient d'−600, sur une charnière réfléchissante, ne correspond pas à un taux d'−\$600 par pied carré. L' $h(1200 - gla)$ de la charnière mesure l'écart entre la superficie d'une maison et 1 200 pieds carrés, et −600 multiplié par cet écart correspond à une décote : \$240,000 pour une maison de 800 pieds carrés, qui tend vers zéro à 1 200. Chaque pied carré supplémentaire réduit la décote de \$600, le taux est donc de +\$600. (Avec un coefficient positif, la même articulation apporte une valeur ajoutée en raison de sa petite taille, et son taux est négatif.) La règle, énoncée une fois pour toutes : dans un intervalle, le taux est égal à la somme des coefficients des charnières actives droites *moins* la somme des coefficients des charnières actives inversées, plus le coefficient d'un terme linéaire (une caractéristique saisie sous forme de ligne droite sans nœud), si le modèle en comporte un. Une convention s'y rattache : exactement au niveau d'un nœud, le taux est pris à partir de la droite, de sorte qu'une articulation droite compte à partir de son nœud et qu'une articulation inversée ne compte que strictement en dessous de son nœud ; le taux à 1 200 pieds carrés est donc \$1,000, soit le taux de l'intervalle qui commence à cet endroit, et le programme l'interprète de la même manière.

Pour vérifier qu'un tableau a été correctement construit, il faut examiner la valeur d'ajustement au niveau des nœuds. À 1 200 pieds carrés, chaque articulation est nulle et la valeur correspond à l'ordonnée à l'origine : \$1,520,000 ; à 2 000, elle est de $\$1,520,000 + 1,000 \times 800 = \$2,320,000$; à 2 800, de $\$2,320,000 + 750 \times 800 = \$2,920,000$; et à 800 pieds carrés, pour la plus petite maison du jeu de données, elle est de $\$1,520,000 - 600 \times 400 = \$1,280,000$. De 800 à 1 200, la valeur passe de \$240,000, soit un taux de \$600, ce qui correspond à la première ligne ; un tableau interprétant le coefficient comme le taux présenterait une valeur décroissante de gauche à droite jusqu'au nœud situé à 1 200, ce que les valeurs ne confirment pas. La figure2 représente les quatre nœuds, leur somme et le taux dans chaque plage, le coefficient −600 étant indiqué aux endroits où le lecteur pourrait le confondre avec le taux.

Le programme gère deux autres complications, et un tableau de conversion doit :

- **Interactions.** Earth, lors de son passage en avant (l'étape qui ajoute les termes un par un ; Friedman 1991, §3), affiche un produit tel que $h(gla - 1500) \times lot$ lorsque le taux d'une caractéristique dépend d'une autre : une *interaction*. Le taux correspondant à « lot » diffère alors selon que la surface habitable est supérieure ou inférieure à 1 500 pieds carrés, et le

tableau des taux d'une caractéristique est toujours celui qui *résulte* des autres, dont les valeurs sont fixées à celles du bien à évaluer. Le programme prend un point, une valeur pour chaque autre caractéristique, et calcule chaque taux à ce point.

- **Niveau par rapport à l'ajustement.** L'ordonnée à l'origine et les termes des autres caractéristiques définissent le niveau à partir duquel la contribution est mesurée ; ils s'annulent à chaque ajustement. Le programme présente les contributions sous la forme des termes de la caractéristique seule, de sorte que deux contributions diffèrent exactement du montant de l'ajustement.

Deux cas : modèles sans et avec interactions

La règle énoncée dans la section précédente peut s'énoncer sous la forme de deux théorèmes, dont la démonstration figure ci-dessous ; ceux-ci apportent une réponse claire à une question que tout lecteur d'un modèle MARS finit tôt ou tard par se poser : de combien de termes imprimés un taux dépend-il réellement ? La réponse se divise en deux cas.

Cas 1 : sans interactions

Contexte. Soit $h(u) = \max(0, u)$ pour un u réel. Un modèle MARS sans interactions sur les caractéristiques $x, z_1, \dots, z_q$ a donné une valeur d'ajustement

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

où β_0 est l'ordonnée à l'origine, chaque g_k est une fonction de z_k seule, et g regroupe tous les termes de x :

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

Ici, $t_1 < t_2 < \dots < t_K$ sont les nœuds distincts des charnières dans x ; a_j est le coefficient de la charnière droite à t_j et b_j celui de la charnière réfléchie à t_j , l'un et l'autre étant 0 lorsque le modèle ne comporte pas ce terme ; et c est le coefficient d'un terme linéaire nu $c \cdot x$, un élément saisi sous forme de droite sans nœud, que le programme « earth » ne génère que pour les éléments nommés dans `linpreds` ou, pour tout élément, lorsque `Auto.linpreds` est activé (voir la remarque après le Théorème 1) ; $c = 0$ lorsqu'il n'y a pas de tel terme, comme dans tous les modèles de ce supplément. Posons $t_0 = -\infty$ et $t_{K+1} = +\infty$ et appelons $S_i = (t_i, t_{i+1})$, pour $i = 0, 1, \dots, K$, les *segments* de x ; t_i est le nœud gauche de S_i et t_{i+1} son nœud droit. Sur chaque S_i , chaque terme de g est soit identiquement nul, soit linéaire, de sorte que g est linéaire sur S_i ; sa pente à cet endroit, qui est également $\partial f / \partial x$ sur S_i car aucun autre terme de f ne dépend de x , est le *taux* sur S_i et s'écrit r_i . Le tableau des taux de x est la liste $r_0, r_1, \dots, r_K$. (Si $K = 0$ il n'y a qu'un seul segment, $(-\infty, +\infty)$, et le tableau correspond au taux unique $r_0 = c$.) À un nœud lui-même, g n'a pas de dérivée : ses dérivées gauche et droite sont respectivement r_{i-1} et r_i , qui diffèrent de $a_i + b_i$ (Théorème 2 ci-dessous) ; ainsi, la valeur indiquée par un tableau de taux à t_i est une convention, prise ici à partir de la droite, et non une valeur de g' .

Théorème 1 (taux sur un segment). Avec la notation du cadre, pour tout $i \in \{0, 1, \dots, K\}$, on a

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

une somme vide étant définie par 0. En particulier, r_i dépend d'au plus K des coefficients de charnière $2K$, un par nœud, ainsi que de c .

Démonstration du théorème 1. Fixons i et un point $x \in S_i$, de sorte que $t_i < x < t_{i+1}$. Considérons chaque terme de la somme g dans un voisinage de x suffisamment petit pour se trouver à l'intérieur de S_i . Le terme nu $c \cdot x$ a pour dérivée c . Pour la charnière de droite en t_j : si $j \leq i$ alors $t_j \leq t_i < x$, donc $h(x - t_j) = x - t_j$ sur le voisinage et sa dérivée est 1 ; si $j \geq i + 1$ alors $x < t_{i+1} \leq t_j$, donc $h(x - t_j) = 0$ en ce point et sa dérivée est 0. Pour la charnière réfléchie en t_j : si $j \geq i + 1$ alors $x < t_j$, donc $h(t_j - x) = t_j - x$ et sa dérivée est -1 ; si $j \leq i$ alors $t_j < x$, donc $h(t_j - x) = 0$ et sa dérivée est 0. Une somme finie de fonctions dérivables en x est dérivable en ce point par la somme des dérivées, ce qui donne

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

la formule indiquée. Pour le comptage : les ensembles d'indices $\{j : j \leq i\}$ et $\{j : j \geq i + 1\}$ sont disjoints et épuisent ensemble $\{1, \dots, K\}$, de sorte que chaque nœud t_j contribue exactement à l'un des éléments a_j et b_j , mais jamais aux deux ; par conséquent, il y a au plus K coefficients de charnière, et r_i est identique en chaque point de S_i . ■

La limite est atteinte : si $b_j = 0$ pour tout j , le segment le plus à droite S_K utilise l'intégralité de K de la a_j .

Remarque (lorsque « c » n'est pas égal à zéro). earth peut représenter une caractéristique sous la forme d'une ligne droite unique sans nœud, le paramètre « $c \cdot x$ » du paramétrage, de deux manières différentes, et cette différence a son importance. « `linpreds` », une liste de caractéristiques nommées telles que « `linpreds = c("GLA")` », impose à chaque caractéristique de la liste d'être représentée de manière linéaire, même lorsqu'un ajustement avec des nœuds serait préférable ; c'est à l'analyste de décider, mais cela peut nuire à la qualité de l'ajustement. « `Auto.linpreds` », un simple commutateur activé par défaut, n'impose rien : il modifie uniquement la manière dont earth affiche une caractéristique dont l'ajustement s'est avéré véritablement linéaire, c'est-à-dire un seul point d'inflexion dont le nœud se situe à la plus petite valeur observée, en l'affichant sous la forme d'un coefficient et du nom de la variable plutôt que sous la forme d'un point d'inflexion ; l'ajustement reste identique. Dans les deux cas, earth affiche le terme sous la forme du nom de variable nu accompagné de son coefficient, `221.4 * GLA` dans le format de résumé d'earth, comme dans l'exemple ci-dessous. Theorem 1 gère cela via c , qui ajoute au taux sur chaque segment. Par exemple, l'appel ‘

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

‘ intègre chaque caractéristique de manière linéaire (avec ‘ `degree = 1`’, de sorte qu’aucun produit des termes linéaires ne soit ajusté non plus) et affiche, sous la forme de résumé d’earth (résultat à titre d’illustration, et non un modèle ajusté), ‘

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

‘ Pour la surface habitable, il s’agit du paramétrage avec ‘ $c = 221.4$ ’ et sans articulations, ‘ $K = 0$ ’ : il y a un seul segment, et le Théorème 1 donne un taux unique $r_0 = c = \$221.40$ par pied carré sur toute la plage des données ; la superficie du terrain, le nombre de salles de bains et de chambres s’interprètent de la même manière, à \$11.80 par pied carré de terrain, \$34,872 par salle de bains et \$17,641 par chambre. Un tel modèle est une régression ordinaire au sens du présent supplément, avec un taux par caractéristique. Sur l’ensemble des données, un tel terme correspond à une droite à une seule articulation vers la droite, dont le nœud se situe à la plus petite valeur observée. Le Journal utilise « earth » avec `Auto.linpreds = FALSE` ; ainsi, dans ses modèles, une caractéristique dont le graphique de la variable est rectiligne sur l’ensemble des données apparaît comme cette articulation unique, `h(gla-800) 700` par exemple, le nœud correspondant à la plus petite maison des données, avec $c = 0$, $K = 1$ et $a_1 = 700$: le taux est de 0 en dessous du nœud, sur un segment ne comportant aucune vente, et de 700 au-dessus de celui-ci, de sorte que le tableau des taux correspondant aux données ne comporte qu’une seule ligne. Il en découle un corollaire ; la variation du taux au niveau d’un nœud est le Théorème 2 ci-dessous.

Corollaire 1 (le signe réfléchi). Une charnière réfléchissante de coefficient b_j apporte $-b_j$ au taux sur chaque segment situé à gauche de t_j et 0 à droite. Dans l’exemple présenté, $b_1 = -600$ à $t_1 = 1,200$ apporte +600 à r_0 , ce qui constitue l’ensemble du piège autour duquel s’articule ce supplément.

Démonstration. Voir la deuxième somme du théorème 1. ■

Un exemple à sept nœuds, et l’hypothèse qui arrête les lecteurs

Ce décompte est important en raison d’une hypothèse que presque tout le monde fait lorsqu’on examine un modèle MARS imprimé : celle selon laquelle le coefficient imprimé à côté d’un nœud correspond à la pente du segment dont ce nœud constitue le nœud gauche. Ce n’est pas le cas. **Le coefficient d’une articulation correspond à un ajustement de la pente venant de la gauche, et non à la pente elle-même** ; la pente d’un segment est donc égale à la pente du premier segment, à laquelle s’ajoute la somme cumulée de tous ces ajustements au niveau de son nœud de gauche et de tous les nœuds situés plus à gauche. Un lecteur qui part de cette hypothèse erronée effectue le calcul mentalement, ou à l’aide d’un petit programme ou d’une formule de tableur, constate que les chiffres ne reproduisent pas les valeurs d’ajustement, et abandonne, concluant que MARS est opaque. Ce n’est pas opaque ; c’est cumulatif.

Théorème 2 (le cumul). En reprenant la notation du cadre du Cas 1, pour tout $i \in \{1, \dots, K\}$, on a

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

Par conséquent, on a $r_i = r_0 + \sum_{j=1}^i (a_j + b_j)$ pour tout i , où $r_0 = c - \sum_{j=1}^K b_j$ est le taux du segment le plus à gauche et $r_K = c + \sum_{j=1}^K a_j$ est le taux du segment le plus à droite.

Démonstration du théorème 2. Soit $i \in \{1, \dots, K\}$. D'après le théorème 1,

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

En soustrayant, l' c s'annule ; les deux sommes des a diffèrent d'un seul terme, a_i ; et comme $\sum_{j=i}^K b_j = b_i + \sum_{j=i+1}^K b_j$, les deux sommes des b diffèrent d'un seul terme, b_i , qui est soustrait dans r_{i-1} mais pas dans r_i . D'où $r_i - r_{i-1} = a_i + b_i$. En ajoutant cette identité pour $j = 1, \dots, i$ à la place de i , les membres de gauche se télescopent : $(r_1 - r_0) + (r_2 - r_1) + \dots + (r_i - r_{i-1}) = r_i - r_0$, ce qui donne la formule de somme. Les valeurs de r_0 et r_K sont le Théorème 1 à $i = 0$, où la somme des a est vide, et à $i = K$, où la somme des b est vide. ■

Corollaire 2 (règle de lecture). Dans le cadre du cas 1, pour tout $i \in \{1, \dots, K\}$, le débit sur le segment dont le nœud gauche est t_i est $r_i = r_{i-1} + a_i + b_i$: les coefficients indiqués à un nœud s'ajoutent au débit provenant de la gauche, et ne constituent pas eux-mêmes le débit du segment qui commence à cet endroit. Les deux coïncident, $r_i = a_i + b_i$, exactement lorsque $r_{i-1} = 0$; en particulier, pour un modèle dont les seules articulations sont des articulations de droite et qui ne comporte aucun terme linéaire, cela se produit à $i = 1$ et, d'après le Théorème 2, à aucun nœud ultérieur à moins que la somme des coefficients aux nœuds qui le précèdent ne soit égale à zéro.

Démonstration. La première égalité correspond au théorème 2. Il en découle que « $r_i = a_i + b_i$ » si et seulement si « $r_{i-1} = 0$ ». Si « $b_j = 0$ » pour tout « j » et « $c = 0$ », le théorème 2 donne « $r_0 = 0$ », donc « $r_1 = a_1$ », et « $r_{i-1} = \sum_{j<i} a_j$ », ce qui n'est «0» que lorsque la somme de ces coefficients est égale à zéro. ■

Corollaire 3 (l'ajustement en tant qu'intégrale). Dans le cadre du Cas 1, g est continue en chaque nœud, car chaque charnière l'est, tout comme $f(\cdot, z)$ pour chaque z ; c'est ce qui fait que la valeur en un nœud est la même, qu'on y parvienne par la gauche ou par la droite, et ce qui permet de passer d'une quantité à une autre. Soient x_c et x_s les quantités de x pour un comparable et un sujet, les autres caractéristiques étant maintenues aux mêmes valeurs pour les deux, de sorte que $g(x)$, les termes de f qui impliquent x , soit la contribution de x à la quantité x . Écrivons $u_1 = \min(x_c, x_s)$ et $u_2 = \max(x_c, x_s)$, ainsi que $|I|$ pour la longueur d'un intervalle I . Alors, l'ajustement $g(x_s) - g(x_c)$, c'est-à-dire la valeur ajoutée à la contribution du comparable pour la rendre égale à celle du sujet, est

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

En d'autres termes : l'ajustement est l'intégrale du taux entre les deux quantités, et comme le taux est constant sur chaque segment, l'intégrale est la somme, sur l'ensemble des segments rencontrés, du produit du taux par la largeur.

Démonstration. La fonction g est continue et par morceaux, avec un nombre fini de morceaux ; elle est donc absolument continue. D'après le théorème fondamental du calcul, on a $g(x_s) - g(x_c) = \int_{x_c}^{x_s} g'(u) du$, la dérivée existant partout sauf aux nœuds, qui forment un ensemble fini. Sur chaque S_i , l'intégrande est égale à r_i d'après le théorème 1 ; ainsi, $\int_{u_1}^{u_2} g'(u) du$ est la somme, sur i , de r_i multiplié par la longueur de $S_i \cap [u_1, u_2]$, les segments qui ne se rencontrent pas contribuant à la longueur 0 ; et $\int_{x_c}^{x_s} = \text{sign}(x_s - x_c) \int_{u_1}^{u_2}$ en raison de l'orientation de l'intégrale. ■

Le tableau des taux est g' ; l'ajustement, calculé directement comme $g(x_s) - g(x_c)$, est égal à son intégrale entre les deux quantités ; et les théorèmes 1 et 2 précisent de quoi se compose l' g' .

En pratique, l'ajustement est toujours calculé comme cette différence de contributions, ce qui ne nécessite aucun taux ; le corollaire explique pourquoi les taux, lus plage par plage, le reproduisent, ce qui en fait une explication de l'ajustement et un contrôle de celui-ci, et non un moyen de le calculer.

Prenons la surface habitable dans un ensemble de données dont la plus petite maison mesure 800 pieds carrés, avec sept nœuds, à 1 200, 1 600, 2 000, 2 400, 2 800, 3 200 et 3 600 pieds carrés, et neuf termes dans le modèle, dont deux correspondent à des charnières, chacune associée à une charnière droite au même nœud : à 1 200 et à 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

Sept nœuds forment huit segments. Le taux de chaque segment correspond à la somme des coefficients des charnières actives de droite moins la somme des coefficients des charnières actives de réflexion :

segment (sf)	termes actifs	taux (\$/sf)
800 à 1 200	$h(1200 - gla)$, $h(2400 - gla)$	$500 + 200 = 700$
1 200 à 1 600	$h(gla - 1200)$, $h(2400 - gla)$	$700 + 200 = 900$
1 600 à 2 000	charnières à droite à 1 200, 1 600 ; $h(2400 - gla)$	$700 - 100 + 200 = 800$
2 000 à 2 400	charnières à droite, de 1 200 à 2 000 ; $h(2400 - gla)$	$700 - 100 + 150 + 200 = 950$
2 400 à 2 800	charnières à droite entre 1 200 et 2 400	$700 - 100 + 150 - 250 = 500$
2 800 à 3 200	charnières à droite, de 1 200 à 2 800	$500 - 300 = 200$
3 200 à 3 600	charnières à droite, de 1 200 à 3 200	$200 - 150 = 50$
supérieur à 3 600	les sept charnières de droite	$50 + 50 = 100$

Trois éléments à relever dans le tableau. Le taux du segment le plus à droite, $700 - 100 + 150 - 250 - 300 - 150 + 50 = 100$, combine les coefficients des sept charnières de droite, à raison d'une par nœud, ce qui correspond à la limite du théorème 1 avec $K = 7$; les deux charnières réfléchies n'y jouent aucun rôle, car une charnière réfléchie est plate à droite de son propre nœud et le segment le plus à droite se trouve à droite de chaque nœud. Le segment le plus à gauche, de 800 jusqu'à son nœud de droite à 1 200, est l'image miroir : seules les articulations réfléchies comptent, avec leurs signes inversés, car chaque nœud se trouve au niveau ou à droite de son nœud de droite. Et à chaque nœud comportant une paire, le pas est la somme des deux coefficients : à 1 200, $-500 + 700 = +200$, de 700 à 900 ; à 2 400, le taux passe de 950 à 500, soit une variation de -450 , ce qui correspond exactement à la somme des deux coefficients à ce nœud, -250 et -200 : franchir le nœud active la charnière de droite et désactive la charnière réfléchie, et ces deux changements agissent dans le même sens.

On peut obtenir le même résultat sur le segment le plus à droite en suivant un parcours : on commence à 700 sur le premier segment, puis on passe de 800 à 1 200, et on ajoute, à chaque nœud

franchi, la somme des coefficients qui s’y trouvent : $+200, -100, +150, -450, -300, -150, +50$. Cet itinéraire passe par les huit segments et les neuf termes, et aboutit au même 100. La somme directe est plus courte et correspond à ce que fait le programme ; le parcours est ce qu’un lecteur devrait effectuer une fois, à la main, pour constater que le coefficient associé à un nœud correspond à un pas et non à un niveau. L’ -450 , à 2 400, correspond à deux demi-étapes : le franchissement du nœud laisse derrière lui la charnière réfléchie $h(2400 - gla)$, ainsi que l’ $+\$200$ qu’elle a ajoutée sous le nœud ; et il fait intervenir la charnière droite $h(gla - 2400)$, dont l’ -250 commence à cet endroit. $-200 - 250 = -450$, et le taux passe de $\$950$ à $\$500$.

La figure 4 illustre le même principe pour un modèle à cinq nœuds, suffisamment simple pour être vérifié à l’œil nu. Les barres représentent les coefficients imprimés aux nœuds ; l’escalier correspond au taux sur chaque segment ; et l’escalier commence au taux du premier segment, $\$900$, défini par la charnière réfléchissante, chaque marche correspondant à la somme des barres à ce nœud. Le taux du segment le plus à droite, $650 - 250 + 150 - 200 - 150 = 200$, utilise les cinq charnières de droite ; la charnière réfléchissante à 1 200, indiquée par -900 , définit le premier segment à $+900$ et n’intervient sur aucun autre élément.

Deux remarques s’imposent concernant la forme de ce graphique, car elles portent sur les maisons plutôt que sur MARS. Premièrement, l’axe commence à 800 pieds carrés et non à zéro, et le graphique part de l’axe : tous les graphiques d’un modèle d’ajustement présentés dans ce supplément, à l’instar des courbes tracées par le logiciel de modélisation, s’étendent de la plus petite maison des données à la plus grande et ne représentent rien en dehors de cette plage. Chaque zone de marché possède une maison la plus petite, en dessous de laquelle aucune vente n’est enregistrée, et le graphique variable ne montre rien en dessous de ce seuil : pas zéro, mais aucune valeur étayée par les ventes ; la formule renverrait tout de même un chiffre à cet endroit, et ce chiffre est le résultat d’une extrapolation. Les chiffres utilisent un seuil minimal de 800 pieds carrés, mais ce seuil est une caractéristique de la zone de marché, et non une constante : différentes zones de marché ont des tailles minimales différentes pour les biens immobiliers répertoriés dans le MLS, et c’est l’ensemble de données, et non l’expert immobilier, qui détermine où il se situe. Deuxièmement, la forme varie d’un marché à l’autre : dans l’exemple côtier, la fourchette la plus raide se situe entre 1 200 et 2 000 pieds carrés, et dans la figure 4, il s’agit du premier segment. Dans une zone où les prix sont élevés, comme Atherton ou Hillsborough, le segment le plus raide de la courbe de la surface habitable (GLA) peut se situer au-delà de 4 000 ou 5 000 pieds carrés, là où se trouvent les acheteurs. De même, un quartier dans lequel un promoteur a construit des maisons d’une seule taille présentant des problèmes à une certaine période pourrait afficher une forte baisse du taux sur une fourchette intermédiaire, comme le montre le graphique entre 1 600 et 2 000. Sur une zone de marché plus vaste, les courbes variables sont variées, et le tableau des taux permet à l’expert immobilier d’interpréter ce que chacune d’entre elles indique. Les figures 5 et 6 représentent deux de ces courbes telles qu’elles seraient imprimées, afin que le lecteur puisse établir chaque tableau de taux selon la règle et le vérifier par rapport à la courbe en escalier.

Cas 2 : interactions

Un terme d’interaction est un coefficient multiplié par le produit de facteurs portant sur différentes variables, chaque facteur étant soit une variable charnière, soit la variable elle-même, par exemple $\gamma \cdot h(gla - 1500) \cdot lot$. Tout ce qui figurait dans le cas n° 1 est conservé, à l’exception d’un mot : le taux d’un segment est désormais une fonction des autres variables, et non plus une constante. L’énoncé et sa démonstration suivent, puis l’impact de cette modification sur les calculs de l’évaluateur.

Paramétrage. Un modèle MARS avec interactions sur les caractéristiques $x, z_1, \dots, z_q$ a donné la

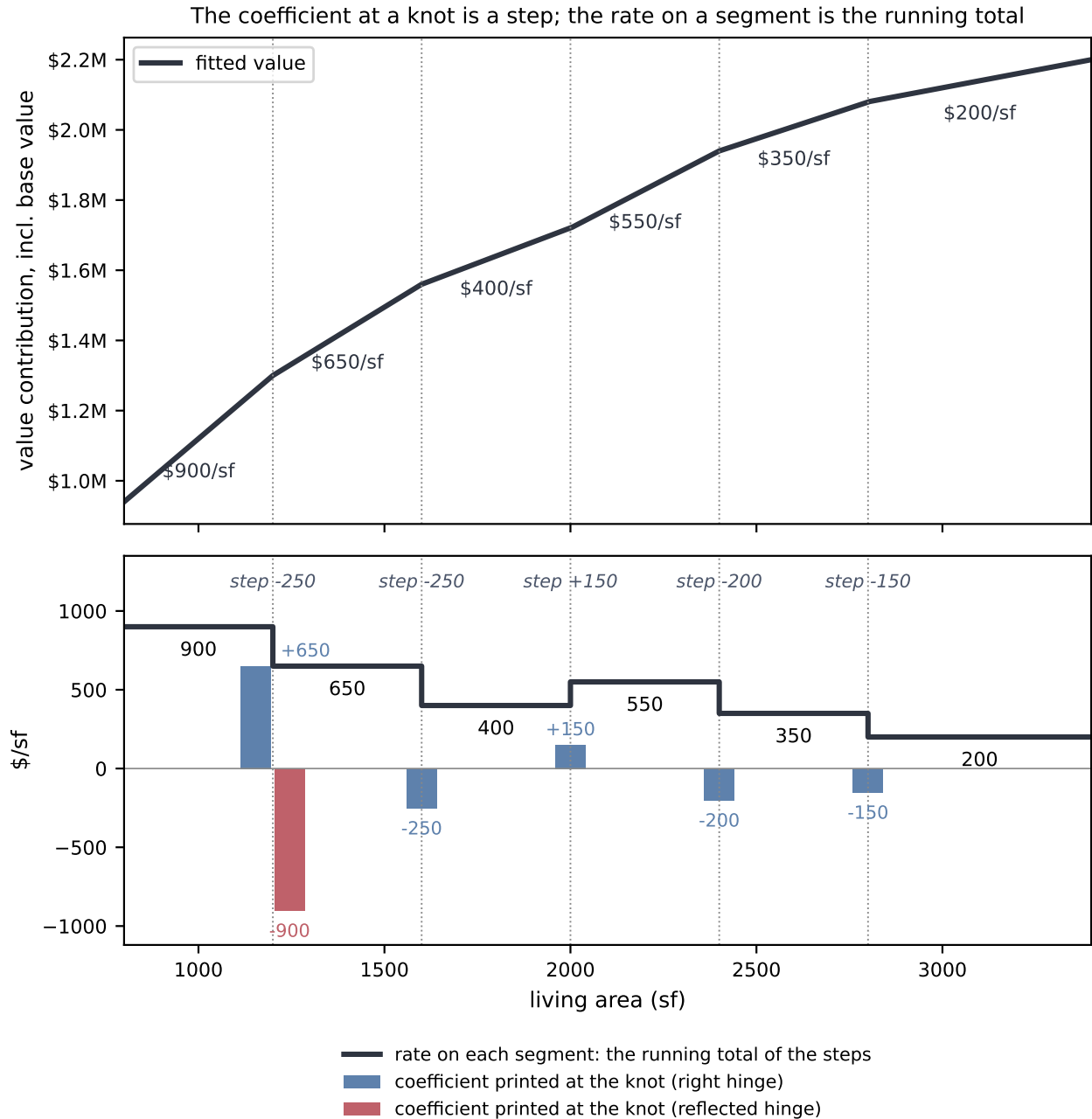


Figure 4: Un modèle à cinq nœuds basé sur des données commençant à 800 pieds carrés : $-900 h(1200 - \text{gla}) + 650 h(\text{gla} - 1200) - 250 h(\text{gla} - 1600) + 150 h(\text{gla} - 2000) - 200 h(\text{gla} - 2400) - 150 h(\text{gla} - 2800)$. Ci-dessus, la valeur d'ajustement avec le taux indiqué sur chaque segment ; il n'y a pas de valeur inférieure à 800 pieds carrés car il n'existe pas de maisons de cette superficie. Ci-dessous, le coefficient affiché à chaque nœud sous forme de barre et le taux sur chaque segment sous forme d'escalier : chaque marche de l'escalier correspond à la somme des barres à ce nœud (une barre ici, deux lorsqu'un nœud comporte une paire, Théorème 2), de sorte que le taux sur un segment correspond au taux du premier segment plus le cumul des marches situées à sa gauche, et non au coefficient qui se trouve à côté : chaque coefficient ajuste le taux provenant de la gauche.

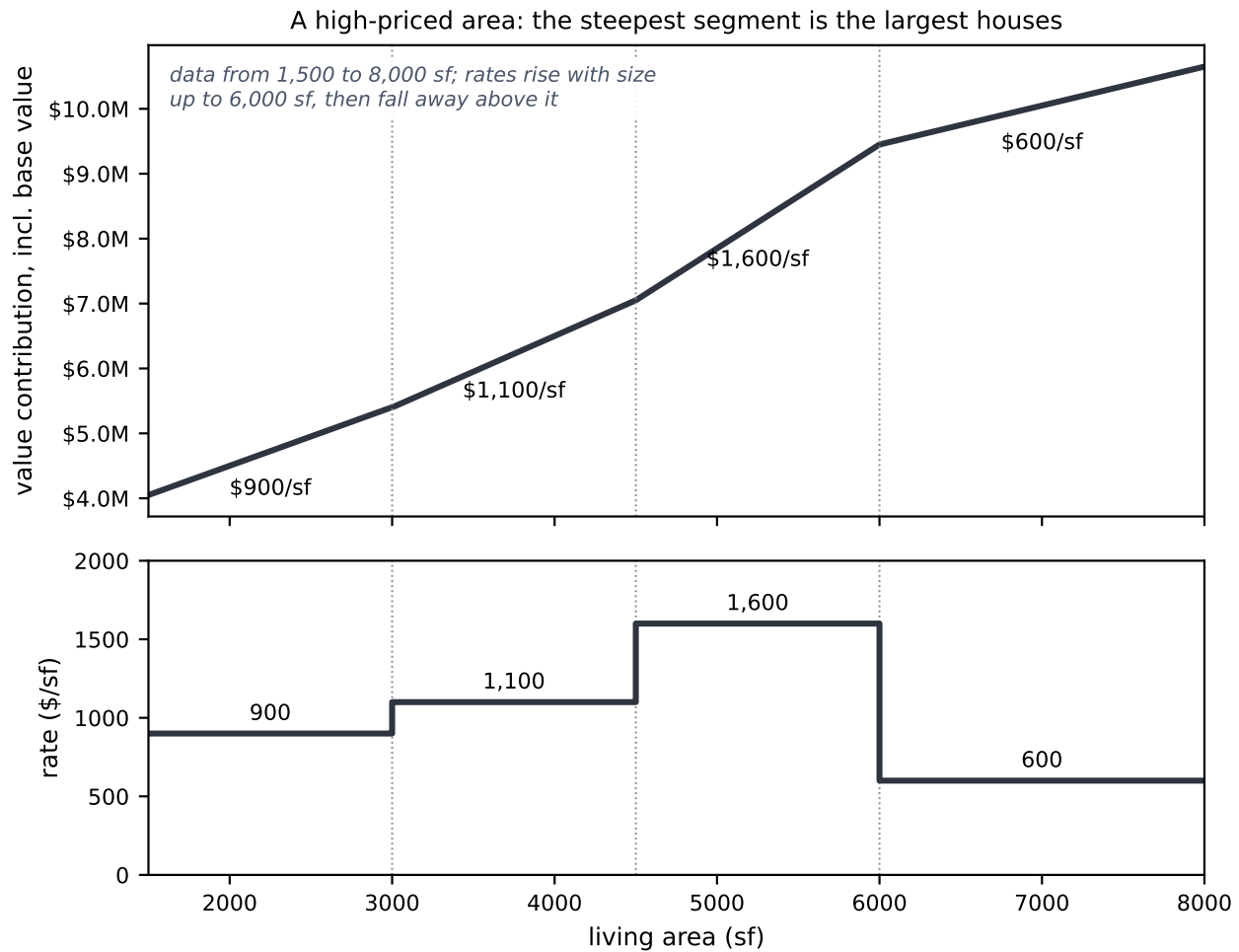


Figure 5: Un quartier où les prix sont élevés, pour des surfaces comprises entre 1 500 et 8 000 pieds carrés : $-900 h(3000 - gla) + 1100 h(gla - 3000) + 500 h(gla - 4500) - 1000 h(gla - 6000)$, croisez la route \$5,400,000 au premier nœud. Tarifs \$900 , \$1,100 , \$1,600 et \$600 par pied carré : la tranche la plus forte se situe entre 4 500 et 6 000 pieds carrés, là où se trouvent les acheteurs ; la première tranche est moins prononcée, et au-delà de 6 000 pieds carrés, le tarif diminue. La règle reste inchangée ; seul le marché évolue.

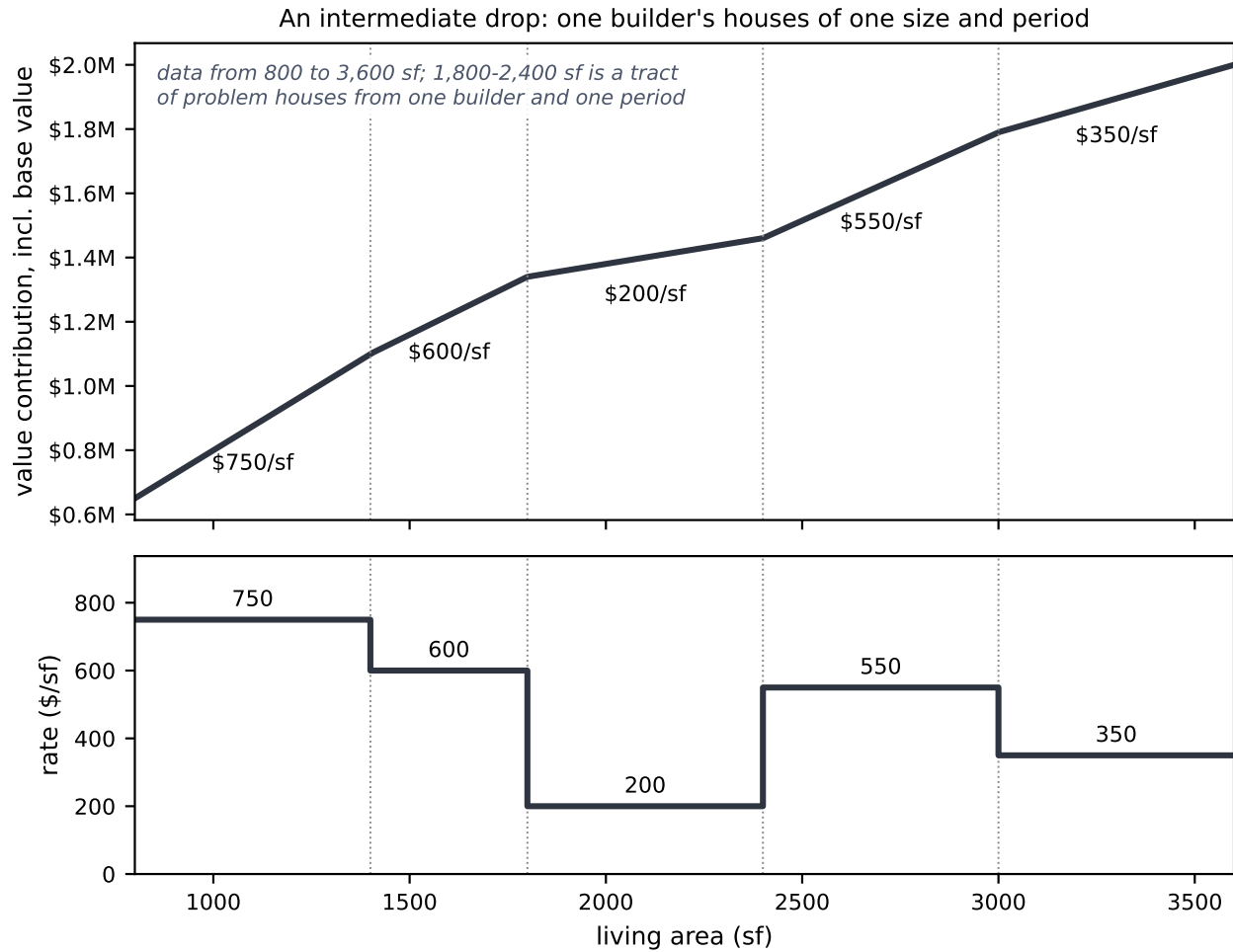


Figure 6: Une baisse intermédiaire, pour des surfaces comprises entre 800 et 3 600 pieds carrés : $-750 h(1400 - \text{gla}) + 600 h(\text{gla} - 1400) - 400 h(\text{gla} - 1800) + 350 h(\text{gla} - 2400) - 200 h(\text{gla} - 3000)$, croisez \$1,100,000 au premier nœud. Taux : \$750 , \$600 , \$200 , \$550 et \$350 par pied carré. Supposons que le segment allant de 1 800 à 2 400 corresponde à un ensemble de maisons de cette taille, construites par un même constructeur au cours d'une même période, avec des problèmes de prix du marché (une illustration fictive) ; le taux y chute à \$200 puis remonte au-delà. Un lecteur qui prend le coefficient -400 à 1 800 pour le taux de ce segment, ou qui suppose que les taux ne baissent qu'avec la taille, se trompe sur ce lotissement.

valeur d'ajustement suivante :

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

où M correspond au nombre de termes autres que l'intercepte, β_0 représente l'intercepte, et le modèle est constitué de l'intercepte plus la somme de ses termes, le terme m étant $\gamma_m \prod_{k \in F_m} B_{m,k}(z_k)$: son coefficient γ_m multiplié par un facteur pour chaque caractéristique du terme. Ici, $z = (z_1, \dots, z_q)$ et, afin qu'un indice parcoure chaque caractéristique, z_0 désigne x . Les caractéristiques apparaissant dans le terme m forment l'ensemble $F_m \subseteq \{0, 1, \dots, q\}$; aucune caractéristique n'apparaît deux fois dans un même terme, et chaque facteur $B_{m,k}$ est une fonction d'une seule caractéristique z_k , de l'un des trois types suivants : la caractéristique elle-même, z_k ; une charnière droite $h(z_k - t)$; ou une charnière réfléchie $h(t - z_k)$. Soient $M_x = \{m : 0 \in F_m\}$ les termes contenant x . Pour $m \in M_x$, $B_{m,0}(x)$ est le facteur du terme dans x , et les autres caractéristiques du terme, celles z_k avec $k \in F_m$, $k \neq 0$, sont les *partenaires* de x dans ce terme. Écrivez

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

pour le produit des facteurs partenaires, un produit vide étant 1. En d'autres termes : une fois les autres caractéristiques du bien à évaluer renseignées, les partenaires d'un terme se multiplient pour donner un nombre ordinaire ; ce nombre multiplié par le coefficient imprimé correspond au coefficient d'un terme de surface habitable simple, et le terme se lit comme dans le cas n° 1. Le produit des partenaires $P_m(z)$ est ce qui reste du terme m lorsque son facteur dans x est soustrait : le terme m est $\gamma_m B_{m,0}(x) P_m(z)$. À un point donné z , il s'agit d'un nombre ; ainsi, sur un segment, le terme agit sur x comme une articulation unique de coefficient $\gamma_m P_m(z)$, et ce produit correspond à la manière dont les partenaires entrent dans le taux. Un terme sans partenaire a $P_m(z) = 1$ et est un terme du Cas 1.

Par exemple, dans l'expression $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ de l'exemple ci-dessous, où x correspond à la surface habitable, le facteur dans x est $h(\text{gla} - 1200)$, le « one partner » correspond au terrain, et $P_m(z) = h(\text{lot} - 6000)$. Pour un terrain de 5 000 pieds carrés $P_m(z) = 0$, ce terme n'ajoute rien au taux ; pour un terrain de 10 000 pieds carrés $P_m(z) = 4,000$, ce terme agit comme un pivot à 1 200 avec un coefficient $0.03 \times 4,000 = 120$, ajoutant \$120 par pied carré au taux au-delà de 1 200 pieds carrés. Rien ici ne limite le nombre de caractéristiques qu'un terme peut contenir ; l'argument « earth's degree » fixe ce maximum, et les théorèmes ci-dessous s'appliquent pour n'importe quel degré. Un terme comportant d caractéristiques possède un graphe de $d+1$ dimensions, que personne ne peut dessiner dès lors que d dépasse deux ; le logiciel n'affiche que des tranches, une ou deux caractéristiques à la fois, les autres étant maintenues fixes, et le Théorème 3 correspond à ce même découpage effectué au niveau du point considéré. $P_m(z)$ est l'endroit où ce découpage a lieu : quel que soit le nombre de partenaires d'un terme, à un moment donné, leur produit est un nombre. Dans le terme à quatre caractéristiques $\gamma h(\text{gla} - 1200) h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$, avec $\gamma = 0.001$, dont le graphique comporte cinq dimensions, $P_m(z) = h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$; pour un terrain de 10 000 pieds carrés, un âge de 20 ans et 3 salles de bains, on a $4,000 \times 10 \times 3 = 120,000$, et le long de la surface habitable, le terme entier forme une articulation à 1 200 avec un coefficient $\gamma \times 120,000 = 0.001 \times 120,000 = 120$, ajoutant \$120 par pied carré : soit le même résultat que le terme à deux caractéristiques ci-dessus, mais avec un coefficient trente fois plus petit. C'est pourquoi le coefficient d'un terme de degré élevé apparaît souvent comme un très petit nombre lorsque ses partenaires sont mesurés en grandes unités : il multiplie le produit de plusieurs valeurs partenaires. Et si l'un des partenaires se situe du côté « zéro » de son nœud (une superficie bien

inférieure à 6 000 pieds carrés ou un âge inférieur à 10 ans), alors $P_m(z) = 0$ et le terme n'apporte aucune contribution, quelles que soient les valeurs des autres partenaires ; plus le degré est élevé, plus il existe de façons pour qu'un terme soit « désactivé » pour un bien donné. Les exemples présentés dans ce supplément utilisent des interactions bidirectionnelles (**degree = 2**).

Soient $t_1 < \dots < t_K$ les nœuds distincts des $B_{m,0}$ qui sont des charnières, avec $t_0 = -\infty$, $t_{K+1} = +\infty$ et les segments $S_i = (t_i, t_{i+1})$ comme dans le cas 1. Pour $m \in M_x$ et $i \in \{0, \dots, K\}$, définissons

$$s_{m,i} = \begin{cases} +1 & \text{si } B_{m,0} \text{ est la variable } x \text{ elle-même, ou une charnière droite } h(x - t_j) \text{ avec } t_j \leq t_i, \\ -1 & \text{si } B_{m,0} \text{ est une charnière réfléchie } h(t_j - x) \text{ avec } t_j \geq t_{i+1}, \\ 0 & \text{sinon.} \end{cases}$$

Un *point* est une valeur z des autres caractéristiques. Étant donné que chaque $B_{m,0}$ est linéaire ou identiquement nulle sur S_i , f est linéaire en x sur S_i pour un z fixe ; sa pente à cet endroit correspond au *taux* sur S_i au point z , noté $r_i(z) = \partial f / \partial x$ pour $x \in S_i$.

Pour visualiser tous les symboles d'un seul coup d'œil, reprenez le modèle de l'exemple ci-dessous tel qu'il apparaît à l'écran : l'ordonnée à l'origine est 900 000, $600 h(\text{gla} - 1200)$, $-350 h(\text{gla} - 2400)$, $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ et 15 lot. Ses caractéristiques sont la surface habitable, $x = z_0 = \text{gla}$, et la superficie du terrain, $z_1 = \text{lot}$, donc $q = 1$; l'ordonnée à l'origine est $\beta_0 = 900,000$ et il y a $M = 4$ termes. Prenons l'exemple d'un bien d'une surface habitable de 2 000 pieds carrés situé sur un terrain de 10 000 pieds carrés :

m	γ_m	F_m	$B_{m,0}(x)$	$B_{m,1}(z_1)$	$P_m(z)$	terme pour le bien
1	600	$\{0\}$	$h(\text{gla} - 1200)$	aucun	1	$600 \times 800 = 480,000$
2	-350	$\{0\}$	$h(\text{gla} - 2400)$	aucun	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	$\{1\}$	aucun	lot	(hors de M_x)	$15 \times 10,000 = 150,000$

Ainsi, $M_x = \{1, 2, 3\}$: le terme 4 ne comporte aucun facteur dans x , ajoute le même montant quelle que soit la surface habitable et n'entre pas dans le taux de x . La valeur d'ajustement est $f = 900,000 + 480,000 + 0 + 96,000 + 150,000 = \$1,626,000$. Les termes 1 et 2 n'ont pas de partenaire, donc $P_1(z) = P_2(z) = 1$; le terme 3 en a un, et $P_3(z) = h(10,000 - 6,000) = 4,000$, le multiplicateur de l'exemple ci-dessus. Le théorème 3, qui s'ensuit, les transforme en taux. La surface habitable comporte les nœuds $t_1 = 1,200$ et $t_2 = 2,400$, et la propriété se trouve sur le segment S_1 entre ces deux nœuds, où les termes 1 et 3 sont actifs et le terme 2 ne l'est pas ($s_{1,1} = s_{3,1} = 1$, $s_{2,1} = 0$) ; on a donc $r_1(z) = 600 \times 1 + 0.03 \times 4,000 = 720$, soit le taux correspondant à un terrain de 10 000 pieds carrés dans le tableau de l'exemple pratique.

Théorème 3 (taux sur un segment, avec interactions). Avec cette notation, pour tout i et tout point z , on a

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

Démonstration du théorème 3. Fixons i , un point z et $x \in S_i$. Un terme de la forme $m \notin M_x$ ne contient aucun facteur de x et est constant en x ; sa dérivée en x est 0. Un terme de la forme $m \in M_x$ est égal à $\gamma_m B_{m,0}(x) P_m(z)$, où $P_m(z)$ ne dépend pas de x . D'après l'analyse par cas de la démonstration du théorème 1, appliquée au facteur unique $B_{m,0}$, la dérivée de $B_{m,0}$ en x est 1 pour la caractéristique elle-même, 1 ou 0 pour une charnière droite selon que son nœud se trouve à t_i ou à gauche de celui-ci, ou à t_{i+1} ou à droite de celui-ci, et -1 ou 0 pour une charnière réfléchie selon

que son nœud se trouve à t_{i+1} ou à droite de celui-ci, ou à t_i ou à gauche de celui-ci ; c'est-à-dire qu'elle est égale à $s_{m,i}$. Ainsi, la dérivée du terme est $\gamma_m s_{m,i} P_m(z)$, et la dérivée de f , une somme finie, est la somme de ces deux termes. La valeur ne dépend pas de l' $x \in S_i$ choisie. ■

Le théorème 1 correspond au cas particulier où aucun terme de M_x n'a de partenaire, de sorte que chaque $P_m(z) = 1$; le regroupement des termes de la somme selon leurs nœuds donne la formule qui y est démontrée. Trois corollaires précisent comment les partenaires changent ; le dernier d'entre eux concerne l'échec de la borne « un par nœud » et le cas où « un terme peut s'annuler », comme le montre l'exemple ci-dessous.

Corollaire 4 (le taux n'est un nombre qu'en un point). Dans le cadre du Cas 2, pour chaque i , le taux r_i est une fonction du point z . Il est constant dans z dès lors que chaque terme $m \in M_x$ avec $s_{m,i} \neq 0$ n'a pas de partenaire. Lorsqu'un terme associé à $s_{m,i} \neq 0$ possède un partenaire, $r_i(z)$ est la fonction $z \mapsto \sum_m \gamma_m s_{m,i} P_m(z)$, qui varie en fonction des valeurs des partenaires, sauf lorsque les termes s'annulent exactement ; deux biens présentant le même x sur un même segment ont alors des taux différents pour x dès lors que les produits de leurs partenaires diffèrent et que ces différences ne s'annulent pas. C'est pourquoi le programme prend un point comme référence, et pourquoi il n'existe pas de tableau de taux pour une seule caractéristique prise isolément.

Démonstration. D'après le théorème 3, $r_i(z) = \sum_{m \in M_x} \gamma_m s_{m,i} P_m(z)$. Si chaque terme de la forme $s_{m,i} \neq 0$ n'a pas de partenaire, chaque « P_m » de ce type est le produit vide 1 et la somme est une constante. Sinon, la somme contient au moins un terme $\gamma_m s_{m,i} P_m(z)$ tel que $\gamma_m s_{m,i} \neq 0$ et P_m soit un produit d'un ou plusieurs facteurs dont chacun est une fonction non constante d'un partenaire ; la somme n'est donc constante en z que si ces termes s'annulent identiquement. ■

Corollaire 5 (sur un segment d' x , le taux est un graphe variable chez chaque partenaire). Dans le cadre du cas 2, fixons i , un partenaire z_p ($p \geq 1$) et les valeurs de toutes les autres caractéristiques. Alors, $z_p \mapsto r_i(z)$ est une fonction continue par morceaux linéaire de z_p dont les ruptures se situent parmi les nœuds des facteurs $B_{m,p}$ des termes de M_x . Le long de x , chaque partenaire étant fixé, le taux ne change qu'aux nœuds $t_1, \dots, t_K$ de x ; c'est dans l'espace conjoint de x et de ses partenaires que la surface des taux est coupée à la fois par les nœuds de x et par les nœuds de chaque partenaire.

Démonstration. Toutes autres caractéristiques étant fixes, chaque $P_m(z)$ est égale à une constante multipliée soit par 1 (si $p \notin F_m$), soit par le facteur unique $B_{m,p}(z_p)$, qui est lui-même z_p , $h(z_p - t)$ ou $h(t - z_p)$: une fonction continue par morceaux linéaire de z_p comportant au plus un point de rupture, situé à son nœud. Une combinaison linéaire finie de telles fonctions, ce que donne le Théorème 3 pour $r_i(z)$, est continue et par morceaux linéaire, avec des ruptures uniquement là où certains termes de la somme présentent une rupture. L'énoncé figurant sur x correspond au Théorème 3 lui-même : pour un z fixe, $r_i(z)$ est un nombre sur chaque S_i et les segments ne se rencontrent qu'aux nœuds de x . ■

Corollaire 6 (la borne du théorème 1 n'est pas respectée). Dans le cadre du cas 2, le nombre de termes comportant un sommand non nul dans la formule du théorème 3 pour $r_i(z)$ n'est limité que par $|M_x|$, le nombre de termes contenant x , et peut dépasser K : un nœud de x peut apparaître dans plusieurs termes de M_x avec différents partenaires, chacun étant un terme distinct. À l'inverse, un terme comportant $s_{m,i} \neq 0$ ne contribue en rien à un point où le produit de son partenaire est nul.

Démonstration. Les termes $\gamma_1 h(x - 1500) \text{lot}$ et $\gamma_2 h(x - 1500) h(\text{age} - 10)$ sont des termes distincts de M_x ayant le même nœud, et sur le segment situé au-dessus de 1,500, ils ont tous deux $s_{m,i} = 1$; ils contribuent donc tous deux, et $r_i(z) = \gamma_1 \text{lot} + \gamma_2 h(\text{age} - 10)$ comporte deux termes

issus d'un même nœud ; l'ajout de nouveaux partenaires ajoute d'autres termes au même nœud sans limite. Pour le contraire, prenons un terme dont le partenaire est une charnière $h(\text{lot} - 6000)$ en un point où $\text{lot} < 6000$: alors $P_m(z) = 0$ et le terme s'annule bien que $s_{m,i} \neq 0$; en un point où $\text{lot} > 6000$, ce n'est pas le cas. ■

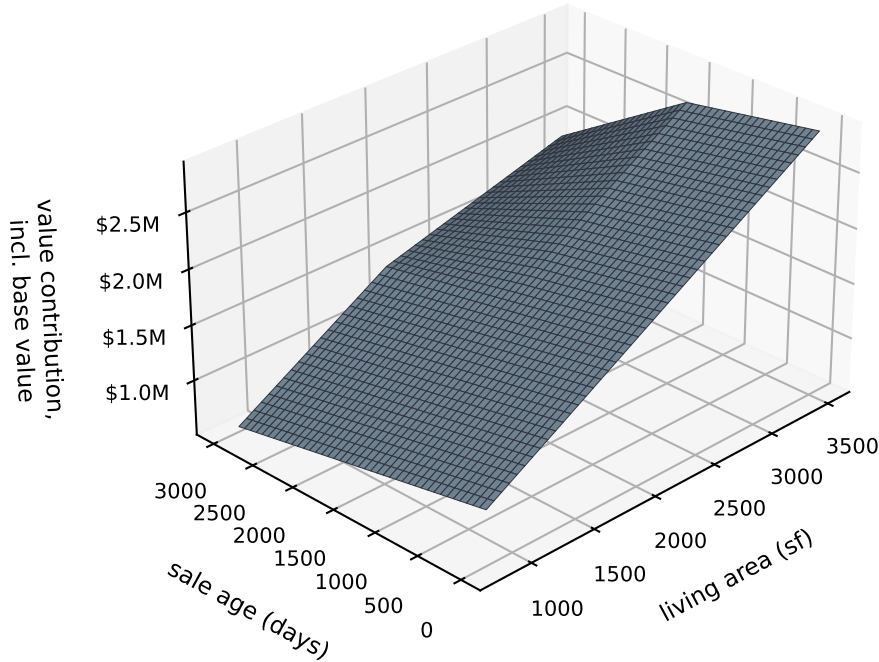
Il convient ici de mentionner une observation pratique tirée des propres évaluations de l'éditeur. Dès lors que les interactions sont autorisées, le graphique de la surface habitable que le logiciel trace pour la variable considérée se présente souvent comme une ligne droite, avec un taux unique, tandis que les courbes qu'un modèle sans interactions aurait introduites dans la surface habitable apparaissent à la place dans les termes d'interaction, généralement la surface habitable en fonction de l'âge de la vente (le délai entre la vente et la date d'entrée en vigueur) ou de la surface habitable en fonction du nombre de salles de bains. Le théorème 3 explique pourquoi la ligne droite ne correspond pas au taux. Le graphique par variable dans earthUI représente les termes sans partenaire ; le graphique de degré un de plotmo, quant à lui, représente une tranche du modèle complet avec les autres variables à leurs médianes. Les termes d'interaction sont représentés sous forme de surfaces ailleurs (les graphiques de degré un et de degré deux de plotmo, Milborrow 2024b, que earthUI affiche côte à côte). Pour tout bien immobilier concret, le taux applicable à la surface habitable correspond à la pente de la droite, à laquelle s'ajoute, pour chaque terme d'interaction actif, son coefficient multiplié par la valeur de son facteur partenaire ; il varie aux nœuds de surface habitable de ces termes, et l'ampleur de cette variation dépend des partenaires : une vente il y a six ans et une vente le mois dernier, ou une maison à deux salles de bains et une à quatre salles de bains, se situent sur des barèmes de surface habitable différents même si le graphique de surface habitable affiche une seule ligne. Un lecteur qui prend cette ligne pour le taux se heurte à nouveau à l'obstacle, à un niveau supérieur ; le tableau des taux au point considéré par le sujet est le moyen de le franchir. La figure7 représente un tel modèle à la manière du logiciel, sous forme de surface en fonction de la surface habitable et de l'ancienneté de la vente, puis en lit le graphique de la variable « surface habitable » à deux anciennetés de vente.

La figure 8 reprend le même principe dans le sens inverse, en partant du marché illustré par la figure 2. Un modèle est construit pour ce marché à partir d'une interaction entre la surface habitable et l'ancienneté de la vente, de sorte que l'effet principal de la surface habitable soit représenté par une ligne droite, \$500 par pied carré, et que les quatre termes d'interaction déterminent les courbes. Le graphique que le logiciel trace de lui-même pour la surface habitable correspond à cette ligne droite. Si l'on considère une vente datant de quatre ans, les taux liés à la surface habitable sont ceux de la figure 2 par dollar, soit \$600, \$1,000, \$750 et \$100 ; pour une vente récente, on obtient la ligne droite ; pour une vente datant de deux ans, les valeurs se situent entre les deux. Le modèle est construit pour illustrer ce point, et non pour s'ajuster aux ventes ; un ajustement réel ne correspondrait pas aussi parfaitement à la figure2 ; mais le mécanisme est celui utilisé par un ajustement réel, et c'est pourquoi un expert qui voit le graphique de la surface habitable se redresser après l'ajout d'interactions n'a pas perdu les courbes. Les courbes se sont déplacées vers les termes d'interaction, où seule une lecture à un point précis permet de les récupérer. Le graphique des effets principaux d'un modèle comportant des interactions ne représente que les termes de la variable elle-même ; il s'agit d'une coupe correspondant à une configuration des partenaires où chaque terme d'interaction est nul (ici, une nouvelle vente), et non de la forme du marché ; il convient d'interpréter le graphique de la variable en question.

La dernière conséquence concerne davantage l'ajustement que le taux, et c'est celle-ci qui se répercute sur la grille des prix de vente.

Théorème 4 (la répartition d'un ajustement dépend de l'ordre). Dans le cadre du cas 2, soit $y = z_p$ ($p \geq 1$) l'une des autres caractéristiques, w représentant les autres, et z_k représentant

Living area by sale age: the surface the software draws



The living-area variable graph, read at two points

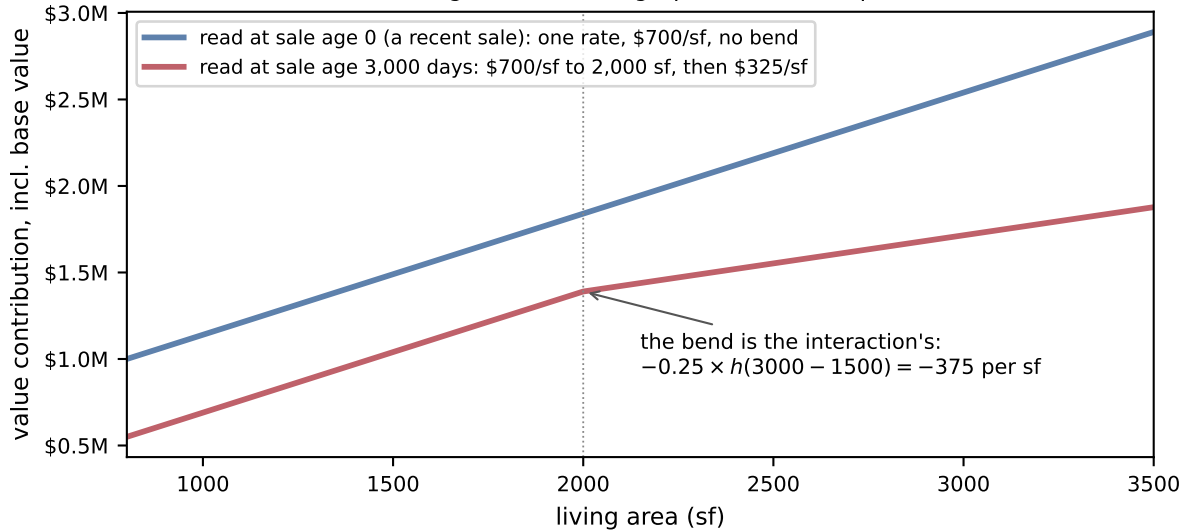


Figure 7: Une interaction entre la surface habitable et l'ancienneté de la mise en vente : $1,000,000 + 700 h(\text{gla} - 800) - 150 \text{ saleage} - 0.25 h(\text{gla} - 2000) h(\text{saleage} - 1500)$, sur des données allant de 800 à 3 500 pieds carrés et des ventes datant de 3 000 jours maximum. Ci-dessus, la surface tracée par le logiciel de modélisation pour ce couple. Ci-dessous, le graphique de la variable « surface habitable » lu en deux points, comme l'exige le théorème 3 : pour toute vente datant de moins de 1 500 jours, le facteur partenaire $h(\text{saleage} - 1500)$ est nul, l'interaction n'apporte rien et le graphique est une ligne droite à \$700 par pied carré ; pour une vente datant de 3 000 jours, le facteur « partenaire » est égal à 1 500, l'interaction soustrait $0.25 \times 1,500 = 375$ par pied carré au-delà de 2 000 pieds carrés, et le taux à ce niveau est de \$325 . L'effet principal de la surface habitable (les termes concernant uniquement la surface habitable, sans partenaire) correspond à la même droite dans les deux cas ; la courbe est due à l'interaction et n'apparaît que lorsque le partenaire est pris en compte.

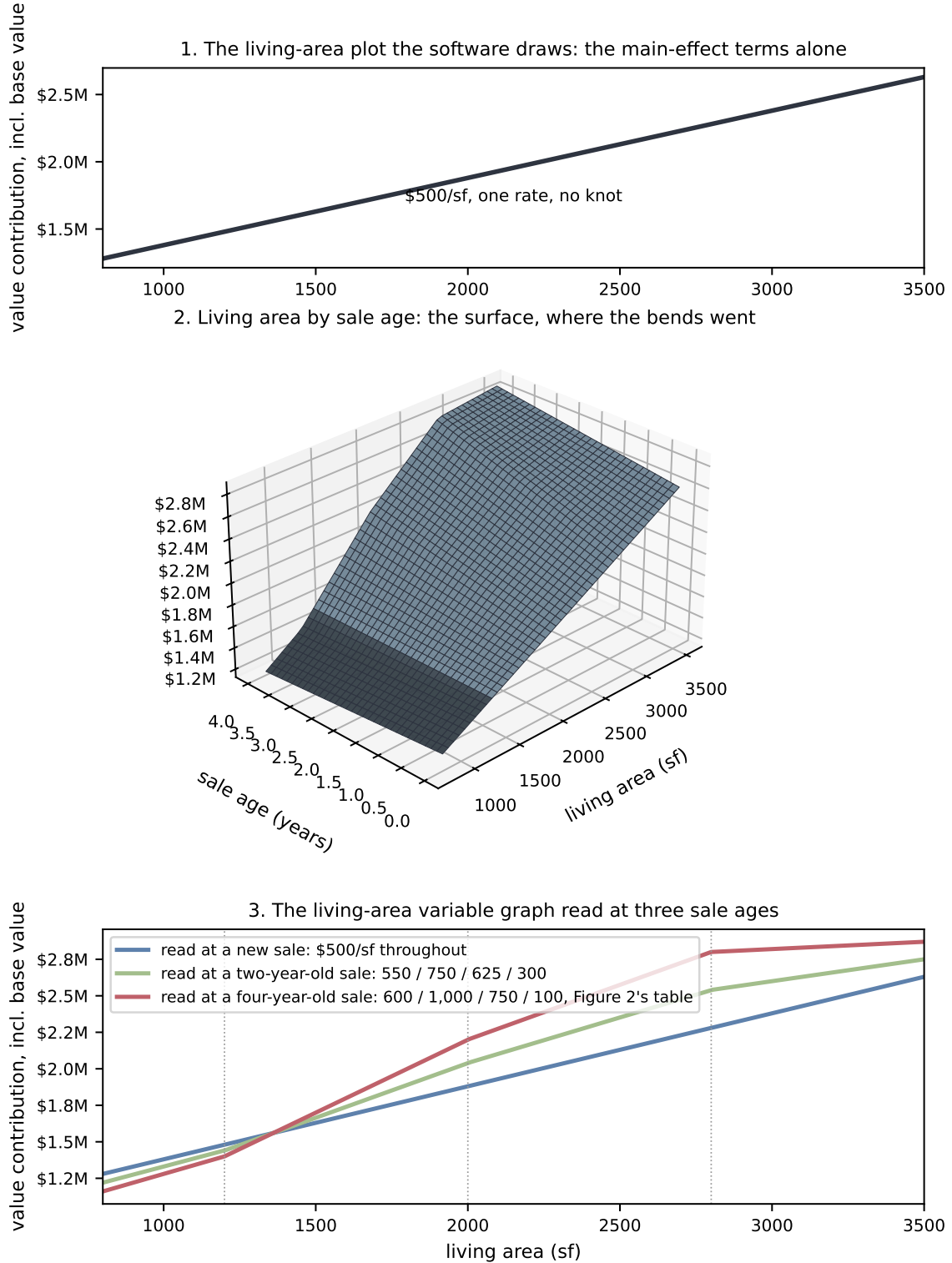


Figure 8: Le marché illustré dans la figure 2, avec une interaction entre la surface habitable et l'ancienneté de vente : $1,280,000 + 500 h(\text{gla} - 800) - 20,000 \text{ saleyrs} + [-25 h(1200 - \text{gla}) + 125 h(\text{gla} - 1200) - 62.5 h(\text{gla} - 2000) - 162.5 h(\text{gla} - 2800)] \times \text{saleyrs}$, l'ancienneté de vente étant exprimée en années, de 0 à 4. Les graphiques sont décrits dans leurs titres ; à quatre ans, le graphique du bas présente exactement les mêmes taux que ceux de la figure 2, \$120,000, mais à un niveau inférieur. Créé à des fins de démonstration, sans ajustement.

$k \notin \{0, p\}$; notons $f(x, y, w)$ pour f avec ses arguments ainsi regroupés. Soit un comparateur à l'adresse (x_c, y_c, w) et un sujet à l'adresse (x_s, y_s, w) , le même w pour les deux. Pour une valeur y^* de y , définissons l'ajustement pour x avec y maintenu à y^* comme suit :

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

Soient $M_{x,y}$ les termes contenant à la fois x et y . Chacun de ces termes est égal à son coefficient multiplié par trois parties, $\gamma_m A_m(x) C_m(y) Q_m(w)$: $A_m(x) = B_{m,0}(x)$ est son facteur dans x , qui est soit x lui-même, soit une articulation dans x ; $C_m(y) = B_{m,p}(y)$ est son facteur dans y , qui est soit y lui-même, soit une articulation dans y ; et $Q_m(w)$ est le produit de ses facteurs dans les caractéristiques restantes, chacune de ces caractéristiques étant soit la caractéristique elle-même, soit une articulation en son sein, et vaut 1 lorsque le terme ne comporte aucun autre facteur. Dans l'exemple ci-dessous, ce terme est $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$, avec $A_m(x) = h(\text{gla} - 1200)$, $C_m(y) = h(\text{lot} - 6000)$ et $Q_m(w) = 1$. Écrivez $\Delta_x(y_s)$ et $\Delta_x(y_c)$ pour l'ajustement de x , en conservant y à la valeur du bien en question y_s et à celle du bien comparable y_c , c'est-à-dire $\Delta_x(y^*)$ à $y^* = y_s$ et à $y^* = y_c$; dans l'exemple ci-dessous, avec x pour la surface habitable et y pour le terrain, $\Delta_x(y_c) = \$240,000$ correspond au terrain de 5 000 pieds carrés du bien comparable et $\Delta_x(y_s) = \$288,000$ à celui de 10 000 pieds carrés du bien en question. Ensuite

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

alors que la différence totale $f(x_s, y_s, w) - f(x_c, y_c, w)$ n'implique aucun choix d' y^* . L'énoncé porte sur cette tranche à deux caractéristiques, un partenaire y , toutes les autres caractéristiques étant maintenues identiques w pour les deux biens ; lorsque trois caractéristiques ou plus diffèrent entre les deux, la répartition du total entre elles dépend du chemin suivi, de la même manière que le long d'une chaîne plus longue de changements portant sur une seule caractéristique, et le théorème s'applique à chaque étape de la chaîne, les caractéristiques non encore déplacées étant maintenues à leurs valeurs actuelles. En particulier, avec un seul terme de ce type $\gamma A(x) C(y)$ et aucun facteur dans w , les deux ajustements pour x diffèrent par $\gamma [A(x_s) - A(x_c)] [C(y_s) - C(y_c)]$, ce qui n'est pas 0 chaque fois que le mouvement dans x modifie A et que les deux propriétés diffèrent en C . Pour $A = h(x-t)$ et $C = y$, cela donne $\gamma [h(x_s-t) - h(x_c-t)] (y_s - y_c)$; pour $A = h(x-t)$ et $C = h(y-s)$, la forme de l'exemple ci-dessous, cela donne $\gamma [h(x_s-t) - h(x_c-t)] [h(y_s-s) - h(y_c-s)]$.

Démonstration du théorème 4. Regroupez les termes de f selon qu'ils contiennent x ou y . Un terme ne contenant ni x ni y prend la même valeur à (x_s, y^*, w) et (x_c, y^*, w) et s'annule dans $\Delta_x(y^*)$. Un terme contenant y mais pas x s'annule de la même manière. Un terme contenant x mais pas y contribue à $\Delta_x(y^*)$ d'un montant qui ne dépend pas de y^* , il s'annule donc dans $\Delta_x(y_s) - \Delta_x(y_c)$. Un terme $m \in M_{x,y}$ apporte $\gamma_m [A_m(x_s) - A_m(x_c)] C_m(y^*) Q_m(w)$ à $\Delta_x(y^*)$; en posant $y^* = y_s$ et $y^* = y_c$ puis en soustrayant, on obtient la somme indiquée. La différence totale correspond à la différence entre deux valeurs d'une seule fonction f et ne fait aucune mention de y^* . ■

Ainsi, la valeur de l'interaction appartient à la paire de caractéristiques, et non à l'une ou l'autre en particulier. Une grille qui ajuste une caractéristique à la fois en fonction des valeurs des biens comparables, et une autre qui l'ajuste en fonction des valeurs du bien à évaluer, s'accordent sur le total mais divergent sur la répartition. Maintenir les autres caractéristiques aux valeurs du bien à évaluer est une convention défendable, et c'est celle qu'utilise le programme ; un rapport doit préciser quelle convention a été utilisée. Une grille résidentielle standard, telle que celle de l'Uniform Residential Appraisal Report (URAR), ne comporte pas de ligne pour une interaction. Effectuées

caractéristique par caractéristique, les deux opérations doivent placer les éléments comparables aux extrémités opposées — la surface habitable ajustée en fonction de la superficie du terrain du bien à évaluer, puis la superficie du terrain ajustée en fonction de la surface habitable du bien comparable, ou l'inverse — afin que l'interaction ne soit comptée qu'une seule fois et que les lignes s'additionnent pour donner la différence totale du modèle. Lorsque les contributions sont calculées directement, une grille peut également inclure l'interaction sous la forme d'une ligne distincte, représentant la différence des termes d'interaction entre le bien à évaluer et le bien comparable, à côté des lignes correspondant à chaque caractéristique ; chaque ligne constitue alors un groupe de termes, et il n'est pas nécessaire de choisir un ordre particulier. Dans l'exemple ci-dessous, les trois lignes sont : surface habitable \$240,000, terrain \$75,000, et surface habitable par terrain \$120,000, dont la somme donne le même résultat \$435,000. En cas d'interactions, le taux sécant du couple varie en fonction de celui-ci de deux manières : d'une part en fonction des plages traversées, et d'autre part en fonction de la position du couple au sein des partenaires.

Un exemple traité avec une interaction

Encore un marché côtier, avec cette fois deux caractéristiques : la surface habitable et la superficie du terrain, ainsi qu'une interaction entre ces deux éléments. 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

Comme l'indique le théorème 3, la surface habitable comporte deux nœuds, 1 200 et 2 400, et trois termes dans $l'M_{\text{gla}}$: les deux articulations simples et l'interaction, dont le partenaire est $h(\text{lot} - 6000)$. Sur le segment allant de 1 200 à 2 400, la charnière simple à 1 200 est active, tout comme la charnière de l'interaction à 1 200, le taux y est donc de $600 + 0.03 h(\text{lot} - 6000)$: un graphique variable en fonction de la superficie du terrain, constant à 600 pour les terrains jusqu'à 6 000 pieds carrés et augmentant de 3 cents par pied carré supplémentaire. Au-delà de 2 400, le seuil situé à 2 400 entre en jeu et le taux est de $250 + 0.03 h(\text{lot} - 6000)$. Le taux propre au terrain est de $15 + 0.03 h(\text{gla} - 1200)$ au-delà de 6 000 pieds carrés et de 15 en dessous ; ainsi, le taux de chaque élément dépend de la quantité de l'autre.

Deux points, deux tableaux. Le barème du programme pour la surface habitable, compris entre 800 et 3 500 pieds carrés, avec d'une part un terrain de 5 000 pieds carrés et d'autre part un terrain de 10 000 pieds carrés. Chaque colonne correspond à la pente du modèle en fonction de la surface habitable, le terrain étant fixé à cette seule valeur, et à aucune autre :

surface habitable (pieds carrés)	prix au lot de 5 000 (\$/sf)	prix au lot de 10 000 (\$/sf)
800 à 1 200	0	0
1 200 à 2 400	600	720
2 400 à 3 500	250	370

Pour le plus petit lot, le facteur de partenariat $h(5000 - 6000)$ est égal à zéro ; l'interaction n'apporte donc rien et le tableau correspond à celui que donnerait un modèle sans interaction ; c'est le cas « où le terme peut s'annuler » du Corollaire 6. Pour la parcelle la plus grande, le partenaire

est 4,000 ; l'interaction ajoute $0.03 \times 4,000 = 120$ à chaque segment supérieur à 1 200, et le même nœud, 1 200, alimente deux termes actifs, ce qui constitue un échec de la borne « un par nœud ».

Passons maintenant à l'ajustement. Le bien de référence présente une superficie de 1 800 pieds carrés sur un terrain de 5 000 pieds carrés, avec une valeur estimée à \$1,335,000 ; le bien à évaluer dispose d'une surface habitable de 2 200 pieds carrés sur un terrain de 10 000 pieds carrés, dont la valeur estimée est de \$1,770,000. Le montant total à expliquer est de \$435,000. Ajustez d'abord la surface habitable en fonction de la superficie du terrain du bien comparable, puis celle du terrain en fonction de la surface habitable du bien à évaluer. Chaque ajustement correspond à la différence entre les deux valeurs comparées ; lorsque le changement se situe au sein d'un même segment, cette différence est égale à la largeur multipliée par le taux, comme l'indique le tableau :

	d'abord la surface habitable, ensuite le terrain	le terrain d'abord, la surface habitable
ajustement de la surface habitable	$400 \times 600 = 240,000$	400 000
ajustement de parcelle	195,000	195 000
total	435,000	435 000

Les ajustements liés à la superficie du terrain correspondent également à des différences de contributions, et les barèmes permettent de les expliquer ; le barème propre au lot change à partir de 6 000 pieds carrés : pour les 2 200 pieds carrés du bien à évaluer, le taux du lot est de \$15 en dessous du nœud et de \$45 au-dessus, donc le passage de 5 000 à 10 000 correspond à $1,000 \times 15 + 4,000 \times 45 = 195,000$, et non à $5,000 \times 45$; pour les biens comparables de 1 800 pieds carrés, on a $1,000 \times 15 + 4,000 \times 33 = 147,000$. Les totaux concordent, comme le stipule le théorème n° 4, et les répartitions diffèrent de \$48,000, ce qui correspond exactement à la formule du théorème : $0.03 \times [h(2200 - 1200) - h(1800 - 1200)] \times [h(10000 - 6000) - h(5000 - 6000)] = 0.03 \times 400 \times 4,000$. Ce \$48,000 correspond à la valeur combinée des 400 pieds carrés supplémentaires de surface habitable *et* des 4 000 pieds carrés supplémentaires de terrain ; il n'appartient à aucune de ces deux caractéristiques, et c'est celle qui est ajustée en second qui en bénéficie. Selon la convention du programme, les autres caractéristiques étant maintenues aux valeurs du bien à évaluer, la colonne de droite indique pour la surface habitable \$288,000 à \$720 par pied carré, et son rapport précise à quel moment le taux a été relevé :

```
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.
```

Ces deux valeurs sont calculées sur la base d'une superficie de 10 000 pieds carrés pour le bien à évaluer ; sur sa propre superficie de 5 000 pieds carrés, la valeur de la surface habitable du bien comparable s'élève à \$360 000.

Un évaluateur qui relance cette opération sur le terrain des biens comparables obtient \$240,000 et n'a pas tort ; l'évaluateur et le programme analysent la même surface en deux points différents, et le rapport doit préciser de quels points il s'agit.

En résumé : sans interactions, un modèle MARS correspond à un tableau de taux par caractéristique, et chaque argument concernant un taux formulé dans le cadre d'une régression classique s'applique segment par segment. Avec des interactions, un taux forme une surface ; le théorème 3 précise exactement comment celle-ci se courbe, et l'expert doit indiquer à quel point elle a été relevée.

Le programme

La fonction « `mars_rates.pl` », reproduite dans son intégralité dans l'annexe, lit le tableau « earth » (imprimé) ou un modèle écrit sous forme de terme Prolog, et génère le tableau des taux, les valeurs aux nœuds, les deux contributions et leur différence, l'ajustement, accompagné d'une explication tranche par tranche, le taux sécant de la paire, ainsi qu'une vérification que la somme des tranches correspond à l'ajustement. Compte tenu de la plage de données, c'est-à-dire la plus petite et la plus grande maison parmi les ventes, il ne signale aucun taux en dehors de celle-ci : l'algèbre d'une charnière tend vers l'infini d'un côté, ce qui n'est pas le cas des ventes, et un taux au-delà de celles-ci relève de l'extrapolation. À titre d'exemple, pour des données allant de 800 à 3 500 pieds carrés :

```
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): 833.33333333333334
Check: segments sum to the adjustment.
```

Deux types de chiffres figurent dans ce bloc. Les contributions correspondent uniquement aux termes relatifs à la surface habitable, l'ordonnée à l'origine et tous les autres termes étant omis : \$600,000 et \$1,100,000 diffèrent exactement du terme d'ajustement. Les valeurs aux nœuds sont des valeurs d'ajustement, avec l'intercepte, ainsi que les termes des partenaires à ce point dans un modèle comportant des interactions ; ici, \$1,520,000 pour 1 200 pieds carrés correspond

à l'intercepte, car tous les termes d'interaction y sont nuls. Les valeurs ajustées au niveau du cas comparable et du cas étudié, \$2,120,000 et \$2,620,000, comportent l'interception de la même manière. Le rapport désigne les valeurs aux nœuds comme des valeurs ajustées afin que le lecteur ne les prenne pas pour des contributions.

C'est ce bloc-là qui doit figurer dans l'annexe d'un rapport : les taux, les nœuds et les calculs qui les relient, sous une forme que le lecteur puisse refaire à la main. Les 69 tests du programme vérifient la règle des signes, la règle d'interaction, la cohérence des segments avec les ajustements, ainsi que l'analyseur syntaxique des résultats de « earth », et sont validés sous Trealla (le Prolog du Journal), sous SWI-Prolog et sous Stryer Prolog. Pour les lecteurs qui travaillent sous R, la même lecture est disponible sans quitter cet environnement : `mars_rates.R`, publié avec le programme, prend un objet 'earth' ou son tableau imprimé et affiche le même bloc, avec les taux limités à la plage de données ; le Prolog reste la référence vérifiée, et le fichier R effectue ses propres auto-vérifications par rapport aux exemples de ce supplément. Le programme, ses tests et le script de la figure sont publiés avec ce numéro sous forme de fichier téléchargeable en complément du PDF, sous licence GNU Affero General Public License, version 3 ou ultérieure.

Pourquoi MARS, et ce que ce supplément laisse à d'autres entrées

La revue utilise le modèle MARS pour trois raisons, et ce sont ces raisons qui permettent l'existence de ce supplément. Objectivité : les nœuds sont placés là où les ventes les positionnent, selon une procédure définie et une règle d'élagage précisée (l'étape ultérieure qui supprime les termes qui ne méritent pas leur place), et rien concernant la forme d'un graphe variable n'est décidé à l'avance au-delà de ces paramètres définis (le degré maximal d'interaction, la pénalité d'élagage, l'espacement minimal entre les nœuds). Détectabilité : le passage avant explore chaque caractéristique et chaque nœud candidat, y compris les interactions jusqu'au degré défini, et rend compte de ce qu'il a trouvé, ce qu'aucune méthode partant d'une forme choisie ne peut faire. Interprétabilité : des taux constants sur les plages choisies par les données, et des interactions construites à partir des mêmes articulations, permettent d'écrire le modèle sous forme de tableau de taux et d'expliquer un ajustement comme le produit du taux par la largeur. Deux autres méthodes produisent également des taux qui varient sur la plage d'une caractéristique, et elles sont différentes l'une de l'autre. La première, la régression segmentée (« broken-stick »), prend une variable et un nombre de ruptures choisies par l'analyste, puis estime l'emplacement de ces ruptures, avec une erreur type (mesure de la variation possible de l'estimation en fonction d'autres ventes) pour chacune d'entre elles ; entre les ruptures, le taux est constant, comme dans le modèle MARS. La seconde, un modèle additif à contrainte de forme, effectue l'ajustement d'une courbe lisse au lieu de segments rectilignes, contrainte à une forme imposée par l'analyste — par exemple, que la valeur ne diminue jamais à mesure que la surface habitable augmente, ou que chaque pied carré supplémentaire apporte une augmentation moindre que le précédent. Aucune de ces deux méthodes n'est utilisée ici. Chacune répond à une question que ce supplément ne pose pas, et chacune renonce à au moins l'une des trois propriétés : la régression segmentée renonce à la découvrabilité, car la variable et le nombre de seuils sont choisis à l'avance ; le modèle à contrainte de forme sacrifie l'interprétabilité, car une courbe ne présente pas de taux constants à tabuler, et l'objectivité, car sa forme est imposée. Ce supplément analyse le modèle produit par MARS ; le choix de la méthode, ainsi que la stabilité des nœuds et des taux que MARS détermine d'un ensemble de ventes à l'autre, sont des bien à évaluer.

Ce que cela ne règle pas

Les taux sont ceux du modèle, et les quatre principes statistiques s’y appliquent tels quels : un taux ne se rapproche de l’espérance conditionnelle que dans la mesure où la forme du modèle le permet (*espérance conditionnelle*) ; il peut refléter la valeur d’une caractéristique omise (*biais de variable omise*) ; elle peut être faussée lorsqu’une caractéristique évolue avec l’erreur structurelle (*Endogénéité*) ; et elle peut ne pas être déterminée du tout par les ventes, quel que soit leur nombre, lorsque deux caractéristiques évoluent toujours ensemble (*Identification des paramètres*). MARS détecte un changement de taux uniquement lorsque les ventes présentent le contraste nécessaire pour le mettre en évidence ; un nœud est une affirmation concernant le marché et sa validité dépend de la qualité des ventes de part et d’autre de celui-ci. Ce que ce supplément apporte est plus précis : étant donné le modèle, voici comment interpréter ce qu’il indique.

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

$\hookrightarrow$ 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).

```