

MARS’ın insan tarafından anlaşılmasının önündeki engel (The Barrier to Human Understanding of MARS)

Bir MARS modelini birim oranlar, değer katkıları ve düzeltmeler olarak okumak

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

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

Bu, Temeller bölümüne yönelik bir ek olup, bir kavram maddesi değildir. Başlığı, editörün yıllar süren dolaylı gözlemleri sonucunda farkına vardığı bir engeli ifade etmektedir: Regresyon katsayısını sorunsuzca okuyan değerlendirme uzmanları, bir MARS programının yazdirdığı “menteşe terimleri” listesinde dururlar; çünkü o listede oran gibi görünen hiçbir şey yoktur. Bu ekin ardından gelen dört istatistiksel madde (*Koşullu Beklenti*, *Atlanan Değişken Önyargısı*, *Endojenlik* ve *Parametre Tanımlama*), regresyonun bir özelliğe atadığı oran hakkında tartışmaktadır. Derginin kendi modellemesinde çok değişkenli uyarlanabilir regresyon spline’ları (MARS; Friedman 1991) ¹ kullanılmaktadır; bu yöntem, özellik başına tek bir oran değil, aralık başına bir oran raporlar ve bir MARS programının yazdirdığı sonuçlardan değerlendirme uzmanlarının savunabileceği oran tablosuna uzanan yolda, elle oluşturulan tabloların hataya yol açtığı bir adım bulunmaktadır. Bu ek, söz konusu süreci, hataya yol açan adımı ve bu süreci izleyerek kendi çalışmasını kontrol eden bir programı sunmaktadır.

Biçimsel tanım. Bir MARS modelinde, tahmin edilen değer bir kesim noktası artı terimlerin toplamından oluşur; her terim, bir katsayının özelliklerin menteşe fonksiyonlarının çarpımıyla çarpılmasıyla elde edilir. Tahmin edilen değer, modelin fiyat tahmini niteliğindedir. Diğer özellikler sabit tutulduğunda, tek bir özellik için bu değer bir dizi düz segmentten oluşur. Segmentlerin kesiştiği noktalar düğümlerdir. Özelliğin aralığı boyunca değer katkısının grafiği, *değişken* grafiğidir. Her aralıktaki eğimi, orandır: o aralıktaki özelliğin birim başına uyumda olan değerdeki değişim; bir özellik miktarının *değer katkısı*, o miktarda o özellikteki model terimlerinin toplamıdır; örnekte tüm bu terimler 1.200 fit kare değerinde sıfırdır, dolayısıyla katkı buradan itibaren ölçülür ve bu değer altında negatiftir; bir modelin böyle bir miktara sahip olması gerekmez ve yalnızca katkıları arasındaki farklar, yani düzeltmeler, anlam taşır; ve bir emsalden bir konuya yapılan *düzeltilme*, emsalin katkısını konunun katkısına eşit hale getirmek için eklenen miktardır; bu, iki miktar arasındaki aralıklarda oran çarpı genişliğin toplamıdır. Bu şekilde tanımlanan oran modele aittir; bu oran, model satışların adil bir açıklaması olarak kabul edildikten sonra piyasanın ödediği oran olarak okunur ve bu eklemenin ardından gelen kayıtlar, bu kabulün ne zaman haklı olduğuyla ilgilidir.

Bu ek neden var

earthUI’de (Craytor 2026) veya doğrudan R paketi earth’te (Milborrow 2024a) bir MARS modeline uyumlu bir değerlendirme uzmanına, aşağıdaki gibi bir terim listesi sunulur:

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.

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

Bunu kimse raporlamaz. Satış tablosuna işlenen şey, düzeltmedir: söz konusu mülkün değer katkısından emsal mülkün değer katkısının çıkarılması. Metrekare başına fiyatlar, MARS modelini gerekçelendirme sürecinin bir adımı olarak raporda ele alınır. Yazılı şartlardan bu rakamlara giden yol kısadır, ancak elle oluşturulan tabloların hataya yol açabileceği bir adım vardır ve modelin ürettiği rakamlar bir denetçiye açıklanamıyorsa, model kullanmanın tüm anlamı kaybolur.

Basit anlatım: dört aralık, dört birim oran

Yaşam alanını tek bir doğrusal terim olarak içeren bir doğrusal regresyon, tüm piyasa genelinde bu alan için tek bir oran verir; geleneksel değerlendirme yöntemlerindeki regresyonlar işte bunu yapar ve bu ekte “sıradan regresyon” terimi de bunu ifade eder. MARS, satış verilerinin bir değişiklik gösterdiği durumlarda fiyatın değişmesine izin verir. Örnekteki kıyı bölgesindeki emlak piyasasında, 1.200 fit kareye kadar olan yaşam alanlarının fiyatı fit kare başına \$600, 1.200 ile 2.000 fit kare arasında olanlar için fit kare başına \$1,000, 2.000 ile 2.800 fit kare arasında olanlar için \$750 ve bunun üzerindeki alanlar için ise sadece \$100 olabilir. İşte aradaki tek fark budur ve bu, bir emlakçının hemen kavrayacağı bir şeydir: Piyasa, küçük bir evi bir aile evine dönüştüren alana en yüksek bedeli öder; bir sonraki odaya daha az, mahalledeki alıcıların istediğinin ötesindeki alana ise neredeyse hiçbir şey ödemez.

Oranlar göz önüne alındığında, geri kalan her şey bir değerlendirme uzmanının zaten yaptığı aritmetik hesaplamalardan ibarettir. 1.800 fit kare alana sahip bir emsal ile 2.400 fit kare alana sahip bir değerlendirme nesnesi arasında 600 fit kare fark vardır ve 2.000 fit kare düğümü bu ikisinin arasında yer alır. \$1,000 adresinde 200 fit kareyi, ardından \$750 adresinde 400 fit kareyi aşıyor \$2,620,000 – \$2,120,000 = \$500,000; bu da toplamda \$500,000 tutarındadır \$1,100,000 – \$600,000. Şekil 3 bunu göstermektedir: 800 fit kare büyüklüğündeki \$1,280,000 adresindeki (verilerdeki en küçük ev) evden, 3.500 fit kare büyüklüğündeki \$3 million adresindeki (en büyük ev) eve kadar uzanan bir aralıkta, modelin bu özelliklere sahip bir ev için verdiği fiyat (uyum değeri) 2.800 fit kare sınırından önce en büyük artış göstermektedir.

İki noktadan geçen düz bir çizginin bu hareket için kullanabileceği tek oran, bu iki noktayı birbirine bağlayan orandır: 600 fit kare için \$500,000, yani fit kare başına yaklaşık \$833: bu ikilinin *sekant oranı*, o tek hareket için geçerli veya harmanlanmış oran: modeldeki oranların, geçilen her aralığın genişliği ile ağırlıklandırılmış ortalaması. Bu, modelin oranlarından biri değildir ve bir sonraki çift için farklı bir ortalama olacaktır. Yaşam alanı için tek bir oran içeren sıradan bir regresyon, her çiftte o tek sayıyı uygulamak zorundadır; bu nedenle bu çift bazlı ortalamaları yeniden üretmez; tek bir oranın yaptığı da budur: modelin piyasanın gerçekleştirdiğini söylediği bir değişiklik üzerinden ortalama alır.

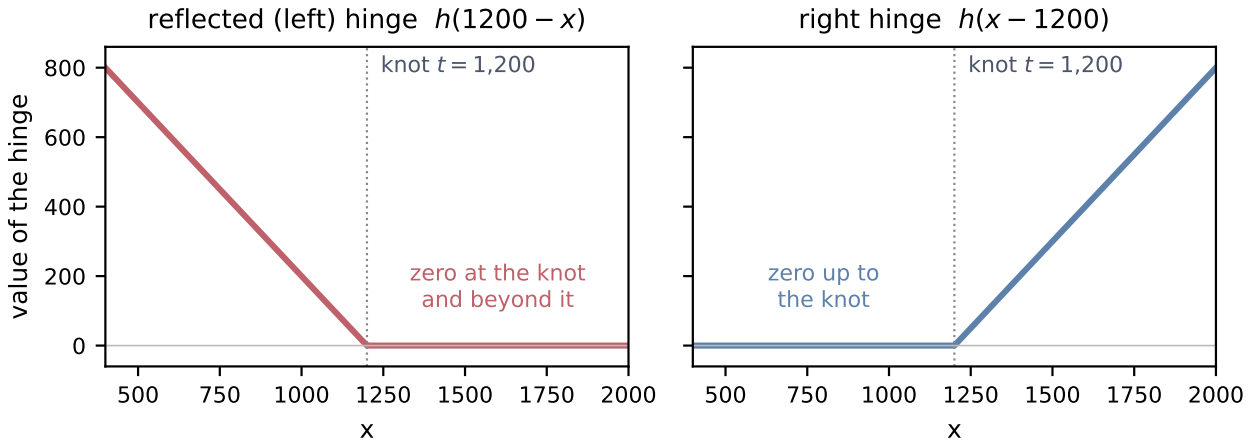
Burada duran bir okuyucu yöntemi kavramıştır: iki katkının farkını alın, ardından her aralığın oranını okuyarak, bu oranlarda konu ile emsal örneği adım adım izleyerek ve toplama işlemini gerçekleştirerek bunu açıklayın. Aşağıdaki açıklamalar, oranların neden bu şekilde belirlendiğini, tek tuzağın nerede yattığını, özellikler birbiriyle etkileşime girdiğinde nelerin değiştiğini ve bir programın hesaplamaları nasıl kontrol ettiğini öğrenmek isteyen okuyucular içindir; oran tablosu ve düzeltmeler değişmez.

Ayrıntılı anlatım: menteşe fonksiyonları ve gözden kaçan işaret

Bir değişkenin değerinin, değişkenlik aralığı boyunca yaptığı katkıyı gösteren grafiğe *değişken grafiği* denir. Bu, günlük hayattaki anlamıyla düzgün bir şekilde kıvrılan bir eğri değildir; düğümler arasında düz olan ve yalnızca bu noktalarda kıvrılan, bölümlere ayrılmış doğrusal bir grafikdir. MARS bunu menteşelerden oluşturur; Friedman (1991) bunları kesik üs fonksiyonları olarak yazmıştır ve daha sonraki literatürde “menteşe” adı benimsenmiştir (Milborrow, tarihsiz); Hastie, Tibshirani ve Friedman (2009, §9.4), bir düğümdeki bu ikiliye “yansıtılmış çift” adını vermektedir.

Tanım (menteşe). Menteşe, tek bir kesinti noktasına sahip tek değişkenli bir fonksiyondur x ; bu kesinti noktasına “düğüm” t denir ve fonksiyonun değeri, düğümün bir tarafında sıfır, diğer tarafında ise düğümden uzaklığa eşittir. Her düğümde, her bir taraf için birer tane olmak üzere iki adet menteşe bulunur; bunlar Şekil 1’de gösterilmiştir:

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



Şekil 1: Düğümlü $t = 1,200$ noktasında olan iki menteşe fonksiyonu: yansıtılmış (sol) menteşe $h(1200 - x)$; bu fonksiyon soldan sağa doğru azalır, düğüm noktasında sıfıra ulaşır ve düğümün ötesinde de sıfır kalır; sağ menteşe $h(x - 1200)$; bu fonksiyon düğüm noktasına kadar sıfırdır ve düğümün ötesinde soldan sağa doğru artar. Bunlar temel menteşelerdir; bir modelde her biri, yönü tersine çevirebilen bir katsayı ile çarpılır.

İlki, $h(x - t)$, *sağ* menteşedir: düğüme kadar sıfırdır ve düğümün ötesinde soldan sağa doğru artar. İkincisi, $h(t - x)$, *yansıtılmış* (veya sol) menteşedir: düğüme kadar soldan sağa doğru azalır ve düğümden ve düğümün ötesinde sıfırdır. Her ikisi de aynı anlamda menteşelerdir; aralarındaki tek fark, düğümün hangi tarafında sıfırdan farklı olduklarıdır. Bu yönler, Şekil 1’deki x ekseninden çizilen yalın menteşeleri tanımlar ve bundan öte bir anlamı yoktur. Bir modelde bir menteşe, negatif de olabilen bir katsayı ile çarpılır ve diğer terimlere eklenir; bu nedenle, sağ menteşe üzerine kurulu bir terim soldan sağa doğru azalabilir ve yansıtılmış menteşe üzerine kurulu bir terim artabilir; değişken grafiği ise diğer terimlerin belirlediği herhangi bir yükseklikte bir düğümlerle kesişir. Bu ekte belirtilen her yön soldan sağa doğru okunur. Mafsal terimleri tek başına kullanılmaz. Bir MARS modelinde her bir mafsal bir katsayı ile çarpılır ve terimler toplanır; modelin eğimleri, yani oranlar, tek bir mafsaldan değil, bu katsayılardan hesaplanır: birikimli olarak, aralık aralığa, her aralığın

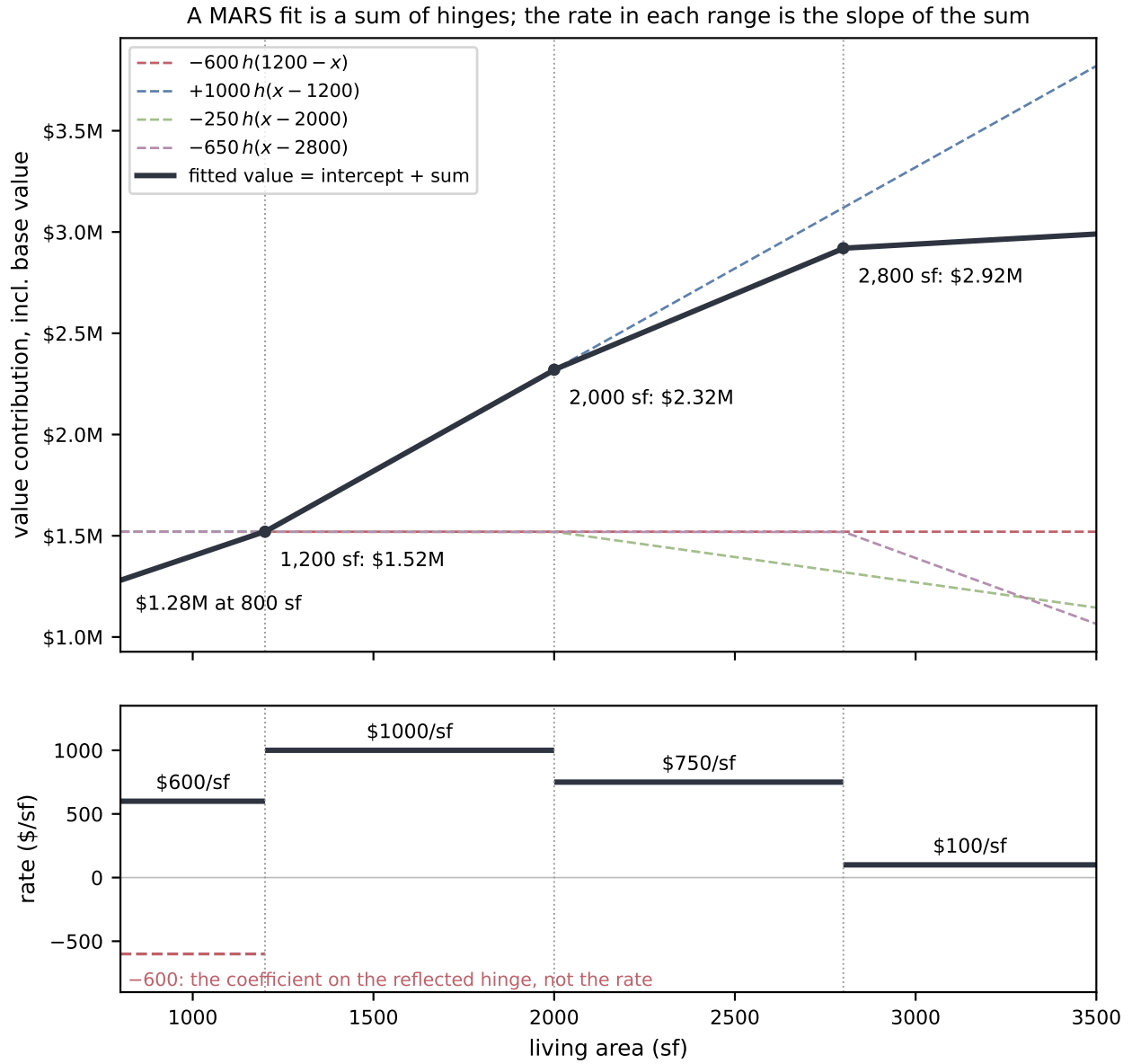
oranı, aralarında düğüm bulunan menteşelerin katsayılarıyla düzeltilmiş bir önceki aralığın oranıdır. Mafsal, MARS'ın tek bir değişken için sahip olduğu tek yapı taşıdır; bir değişken grafiğindeki her kıvrım, bir mafsalın düğümüdür ve her düz bölüm, o noktada ya sıfır ya da doğrusal olan terimlerin toplamıdır. **Aşağıdaki tüm açıklamalarda akılda tutulması gereken nokta şudur: bir mafsalın katsayısı, eğim değildir. Bu, eğime yapılan bir düzeltmedir; o menteşenin düğüm noktasında, soldan gelen eğime eklenen miktardır.** Soldan sağa doğru okunduğunda, bir değişken grafiği ilk segmentinin oranından başlar ve her düğüm, o noktada bulunan katsayılarla bu oranı ayarlar. Okuma kuralı budur ve aşağıdaki Teorem 1 ve 2 bunun kanıtıdır. Model, kesim noktası artı her katsayının kendi menteşesiyle çarpımıdır ve herhangi bir aralıktaki oran, bu toplamın eğimidir.

Yukarıdaki basılı terimleri ele alalım; Şekil 2 her birini toplamlarıyla birlikte ve altında her aralığa ait oranı gösterirken, Şekil 3 ise bu oranların emsal bir öge ile incelenen öge arasındaki düzeltmeyi nasıl açıkladığını göstermektedir. 1.200, 2.000 ve 2.800'deki düğümler dört aralık oluşturur. Her aralıkta, düğümü aralığın üzerinde veya solunda yer alıyorsa bir sağ menteşe, düğümü aralığın üzerinde veya sağında yer alıyorsa bir yansıtılmış menteşe etkindir — bir sağ menteşe, düğümünün ötesindeki her aralıkta, bir yansıtılmış menteşe ise düğümünden önceki her aralıkta etki eder. Oran, aktif menteşelerin katsayılarından hesaplanır: aktif sağ menteşelerin katsayılarını toplayın ve aktif yansıtılmış menteşelerin katsayılarını çıkarın. Ortaya çıkan tablo, diğer özellikler belirli bir değer kümesinde tutulduğunda (program konunun değerlerini kullanır) modelin her aralıkta gösterdiği orandır; etkileşimlerin olmadığı bir modelde bu koşul hiçbir şeyi değiştirmez, ancak aşağıdaki 2. Durum'da her şeyi değiştirir:

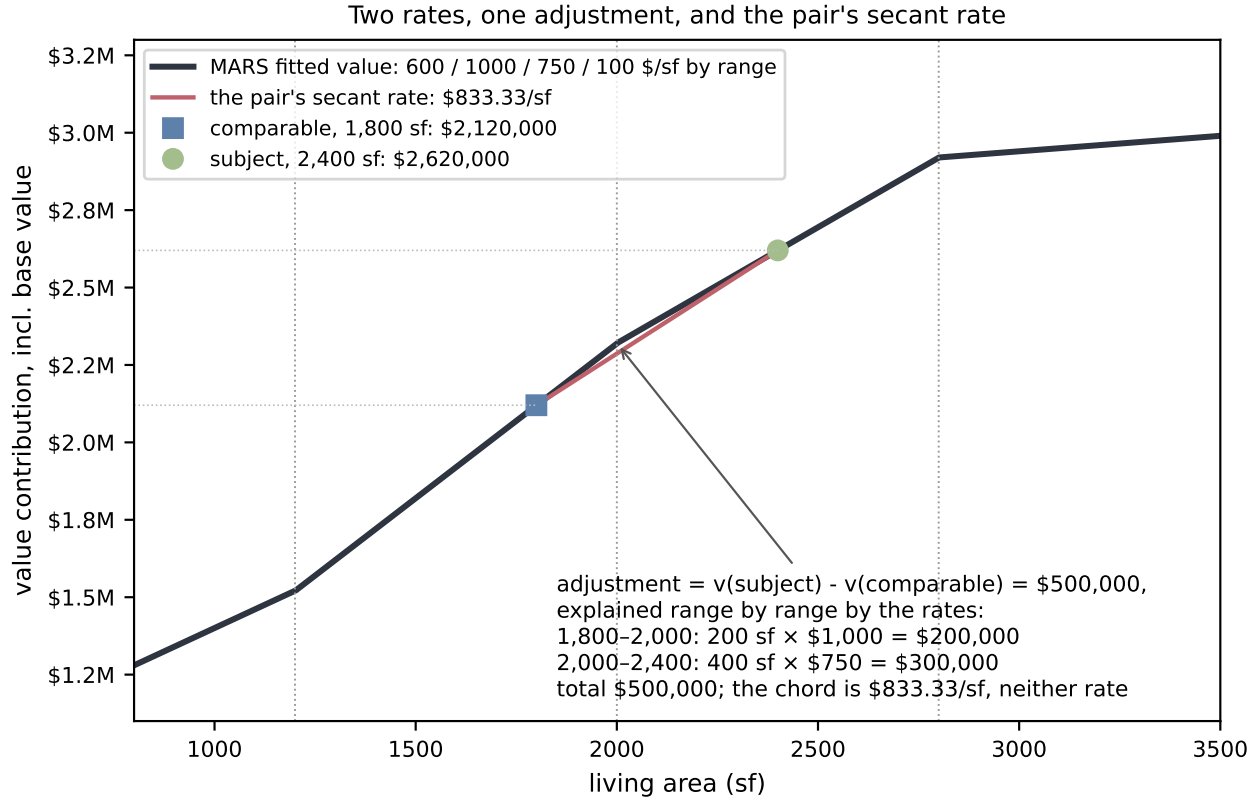
alan (sf)	aktif menteşeler	oran (\$/sf)
1.200'ün altında	$h(1200 - gla)$ sadece	$-(-600) = +\mathbf{600}$
1.200 ila 2.000	$h(gla - 1200)$	$\mathbf{1000}$
2.000 ila 2.800	$h(gla - 1200), h(gla - 2000)$	$1000 - 250 = \mathbf{750}$
2.800'ün üzerinde	üç sağ menteşe de	$1000 - 250 - 650 = \mathbf{100}$

İlk satır, gözden kaçan adımdır. Yansıtılmış bir menteşedeki -600 katsayı, $-\$600$ bir fit kare başına oran değildir. $h(1200 - gla)$ menteşesi, bir evin 1.200 fit kareye ne kadar yetmediğini ölçer ve -600 bu eksikliğin katı bir indirimdir: $\$240,000$ 800 fit karelik bir ev için, 1.200'de sıfıra iner. Eklenen her fit kare, indirimin $\$600$ kısmını ortadan kaldırır, dolayısıyla oran $+\$600$ olur. (Pozitif bir katsayıyla, aynı menteşe daha küçük olması nedeniyle değer katar ve oranı negatiftir.) Kural, bir kez belirtilmiştir: bir aralıkta, oran, aktif sağ menteşelerin katsayılarının toplamından aktif yansıyan menteşelerin katsayılarının toplamının çıkarılmasıyla elde edilir; ayrıca, modelde varsa, doğrusal terimin katsayısı (düğüm içermeyen bir düz çizgi olarak girilen bir özellik) eklenir. Bununla birlikte bir kural vardır: tam olarak bir düğüm noktasında oran sağdan alınır; böylece bir sağ menteşe düğüm noktasından itibaren sayılır ve bir yansıtılmış menteşe ise yalnızca düğüm noktasının tam altında sayılır; dolayısıyla 1.200 fit kare'deki oran, oradan başlayan aralığın oranı olan $\$1,000$ 'dir ve program da bunu aynı şekilde okur.

Bir tablonun doğru şekilde oluşturulup oluşturulmadığının kontrolü, düğüm noktalarındaki uyum değeriyle yapılır. 1.200 fit kare değerinde her bir menteşe sıfırdır ve değer kesim noktasıdır: $\$1,520,000$; 2.000 değerinde ise $\$1,520,000 + 1,000 \times 800 = \$2,320,000$; 2.800 değerinde $\$2,320,000 + 750 \times 800 = \$2,920,000$; ve verilerdeki en küçük ev olan 800 fit kare değerinde ise $\$1,520,000 - 600 \times 400 = \$1,280,000$ 'dir. 800'den 1.200'e kadar değer $\$240,000$ 'e yükselir; bu, $\$600$ 'lük bir orandır ve bu, ilk satırdır; katsayıyı oran olarak okuyan bir tabloda, değer 1.200'deki düğüme kadar soldan sağa doğru azalırdı, ancak değerler bunun böyle olmadığını göstermektedir.



Şekil 2: Örnekteki dört menteşe terimi, bunların toplamı ve her aralıktaki oran. Yansıtılmış menteşe $h(1200 - gla)$ üzerindeki katsayı -600 , okuyucunun bunu oranla karıştırabileceği bir yerde işaretlenmiştir; 1.200 fit kare altındaki oran $+600$ 'dir.



```

Model (v = value contribution, gla = living area in sf):
v = 1,520,000 - 600·h(1200-gla) + 1000·h(gla-1200) - 250·h(gla-2000) - 650·h(gla-2800)
as earth prints it:
(Intercept) 1520000 h(1200-gla) -600 h(gla-1200) 1000 h(gla-2000) -250 h(gla-2800) -650
Adjustment from the comparable (1,800 sf) to the subject (2,400 sf) =  $v(2,400) - v(1,800) = \$500,000$ ;
explained range by range by the rates:
1800-2000 sf: 200 sf  $\times$  $1000/sf = $200,000
2000-2400 sf: 400 sf  $\times$  $750/sf = $300,000
total = $500,000 secant rate for the pair:  $500,000 / 600 = \$833.33/\text{sf}$ 

```

Şekil 3: 1.800 fit kare büyüklüğündeki emsal nesne ile 2.400 fit kare büyüklüğündeki konu nesnesine göre hesaplanan değer. Düzeltme, yani iki katkının farkı, \$500,000 olup, yürürlükteki oranlar bunu aralık aralık açıklamaktadır; iki noktayı birleştiren kord, yaklaşık \$833 bir fit kare, bu çiftin sekant oranıdır; bu, bir düz çizginin sadece bu çift için ihtiyaç duyacağı tek orandır; sıradan bir regresyon, her çift için tek bir oran kullanılmalıdır.

Şekil2, dört menteşeyi, bunların toplamını ve her aralıktaki oranı gösterir; okuyucunun bunu oranla karıştırabileceği yerlerde -600 katsayısı işaretlenmiştir.

Programın başa çıktığı iki ek karmaşıklık ve bir el tablosu şunları içermelidir:

- **Etkileşimler.** earth programı, ileri geçiş aşamasında (terimlerin tek tek eklendiği aşama; Friedman 1991, §3), bir özelliğin oranı başka bir özelliğe bağlı olduğunda (bir *etkileşim* durumunda) $h(\text{gla} - 1500) \times \text{lot}$ gibi bir çarpım sonucunu verir. Bu durumda, “lot”un oranı 1.500 fit kare yaşam alanı üzerinde ve altında farklılık gösterir; bir özelliğin oran tablosu, her zaman diğerlerinin sabit tutulduğu (konu değerlerinde tutulduğu) tablodur. Program bir noktayı ve diğer tüm özelliklerin değerlerini alır ve bu noktada her oranı hesaplar.
- **Seviye ve düzeltme.** Kesim noktası ve diğer özelliklerdeki terimler, katkının ölçüldüğü seviyeyi belirler; bunlar her düzeltmede birbirini iptal eder. Program, katkıları yalnızca o özelliğin terimleri olarak raporlar; böylece iki katkı, tam olarak düzeltme kadar farklılık gösterir.

İki durum: etkileşimsiz ve etkileşimli modeller

Bir önceki bölümdeki kural, aşağıda kanıtlanan iki teorem olarak ifade edilebilir ve bu teoremler, bir MARS modelini okuyan her okuyucunun er ya da geç sorduğu bir soruya net bir yanıt verir: Bir oran, basılı terimlerin kaç tanesine gerçekten bağlıdır? Cevap, iki duruma ayrılır.

Durum 1: etkileşim yok

Ayar. $h(u) = \max(0, u)$ olsun; burada u gerçek bir sayıdır. $x, z_1, \dots, z_q$ özellikleri üzerinde etkileşim içermeyen bir MARS modelinin uyum değeri şöyledir:

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

Burada β_0 kesim noktasıdır; her bir g_k yalnızca z_k ’in bir fonksiyonudur; ve g , x ’daki tüm terimleri bir araya getirir:

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

Burada $t_1 < t_2 < \dots < t_K$, x içindeki menteşelerin farklı düğümleridir; a_j , t_j ’daki sağ menteşenin katsayısıdır ve b_j , t_j ’daki yansıyan menteşenin katsayısıdır; modelde böyle bir terim olmadığında her ikisi de 0 olur; ve c , düğüm içermeyen bir düz çizgi olarak girilen bir özellik olan $c \cdot x$ ’in çıplak doğrusal teriminin katsayısıdır; bu, earth tarafından yalnızca `linpreds`’de adı geçen özellikler için veya herhangi bir özellik için `Auto.linpreds` açık olduğunda üretilir (Teorem 1’den sonraki açıklamaya bakınız); bu ekteki her modelde olduğu gibi böyle bir terim olmadığında $c = 0$ olur. $t_0 = -\infty$ ve $t_{K+1} = +\infty$ olsun ve $i = 0, 1, \dots, K$ için x ’nin *segmentleri* olan $S_i = (t_i, t_{i+1})$ olarak adlandıralım; t_i , S_i ’nin sol düğümü ve t_{i+1} ise sağ düğümüdür. Her bir S_i üzerinde, g ’nin her bir terimi ya tam olarak sıfırdır ya da doğrusaldır; dolayısıyla g , S_i üzerinde doğrusaldır; buradaki eğimi, f ’nin başka hiçbir teriminin x ’a bağlı olmaması nedeniyle S_i üzerinde de $\partial f / \partial x$ olduğundan, S_i üzerindeki orandır ve r_i olarak yazılır. x ’nin oran tablosu, $r_0, r_1, \dots, r_K$ listesidir. ($K = 0$ adresinde tek bir segment varsa, $(-\infty, +\infty)$ adresindeki tablo, $r_0 = c$ adresindeki tek orandır.) Bir düğüm noktasında g adresindeki oranın türevi yoktur: sol ve sağ türevleri sırasıyla r_{i-1} ve r_i olup, bunlar

$a_i + b_i$ (aşağıdaki Teorem 2) kadar farklılık gösterir; dolayısıyla t_i adresinde bir oran tablosunun gösterdiği değer, burada sağdan alınan bir kural olup, g' adresindeki bir değer değildir.

Teorem 1 (bir segmentteki oran). Bu bağlamdaki gösterimlere göre, her $i \in \{0, 1, \dots, K\}$ için,

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

buradaki boş toplam 0'dir. Özellikle, r_i , $2K$ 'daki menteşe katsayılarından en fazla K 'a (her düğüm için bir tane) ve c 'a bağlıdır.

Teorem 1'in Kanıtı. i ve $x \in S_i$ noktasını sabit alalım; böylece $t_i < x < t_{i+1}$ olur. g ifadesinin her bir terimini, S_i içinde kalacak kadar küçük bir x mahalle üzerinde ele alalım. $c \cdot x$ teriminin türevi c 'dir. t_j 'daki sağ menteşe noktası için: eğer $j \leq i$ ise o zaman $t_j \leq t_i < x$ olur, dolayısıyla bu komşuluk bölgesinde $h(x - t_j) = x - t_j$ olur ve türevi 1'dir; eğer $j \geq i + 1$ ise o zaman $x < t_{i+1} \leq t_j$ olur, dolayısıyla orada $h(x - t_j) = 0$ olur ve türevi 0'dir. t_j 'daki yansıyan menteşe noktası için: eğer $j \geq i + 1$ ise o zaman $x < t_j$ olur, dolayısıyla $h(t_j - x) = t_j - x$ olur ve türevi -1 'dir; eğer $j \leq i$ ise o zaman $t_j < x$ olur, dolayısıyla $h(t_j - x) = 0$ olur ve türevi 0'dir. noktasında türevlenebilir fonksiyonların sonlu bir toplamı, türevlerin toplamı ile o noktada türevlenebilir; bu da

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

belirtilen formülü verir. Sayılamaya gelince: $\{j : j \leq i\}$ ve $\{j : j \geq i + 1\}$ indeks kümeleri birbirinden ayrıdır ve birlikte $\{1, \dots, K\}$ kümesini tamamen kapsar; dolayısıyla her bir düğüm t_j , a_j ve b_j 'dan tam olarak birine katkıda bulunur ve asla her ikisine birden katkıda bulunmaz; bu nedenle en fazla K menteşe katsayısı girer ve r_i , S_i 'in her noktasında aynıdır. ■

Sınır değere ulaşılır: Her j için $b_j = 0$ ise, en sağdaki S_K segmenti, a_j 'in tüm K değerlerini kullanır.

Not (c değeri sıfır olmadığında). earth, bir özelliği düğüm içermeyen tek bir düz çizgi olarak ifade edebilir; ayarın " $c \cdot x$ " terimi bunu iki farklı şekilde yapar ve aradaki fark önemlidir. "linpreds", "linpreds = c("GLA")" gibi adlandırılmış özelliklerin bir listesidir ve düğümlü bir uyum daha iyi olsa bile listedeki her özelliğin doğrusal olarak girilmesini *zorlar*; bu, analistin kararıdır ve uyumun kalitesini düşürebilir. "Auto.linpreds", varsayılan olarak açık olan tek bir anahtardır ve hiçbir şeyi zorlamaz: yalnızca, uyumu gerçekten doğrusal olduğu ortaya çıkan bir özelliğin (en küçük gözlemlenen değerde düğümü bulunan tek bir menteşe) earth tarafından nasıl yazılacağını değiştirir; bunu bir menteşe olarak değil, katsayı ve değişken adı olarak yazdırır; uyum aynı kalır. Her iki durumda da earth, aşağıdaki örnekte olduğu gibi, earth'ün özet biçiminde 221.4 * GLA şeklinde, terimi katsayısı ile birlikte sadece değişken adı olarak yazdırır. Theorem 1, bunu her segmentteki orana ekleme yapan c aracılığıyla işler. Örneğin, şu çağrı:

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

her özelliği doğrusal olarak işler (degree = 1 ile, böylece doğrusal terimlerin çarpımları da uyum sağlanmadan kalır) ve earth'ün özet biçiminde (uyum sağlanmış bir model değil, örnek çıktı olarak) şunu yazdırır:

```

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

```

Yaşam alanı için bu, $c = 221.4$ ayarı ve menteşe içermeyen $K = 0$ ayarıdır: tek bir oran verir; arsa büyüklüğü, banyo ve yatak odaları da aynı şekilde okunur: $\$11.80$ $r_0 = c = \$221.40$ bir fit kare arsa, $\$34,872$ bir banyo ve $\$17,641$ bir yatak odası. Böyle bir model, bu ekte kastedilen anlamda sıradan bir regresyondur; her özellik için tek bir oran vardır. Veriler üzerinde bu tür bir terim, düğümü gözlemlenen en küçük değerde bulunan tek bir sağ menteşe ile aynı çizgidir ve `Journal, Auto.linpreds = FALSE` ile çalışır; bu nedenle modellerinde, değişken grafiği veriler üzerinde düz olan bir özellik, o tek menteşe olarak görünür; örneğin `h(gla-800) 700`, verilerdeki en küçük evdeki düğüm, $c = 0$, $K = 1$ ve $a_1 = 700$: oran, düğümün altında satışın gerçekleşmediği bir aralıkta 0 ve düğümün üzerinde 700 seviyesindedir; dolayısıyla veriler üzerindeki oran tablosu tek satırlıdır. Buradan bir sonuç çıkar; bir düğümdeki oran değişimi, aşağıdaki Teorem 2’de ele alınmaktadır.

Sonuç 1 (yansıma işareti). Katsayısı b_j olan bir yansıma menteşesi, t_j adresinin solundaki her segmente $-b_j$, sağındaki her segmente ise 0 değerini katar. Örnekte, $t_1 = 1,200$ adresindeki $b_1 = -600$, r_0 adresine $+600$ değerini katar; bu da bu ekin temelini oluşturan tuzağın tamamıdır.

Kanıt. Teorem 1’deki ikinci toplamı inceleyin. ■

Yedi düğümlü bir örnek ve okurları yenen varsayım

Bu sayının önemi, neredeyse herkesin basılı bir MARS modeline ilişkin yaptığı bir varsayımdan kaynaklanmaktadır: Bir düğümün yanına basılan katsayının, o düğümün solundaki segmentin eğimi olduğu varsayımı. Oysa durum böyle değildir. **Bir menteşenin katsayısı, soldan gelen eğime yapılan bir düzeltmedir, eğimin kendisi değildir;** dolayısıyla bir segmentin eğimi, ilk segmentin eğimi artı sol düğümündeki ve daha solundaki her düğümdeki bu tür düzeltmelerin toplamıdır. Yanlış varsayımı benimseyen bir okuyucu, hesaplamayı kafasında, küçük bir programda veya bir hesap tablosu formülünde dener, sayıların uyum değerlerini vermediğini fark eder ve MARS’ın anlaşılabilir olduğu sonucuna vararak vazgeçer. MARS anlaşılabilir değildir; birikimseldir.

Teorem 2 (kümülatif toplam). Durum 1’deki gösterim sistemine göre, her $i \in \{1, \dots, K\}$ için,

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

Dolayısıyla, her i için $r_i = r_0 + \sum_{j=1}^i (a_j + b_j)$ geçerlidir; burada $r_0 = c - \sum_{j=1}^K b_j$ en soldaki segmentteki oran, $r_K = c + \sum_{j=1}^K a_j$ ise en sağdaki segmentteki orandır.

Teorem 2’nin Kanıtı. $i \in \{1, \dots, K\}$ sabit alınsın. Teorem 1’e göre,

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

Çıkarma işlemi yapıldığında, c terimi ortadan kalkar; a terimlerinin iki toplamı, a_i terimi kadar farklılık gösterir; ve $\sum_{j=i}^K b_j = b_i + \sum_{j=i+1}^K b_j$ olduğundan, b terimlerinin iki toplamı, b_i terimi kadar farklılık gösterir; bu terim, r_{i-1} ifadesinde çıkarılırken, r_i ifadesinde çıkarılmaz. Dolayısıyla $r_i - r_{i-1} = a_i + b_i$. i yerine $j = 1, \dots, i$ için bu özdeşliği eklediğimizde, sol taraflar birbirine girer:

$(r_1 - r_0) + (r_2 - r_1) + \dots + (r_i - r_{i-1}) = r_i - r_0$, bu da toplam formülünü verir. r_0 ve r_K değerleri, a 'ların toplamının boş olduğu $i = 0$ adresindeki Teorem 1'de ve b 'ların toplamının boş olduğu $i = K$ adresindedir. ■

Sonuç 2 (okuma kuralı). Durum 1 kapsamında, her $i \in \{1, \dots, K\}$ için, sol düğümü t_i olan segmentin oranı $r_i = r_{i-1} + a_i + b_i$ 'dir: bir düğümde yazılı katsayılar, soldan gelen orana eklenir ve bu katsayılar, o noktadan başlayan segmentin oranı değildir. Bu ikisi, tam olarak $r_i = a_i + b_i$ olduğunda, ile çakışır $r_{i-1} = 0$; özellikle, sadece sağ menteşelere sahip ve doğrusal terimi olmayan bir model için bu durum $i = 1$ 'de gerçekleşir ve Teorem 2'ye göre, ondan önceki düğümlerdeki katsayıların toplamı sıfır olmadığı sürece daha sonraki hiçbir düğümde gerçekleşmez.

Kanıt. İlk eşitlik, Teorem 2'dir. $r_i = a_i + b_i$, ancak ve ancak $r_{i-1} = 0$ bundan çıkar. Eğer $b_j = 0$ ise, tüm j ve $c = 0$ için, Teorem 2'ye göre $r_0 = 0$ olur; dolayısıyla $r_1 = a_1$ ve $r_{i-1} = \sum_{j < i} a_j$ olur; bu da, ancak ve ancak bu katsayıların toplamı sıfır olduğunda 0 olur. ■

Sonuç 3 (integral olarak düzeltme). Durum 1'in çerçevesi içinde, g her düğüm noktasında süreklidir; zira her menteşe süreklidir ve dolayısıyla $f(\cdot, z)$ de her z için süreklidir; işte bu, bir düğüm noktasındaki değerin, oraya soldan ya da sağdan ulaşılsa da aynı olmasını sağlar ve oranların bir miktardan diğerine aktarılabilmesini mümkün kılar. x_c ve x_s , bir karşılaştırma nesnesi ve bir denek için x 'in miktarları olsun; diğer özellikler her ikisi için de aynı değerlerde tutulduğunda, $g(x)$ olur; f 'nin x 'yi içeren terimleri, x miktarında x 'nin katkısıdır. Bir aralığın uzunluğu I için $u_1 = \min(x_c, x_s)$, $u_2 = \max(x_c, x_s)$ ve $|I|$ yazalım. O halde, $g(x_s) - g(x_c)$ düzeltmesi, emsal nesnesinin katkısını konu katkısına eşit hale getirmek için eklenen miktar şöyledir:

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

Başka bir deyişle: düzeltme, iki miktar arasındaki oranın integrali olup, oran her segmentte sabit olduğundan integral, yol üzerinde karşılaşılan segmentler üzerinden, oranın genişlikle çarpımının toplamıdır.

İspat. g fonksiyonu süreklidir ve sonlu sayıda parçadan oluşan parça parça doğrusal bir fonksiyondur; dolayısıyla mutlak süreklidir. Bu nedenle, diferansiyel hesabın temel teoremi uyarınca $g(x_s) - g(x_c) = \int_{x_c}^{x_s} g'(u) du$ elde edilir; türev, düğümler hariç her yerde mevcuttur ve düğümler sonlu bir kümedir. Her bir S_i üzerinde, Teorem 1'e göre integrand r_i 'ye eşittir; dolayısıyla $\int_{u_1}^{u_2} g'(u) du$, i üzerindeki r_i 'nin $S_i \cap [u_1, u_2]$ uzunluğuyla çarpımının toplamıdır; kesilmeyen parçalar 0 uzunluğuna katkıda bulunur; ve $\int_{x_c}^{x_s} = \text{sign}(x_s - x_c) \int_{u_1}^{u_2}$, integralin yönelimine göre elde edilir. ■

Oran tablosu g' şeklindedir; doğrudan $g(x_s) - g(x_c)$ olarak hesaplanan düzeltme, bu iki büyüklük arasındaki integraline eşittir; ve Teorem 1 ile 2, g' 'in neyden oluştuğunu belirtir. Uygulamada düzeltme her zaman katkıların farkı olarak hesaplanır; bu işlem için oranlara hiç gerek yoktur; sonuç ise, aralık aralık okunduğunda oranların neden bunu yeniden ürettiğini açıklar; işte bu da oranları düzeltmenin bir açıklaması ve kontrolü haline getirir, onu hesaplamanın bir yolu değil.

En küçük evi 800 fit kare olan ve yedi düğüm içeren bir veri kümesindeki yaşam alanını ele alalım; bu düğümler 1.200, 1.600, 2.000, 2.400, 2.800, 3.200 ve 3.600 fit kare büyüklüğünde evler bulunan bir veri kümesindeki yaşam alanını ele alalım. Modelde dokuz terim vardır; bunlardan ikisi menteşeleri yansıtmaktadır ve her biri aynı düğümdeki bir sağ menteşeyle eşleştirilmiştir: 1.200 ve 2.400'de:

h(1200-g1a) -500

$h(\text{gla}-1200)$	700
$h(\text{gla}-1600)$	-100
$h(\text{gla}-2000)$	150
$h(2400-\text{gla})$	-200
$h(\text{gla}-2400)$	-250
$h(\text{gla}-2800)$	-300
$h(\text{gla}-3200)$	-150
$h(\text{gla}-3600)$	50

Yedi düğüm, sekiz segment oluşturur. Her bir segmentteki oran, aktif sağ menteşelerin katsayılarının toplamından, aktif yansıtılmış menteşelerin katsayılarının toplamının çıkarılmasıyla elde edilir:

segment (sf)	aktif terimler	oran (\$/sf)
800 ila 1.200	$h(1200 - \text{gla}), h(2400 - \text{gla})$	$500 + 200 = 700$
1.200 ila 1.600	$h(\text{gla} - 1200), h(2400 - \text{gla})$	$700 + 200 = 900$
1.600 ila 2.000	sağdaki menteşeler 1.200, 1.600; $h(2400 - \text{gla})$	$700 - 100 + 200 = 800$
2.000 ila 2.400	sağ menteşeler 1.200 ile 2.000 arasında; $h(2400 - \text{gla})$	$700 - 100 + 150 + 200 = 950$
2.400 ila 2.800	sağ menteşeler 1.200 ile 2.400 arasında	$700 - 100 + 150 - 250 = 500$
2.800 ila 3.200	sağ menteşeler 1.200 ile 2.800 arasında	$500 - 300 = 200$
3.200 ila 3.600	sağ menteşeler 1.200 ile 3.200 arasında	$200 - 150 = 50$
3.600'ün üzerinde	yedi menteşenin tamamı sağ tarafta	$50 + 50 = 100$

Tablodan okunması gereken üç şey. En sağdaki segmentin oranı, $700 - 100 + 150 - 250 - 300 - 150 + 50 = 100$, yedi sağ menteşenin katsayılarını birleştirir; her düğüm için bir tane olmak üzere, bu da Teorem 1'in $K = 7$ ile sınırını oluşturur; iki yansıyan menteşe burada rol oynamaz, çünkü yansıyan bir menteşe kendi düğümünün sağında düzdür ve en sağdaki segment her düğümün sağında yer alır. En soldaki segment, 800'den 1.200'deki sağ düğüme kadar olan kısım, bunun ayna görüntüsüdür: her düğüm kendi sağ düğümünün üzerinde veya sağında yer aldığı için, yalnızca yansıtılmış menteşeler dikkate alınır ve işaretleri tersine çevrilir. Ve bir çift taşıyan her düğümde adım, iki katsayının toplamıdır: 1.200'de, 700'den $900 - 500 + 700 = +200$; 2.400'de oran 950'den 500'e düşer; bu, -450 'luk bir değişikliktir ve bu, tam olarak o düğümdeki iki katsayının toplamına eşittir: -250 ve -200 . Düğümü geçmek, sağ menteşeyi açar ve yansıyan menteşeyi kapatır; her iki değişiklik de aynı yönde etki eder.

En sağdaki segmentteki aynı orana yürüyerek de ulaşılabilir: ilk segmentte 700'den başlayın, 800'den 1.200'e ilerleyin ve geçtiğiniz her düğümde, o noktada bulunan katsayıların toplamını ekleyin: $+200, -100, +150, -450, -300, -150, +50$. Bu rota sekiz segmentin ve dokuz terimin tümüne dokunur ve yine aynı 100'e ulaşır. Doğrudan toplam daha kısadır ve programın yaptığı budur; yürüyüş ise, bir düğümün yanındaki katsayının bir seviye değil bir adım olduğunu görmek için okuyucunun elle bir kez yapması gereken şeydir. 2.400'deki -450 iki yarım adımdır: düğümü geçmek, yansıtılmış menteşeyi $h(2400 - \text{gla})$ geride bırakır ve bununla birlikte düğümün altına eklediği $+\$200$ 'i de geride bırakır; ve -250 'si orada başlayan sağ menteşeyi $h(\text{gla} - 2400)$ 'yi getirir. $-200 - 250 = -450$ ve oran $\$950$ 'den $\$500$ 'e düşer.

Şekil 4, gözle kontrol edilebilecek kadar küçük olan beş düğümlü bir model için de aynı sonucu ortaya koymaktadır. Çubuklar, düğümlerde yazan katsayıları; merdiven ise her bir segmentteki oramı temsil eder; merdiven, yansıtılmış menteşe tarafından belirlenen ilk segmentin oramı olan $\$900$ 'den başlar ve her basamak, o düğümdeki çubukların toplamıdır. En sağdaki segmentteki oran,

$650 - 250 + 150 - 200 - 150 = 200$, beş sağ menteşenin tümünü kullanır; 1.200'de bulunan ve -900 olarak basılan yansıyan menteşe, ilk segmenti $+900$ olarak ayarlar ve başka hiçbir şeye dokunmaz.

Bu grafiğin şekliyle ilgili iki hususu belirtmekte fayda var; zira bunlar MARS'tan ziyade evlerle ilgili. İlk olarak, eksen sıfırdan değil 800 fit kare'den başlıyor ve grafik eksenden başlıyor: bu ekteki uyumlu bir modelin her grafiği, modelleme yazılımının çizdiği grafikler gibi, verilerdeki en küçük evden en büyüğe kadar uzanır ve bu aralığın dışında hiçbir şey çizmez. Her piyasa bölgesinde, altında satışın gerçekleşmediği en küçük bir ev vardır ve değişken grafik bunun altındaki değerleri göstermez: sıfır değil, ancak satış verilerinin desteklemediği bir değer; formül yine de orada bir sayı verecektir ve bu sayı ekstrapolasyondur. Şekillerde 800 fit kare alt sınır kullanılmaktadır, ancak bu alt sınır piyasa bölgesinin bir nitelikidir, sabit bir değer değildir: farklı piyasa bölgelerinde MLS'de görünen taşınmazların minimum büyüklükleri farklıdır ve bu sınırın nerede olduğu, değerlendirme uzmanı değil, veri kümesi tarafından belirlenir. İkincisi, eğri şekli pazardan pazara değişir: Kıyı örneğinde en dik aralık 1.200 ila 2.000 fit kare arasındadır ve Şekil4'de bu, ilk segmenttir. Atherton veya Hillsborough gibi yüksek fiyatlı bir bölgede, yaşam alanı (GLA) eğrisinin en dik segmenti, alıcıların bulunduğu 4.000 veya 5.000 fit kareyi aşan bir noktada yer alabilir. Ayrıca, bir inşaat firmasının belirli bir dönemde tek tip büyüklükte ve sorunlu evler inşa ettiği bir mahallede, 1.600 ile 2.000 arasında gösterilen grafikte olduğu gibi, orta aralıkta oranda keskin bir düşüş görülebilir. Daha geniş bir piyasa bölgesi içinde değişken grafikler çeşitlilik gösterir ve oran tablosu, değerlendirme uzmanının belirli bir grafiğin ne anlama geldiğini nasıl yorumladığını gösterir. Şekil 5 ve 6, bu şekillerden ikisini, basılı olarak görüneceği şekilde çizmektedir; böylece okuyucu, kurallara göre her bir oran tablosunu oluşturabilir ve bunu merdiven grafiğiyle karşılaştırabilir.

Durum 2: etkileşimler

Bir etkileşim terimi, farklı değişkenlerdeki faktörlerin çarpımının bir katsayı ile çarpılmasıyla elde edilir; bu faktörlerden her biri bir menteşe ya da değişkenin kendisidir; örneğin, $\gamma \cdot h(\text{gla} - 1500) \cdot \text{lot}$. Durum 1'deki her şey, tek bir kelime hariç, geçerliliğini korur: bir segmentteki oran artık bir sabit değil, diğer değişkenlerin bir fonksiyonudur. Aşağıda bu ifade ve kanıtı yer almaktadır; ardından bunun değerlendirme uzmanlarının hesaplamalarına nasıl yansıdığı ele alınmaktadır.

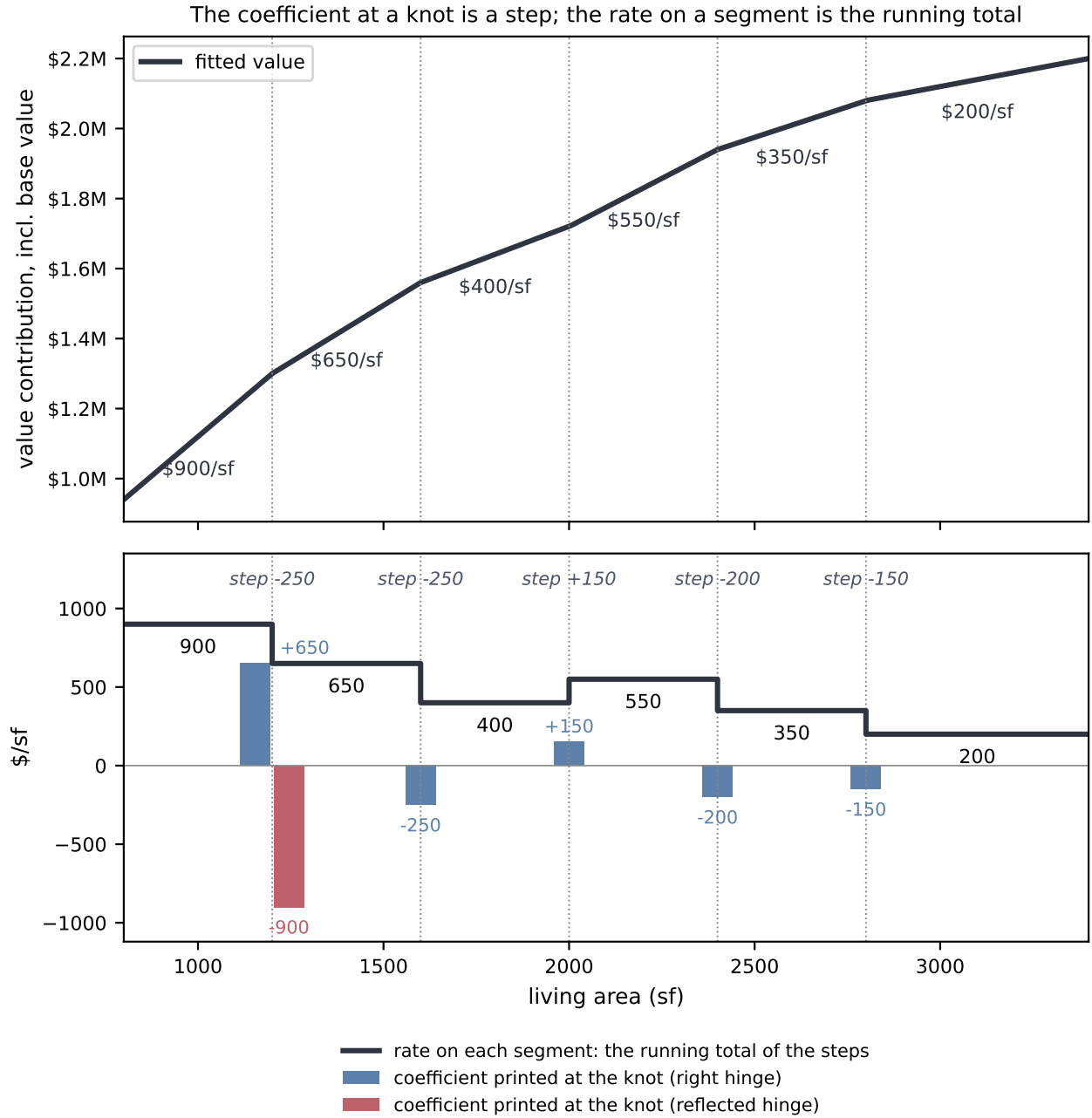
Ayar. $x, z_1, \dots, z_q$ özellikleri üzerinde etkileşimler içeren bir MARS modeli, şu değeri uydurmuştur:

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

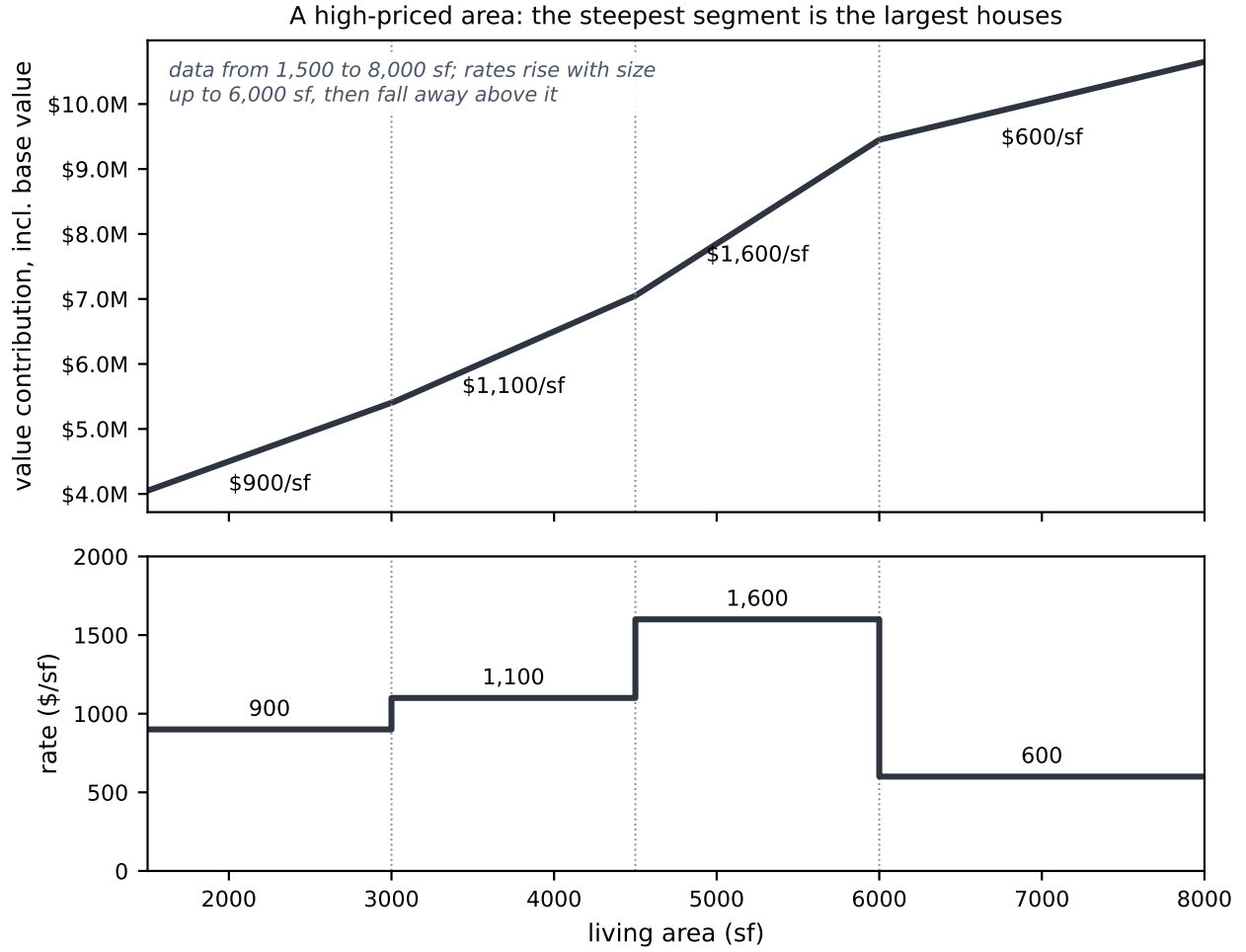
Burada M , kesim noktası dışındaki terimlerin sayısıdır; β_0 kesim noktasıdır ve model, kesim noktası ile terimlerinin toplamından oluşur; m terimi ise $\gamma_m \prod_{k \in F_m} B_{m,k}(z_k)$ şeklindedir: bu terimin katsayısı γ_m , terimdeki her bir özellik için bir faktör ile çarpılır. Burada $z = (z_1, \dots, z_q)$ ve, bir indeksin her özelliği taraması için, z_0, x 'yi temsil eder. m teriminde görünen özellikler $F_m \subseteq \{0, 1, \dots, q\}$ kümesini oluşturur; hiçbir özellik bir terimde iki kez görünmez ve her bir faktör $B_{m,k}, z_k$ özelliğinin bir fonksiyonudur ve üç türden birine aittir: özelliğin kendisi, z_k ; bir sağ menteşe $h(z_k - t)$; veya bir yansıtılmış menteşe $h(t - z_k)$. $M_x = \{m : 0 \in F_m\}$, x içeren terimler olsun. $m \in M_x$ için, $B_{m,0}(x)$, x içindeki terimin faktörüdür ve terimin diğer özellikleri, yani $k \in F_m, k \neq 0$ olan z_k , 0 terimde x 'in ortaklarıdır. Ortak faktörlerin çarpımı için

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

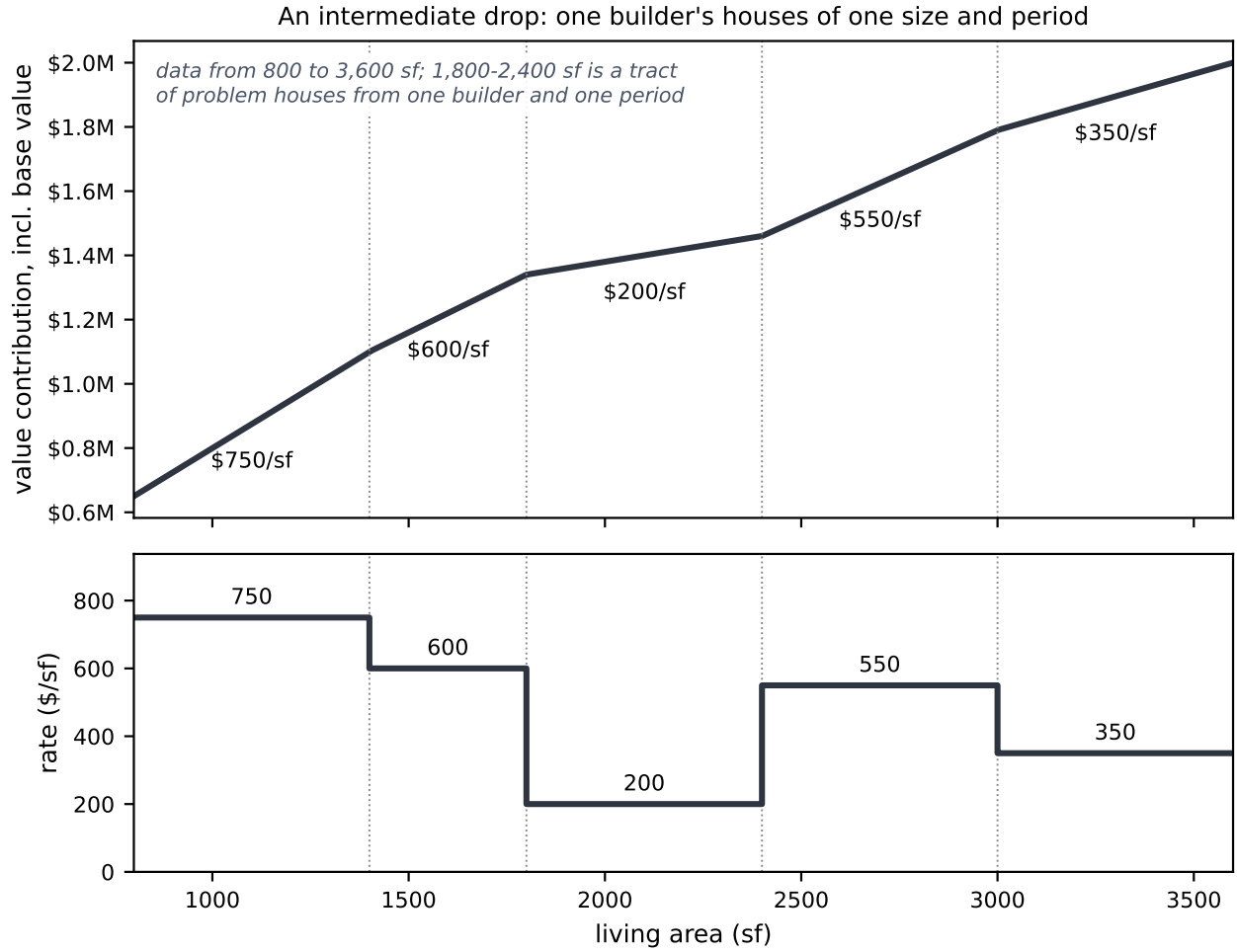
yazın; boş çarpım ise 1 şeklindedir. Başka bir deyişle: konunun diğer özellikleri doldurulduktan sonra, bir terimin ortakları çarpılarak tek bir sıradan sayıya dönüşür; bu sayı, yazdırılan katsayıyla



Şekil 4: 800 fit kare ile başlayan verilere dayalı beş düğümlü bir model: $-900 h(1200 - gla) + 650 h(gla - 1200) - 250 h(gla - 1600) + 150 h(gla - 2000) - 200 h(gla - 2400) - 150 h(gla - 2800)$. Yukarıda, her bir segmente oran yazılmış uyum içinde bulunan değer gösterilmektedir; 800 fit karenin altında bir değer bulunmamaktadır, çünkü bu büyüklükte evler yoktur. Aşağıda, her düğümde bir çubuk olarak gösterilen katsayı ve her segmentteki oran bir merdiven olarak gösterilmiştir: merdivenin her basamağı, o düğümdeki çubukların toplamıdır (burada bir çubuk, bir düğümün bir çift taşıdığı yerde iki çubuk, Teorem 2), dolayısıyla bir segmentteki oran, yanındaki katsayı değil, ilk segmentin oranı artı solundaki basamakların kümülatif toplamıdır: her katsayı, soldan gelen oranı düzeltir.



Şekil 5: 1.500 ila 8.000 fit kare aralığındaki verilere göre pahalı bir bölge: $-900h(3000 - gla) + 1100h(gla - 3000) + 500h(gla - 4500) - 1000h(gla - 6000)$, ilk düğümde \$5,400,000 ile kesişir. \$900, \$1,100, \$1,600 ve \$600 adreslerinde fit kare başına fiyatlar: en dik artış 4.500 ila 6.000 fit kare aralığında görülür; alıcıların yoğun olduğu bu aralıkta fiyatlar en hızlı yükselir, ilk aralıkta artış daha yavaştır ve 6.000 fit kare üzerinde fiyat artışı azalır. Kural değişmez; sadece piyasa değişir.



Şekil 6: 800 ila 3.600 fit kare aralığındaki verilere göre bir ara düşüş: $-750 h(1400 - gla) + 600 h(gla - 1400) - 400 h(gla - 1800) + 350 h(gla - 2400) - 200 h(gla - 3000)$, ilk düğüm noktasında \$1,100,000 ile kesişir. Fiyatlar: \$750, \$600, \$200, \$550 ve \$350 (bir fit kare başına). 1.800 ile 2.400 arasındaki segmentin, tek bir inşaatçı tarafından tek bir dönemde inşa edilmiş, piyasa fiyatlarında sorunlar yaşayan bu büyüklükteki evlerden oluşan bir bölge olduğunu varsayalım (uydurma bir örnek); fiyat orada \$200 seviyesine düşer ve bu seviyenin üzerinde toparlanır. -400 katsayısını 1.800'de o aralığın oranı olarak kabul eden ya da oranların sadece büyüklükle birlikte düştüğünü varsayan bir okuyucu, bu bölgeyi yanlış yorumlamış olur.

çarpıldığında, basit bir yaşam alanı teriminin katsayısı olur ve terim, Durum 1'deki gibi okunur. Eş çarpımı $P_m(z)$, m teriminden x 'deki çarpanı çıkarıldığında geriye kalan kısımdır: m terimi $\gamma_m B_{m,0}(x) P_m(z)$ şeklindedir. Belirli bir noktada z bu bir sayıdır; dolayısıyla bir segment üzerinde terim, x üzerinde $\gamma_m P_m(z)$ katsayısına sahip tek bir menteşe gibi etki eder ve bu çarpım, eşlerin orana nasıl girdiğini gösterir. Eşi olmayan bir terimin $P_m(z) = 1$ değeri vardır ve bu, Durum 1'e ait bir terimdir.

Örneğin, aşağıdaki örnek hesaplamadaki $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ teriminde, yaşam alanı x olarak alındığında, x ifadesindeki çarpan $h(\text{gla} - 1200)$ 'dir; "one partner" arsa anlamına gelir ve $P_m(z) = h(\text{lot} - 6000)$ 'dir. 5.000 fit kare büyüklüğündeki bir arsa için $P_m(z) = 0$ ve bu terim orana hiçbir şey eklemes; 10.000 fit kare büyüklüğündeki bir arsa için $P_m(z) = 4,000$ ve bu terim, 1.200'de $0.03 \times 4,000 = 120$ katsayılı bir menteşe görevi görür ve \$120 1.200 fit karenin üzerindeki orana fit kare başına 120 ekler. Burada bir terimin kaç özellik içerebileceğine dair herhangi bir sınırlama yoktur; earth's **degree** argümanı bu üst sınırı belirler ve aşağıdaki teoremler her derece için geçerlidir. d özelliklere sahip bir terimin $d+1$ boyutlarında bir grafiği vardır; ancak d değeri ikiyi aştığında kimse bu grafiği çizemez; yazılım, diğerleri sabit tutulurken her seferinde bir veya iki özellik içeren dilimler gösterir ve Teorem 3, konunun kendi noktasında yapılan aynı dilimlemedir. $P_m(z)$, bu dilimlemenin gerçekleştiği yerdir: bir terimin kaç tane ortağı olursa olsun, bir noktada bunların çarpımı tek bir sayıdır. Dört özellikli $\gamma h(\text{gla} - 1200) h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$ teriminde, $\gamma = 0.001$ ile, grafiği beş boyutludur, $P_m(z) = h(\text{lot} - 6000) h(\text{age} - 10) \text{baths}$; 10.000 fit kare arsa, 20 yaş ve 3 banyo durumunda $4,000 \times 10 \times 3 = 120,000$ olur ve yaşam alanı boyunca tüm terim, katsayısı $\gamma \times 120,000 = 0.001 \times 120,000 = 120$ olan 1.200'de bir menteşe oluşturur, buna \$120 fit kare eklenir: yukarıdaki iki özellikli terimle aynıdır, ancak katsayısı otuz kat daha küçüktür. Bu nedenle, yüksek dereceli bir terimin katsayısı, ortakları büyük birimlerle ölçüldüğünde genellikle çok küçük bir sayı olarak görünür: birkaç ortak değer çarpımını çarpar. Ve eğer herhangi bir bileşen, düğümünün sıfır tarafında ise (örneğin 6.000 fit kare'nin çok altında bir alan veya 10 yıldan küçük bir yaş), o zaman $P_m(z) = 0$ olur ve diğer bileşenler ne olursa olsun bu terim hiçbir katkı sağlamaz; derece ne kadar yüksekse, belirli bir taşınmazda bir terimin devre dışı kalma olasılığı o kadar artar. Bu ektaki örneklerde iki yönlü etkileşimler kullanılmaktadır (**degree** = 2).

$t_1 < \dots < t_K$, $B_{m,0}$ içinde menteşe olan düğümlerin birbirinden farklı düğümleri olsun; burada $t_0 = -\infty$, $t_{K+1} = +\infty$ ve $S_i = (t_i, t_{i+1})$, Durum 1'deki gibi segmentlerdir. $m \in M_x$ ve $i \in \{0, \dots, K\}$ için şunu tanımlayalım:

$$s_{m,i} = \begin{cases} +1 & \text{eğer } B_{m,0}, x \text{ değişkeninin kendisi ya da } t_j \leq t_i \text{ olan bir sağ menteşe } h(x - t_j) \text{ ise,} \\ -1 & \text{eğer } B_{m,0}, t_j \geq t_{i+1} \text{ olan bir yansıtılmış menteşe } h(t_j - x) \text{ ise,} \\ 0 & \text{aksi halde.} \end{cases}$$

Bir *nokta*, diğer özelliklerin bir değeridir (z). Her $B_{m,0}$, S_i üzerinde doğrusal ya da tam olarak sıfır olduğundan, f , sabit z için S_i üzerinde x 'de doğrusaldır; buradaki eğimi, z noktasında S_i üzerindeki orandır ve $x \in S_i$ için $r_i(z) = \partial f / \partial x$ olarak yazılır.

Tüm sembolleri bir kerede görmek için, aşağıdaki örnek hesaplamadaki modeli, Earth'ün yazdırıldığı haliyle kullanın: kesim noktası 900.000, $600 h(\text{gla} - 1200)$, $-350 h(\text{gla} - 2400)$, $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ ve 15 lot. Özellikleri şunlardır: yaşam alanı, $x = z_0 = \text{gla}$ ve arsa büyüklüğü, $z_1 = \text{lot}$; dolayısıyla $q = 1$; kesim noktası $\beta_0 = 900,000$ 'dir ve $M = 4$ terim bulunmaktadır. 10.000 fit kare arsa üzerinde 2.000 fit kare yaşam alanına sahip bir taşınmazda:

m	γ_m	F_m	$B_{m,0}(x)$	$B_{m,1}(z_1)$	$P_m(z)$	mülkteki terim
1	600	$\{0\}$	$h(\text{gla} - 1200)$	yok	1	$600 \times 800 = 480,000$
2	-350	$\{0\}$	$h(\text{gla} - 2400)$	yok	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	$\{1\}$	yok	lot	$(M_x \text{ içinde değil})$	$15 \times 10,000 = 150,000$

Dolayısıyla $M_x = \{1, 2, 3\}$: 4. terim, x içinde herhangi bir faktöre sahip değildir, yaşam alanından bağımsız olarak aynı miktarı ekler ve x oranına dahil olmaz. Uyum değeri $f = 900,000 + 480,000 + 0 + 96,000 + 150,000 = \$1,626,000$ 'dir. 1. ve 2. terimlerin eş terimi yoktur, dolayısıyla $P_1(z) = P_2(z) = 1$; 3. terimin ise bir eş terimi vardır ve bu, yukarıdaki örnekteki çarpan olan $P_3(z) = h(10,000 - 6,000) = 4,000$ 'dir. Bundan çıkan Teorem 3, bunları orana dönüştürür. Yaşam alanı, $t_1 = 1,200$ ve $t_2 = 2,400$ düğümlerine sahiptir ve taşınmaz, bunların arasındaki S_1 segmenti üzerindedir; burada terimler 1 ve 3 etkindir, terim 2 ise değildir ($s_{1,1} = s_{3,1} = 1$, $s_{2,1} = 0$); dolayısıyla $r_1(z) = 600 \times 1 + 0.03 \times 4,000 = 720$, örnekteki tablodaki 10.000 fit kare taşınmaz için oran budur.

Teorem 3 (bir aralık üzerindeki oran, etkileşimler dikkate alındığında). Bu gösterim sistemine göre, her i ve her z noktası için,

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

Teorem 3'ün Kanıtı. i , z noktası ve $x \in S_i$ sabitlerini belirleyelim. $m \notin M_x$ içeren bir terim, x 'de hiçbir çarpan içermez ve x 'de sabittir; x 'deki türev değeri 0'dir. $m \in M_x$ içeren bir terim, $\gamma_m B_{m,0}(x) P_m(z)$ 'ye eşittir; burada $P_m(z)$, x 'ye bağlı değildir. Teorem 1'in ispatındaki durum analizine göre, tek faktör $B_{m,0}$ 'a uygulandığında, $B_{m,0}$ 'in x 'daki türev değeri, özellik kendisi için 1; düğümü t_i 'un üzerinde veya solunda ya da t_{i+1} 'un üzerinde veya sağında olmasına göre sağ menteşe için 1 veya 0; düğümü t_{i+1} 'un üzerinde veya sağında ya da t_i 'un üzerinde veya solunda olmasına göre yansıyan menteşe için -1 veya 0'dur; yani, $s_{m,i}$ 'ye eşittir. Dolayısıyla terimin türevi $\gamma_m s_{m,i} P_m(z)$ 'dir ve sonlu bir toplam olan f 'nin türevi, bunların toplamıdır. Değer, hangi $x \in S_i$ 'in seçildiğine bağlı değildir. ■

Teorem 1, M_x içindeki hiçbir terimin bir eşi bulunmadığı özel durumdur; dolayısıyla her $P_m(z) = 1$ olur; toplamın terimlerini düğümlerine göre gruplandırmak, orada kanıtlanan formülü verir. Üç sonuç, eşlerin nasıl değiştiğini belirtir; bunlardan sonuncusu, düğüm başına bir sınırlamasının geçerli olmamasıdır ve aşağıdaki örnekte gösterilen "terim sıfırlanabilir" durumudur.

Sonuç 4 (oran, yalnızca bir noktada bir sayıdır). Durum 2 çerçevesinde, her bir i için r_i oranı, z noktasının bir fonksiyonudur. z içinde, $s_{m,i} \neq 0$ içeren her $m \in M_x$ teriminin bir eşi olmadığı durumlarda bu oran sabittir. $s_{m,i} \neq 0$ ile bir terimin bir eşi olduğunda, $r_i(z)$, $z \mapsto \sum_m \gamma_m s_{m,i} P_m(z)$ fonksiyonudur; bu fonksiyon, terimlerin tam olarak birbirini ortadan kaldırdığı durumlar hariç, eşlerin değerlerine göre değişir; aynı segment üzerinde aynı x değerine sahip iki taşınmaz, eşlerinin çarpımları farklı olduğunda ve bu farklar birbirini ortadan kaldırmadığında, x için farklı oranlara sahiptir. Bu nedenle program bir nokta alır ve tek bir taşınmaz için tek başına bir oran tablosu mevcut değildir.

Kanıt. Teorem 3'e göre, $r_i(z) = \sum_{m \in M_x} \gamma_m s_{m,i} P_m(z)$. $s_{m,i} \neq 0$ olan her terimin bir eşi yoksa, bu tür her P_m , boş çarpım 1 olur ve toplam bir sabittir. Aksi takdirde, toplamda en az bir $\gamma_m s_{m,i} P_m(z)$ terimi bulunur; burada $\gamma_m s_{m,i} \neq 0$ ve P_m , her biri bir eşin sabit olmayan bir fonksiyonu olan bir veya daha fazla faktörün çarpımıdır; dolayısıyla, toplamın z sabit olması için bu terimlerin birbirini tamamen ortadan kaldırması gerekir. ■

Sonuç 5 (x aralığındaki bir segmentte, oran her bir eş için değişken bir grafikdir). Durum 2'nin koşulları çerçevesinde, i değerini, bir eş olarak z_p ($p \geq 1$) değerini ve diğer tüm özelliklerin değerlerini sabitleyin. O zaman $z_p \mapsto r_i(z)$, z_p 'nin kesintileri M_x 'daki terimlerin faktörlerinin $B_{m,p}$ düğümleri arasında yer alan, parçalı doğrusal bir sürekli'dir. Her bir eş sabit tutulduğunda x boyunca, oran yalnızca x 'nın $t_1, \dots, t_K$ düğümlerinde değişir; oran yüzeyi, x ve partnerlerinin ortak uzayında hem x 'nin düğümleri hem de her bir partnerin düğümleri tarafından kesilir.

İspat. Diğer özellikler sabit tutulduğunda, her bir $P_m(z)$, 1 (eğer $p \notin F_m$ ise) ya da tek faktör $B_{m,p}(z_p)$ ile çarpılmış bir sabittir; bu faktör ise z_p 'in kendisi, $h(z_p - t)$ ya da $h(t - z_p)$ 'dir: z_p 'in, en fazla bir kesme noktası olan (bu nokta düğüm noktasıdır) kesikli doğrusal bir fonksiyondur. Teorem 3'ün $r_i(z)$ için verdiği gibi, bu tür fonksiyonların sonlu bir doğrusal kombinasyonu, sürekli ve parça parça doğrusaldır; kesintiler yalnızca bazı toplama terimlerinin kesintiye uğradığı yerlerde görülür. x adresindeki ifade, Teorem 3'ün kendisidir: sabit z için, $r_i(z)$ her bir S_i üzerinde tek bir sayıdır ve segmentler yalnızca x 'in düğüm noktalarında kesişir. ■

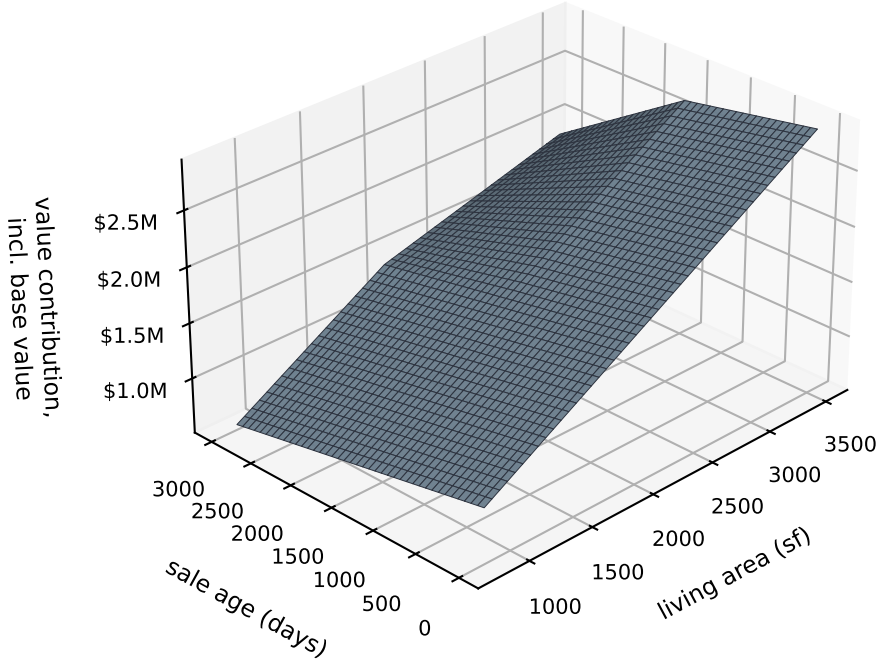
Sonuç 6 (Teorem 1'deki sınır geçerli değildir). Durum 2'de, Teorem 3'ün $r_i(z)$ formülünde sıfırdan farklı bir terim içeren terimlerin sayısı, yalnızca $|M_x|$ ile sınırlıdır; bu, x içeren terimlerin sayısıdır ve K değerini aşabilir: x şeklindeki bir düğüm, her biri farklı bir terim olan farklı ortaklarla M_x formülünün birkaç teriminde görünebilir. Tersine, $s_{m,i} \neq 0$ içeren bir terim, eş teriminin çarpımı sıfır olduğu bir noktada hiçbir katkı sağlamaz.

Kanıt. $\gamma_1 h(x - 1500)$ lot ve $\gamma_2 h(x - 1500) h(\text{age} - 10)$ terimleri, aynı düğüme sahip M_x terimlerinin farklı terimleridir ve 1,500 üzerindeki segmentte her ikisi de $s_{m,i} = 1$ değerine sahiptir; dolayısıyla her ikisi de katkıda bulunur ve $r_i(z) = \gamma_1 \text{lot} + \gamma_2 h(\text{age} - 10)$, tek bir düğümden iki toplama terimi içerir; ortakların eklenmesi, aynı düğümden sınırsız sayıda yeni terim ekler. Tersini ele alalım: ortağı, lot < 6000 olan bir noktada bir menteşe $h(\text{lot} - 6000)$ olan bir terim ele alalım: o zaman $P_m(z) = 0$ olur ve $s_{m,i} \neq 0$ olmasına rağmen terim sıfıra iner; lot > 6000 olan bir noktada ise bu olmaz. ■

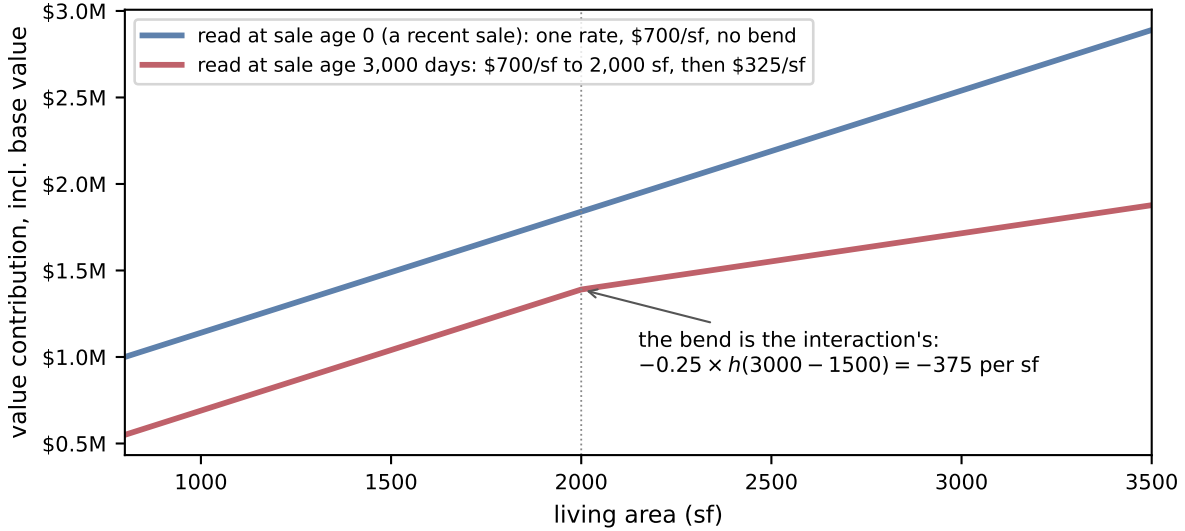
Burada, editörün kendi değerlemelerinden elde ettiği pratik bir gözlem yer almaktadır. Etkileşimlere izin verildiğinde, yazılımın değişken için tek başına çizdiği yaşam alanı grafiği genellikle düz, tek bir oran olarak ortaya çıkar; oysa etkileşim içermeyen bir modelin yaşam alanına ekleyeceği kıvrımlar, bunun yerine etkileşim terimlerinde, tipik olarak satış yaşı (satış ile yürürlük tarihi arasındaki süre) ile yaşam alanı veya banyo sayısı ile yaşam alanı gibi terimlerde görülür. Teorem 3, düz çizginin neden oran olmadığını açıklar. earthUI'deki değişken başına grafik, eşleşmesi olmayan terimleri çizer; buna karşılık plotmo'nun birinci derece grafiği, diğer değişkenleri medyan değerlerinde tutarak tüm modelin bir kesitini çizer. Etkileşim terimleri başka yerlerde yüzeyler olarak çizilir (plotmo'nun birinci ve ikinci dereceden grafikleri, Milborrow 2024b; earthUI bunları yan yana gösterir). Herhangi bir gerçek taşınmazda, yaşam alanı oranı, düz çizginin eğimi artı her aktif etkileşim terimi için katsayısının eş terim faktörünün değeriyle çarpımıdır; bu oran, söz konusu terimlerin yaşam alanı düğümlerinde değişir ve ne kadar değişeceği eş faktörlere bağlıdır: altı yıl önceki bir satış ile geçen ayki bir satış ya da iki banyolu bir ev ile dört banyolu bir ev, yaşam alanı grafiği tek bir çizgi gösterse de farklı yaşam alanı fiyat tablolarına tabidir. Oranı doğrusal bir çizgi olarak kabul eden bir okuyucu, bir üst seviyede yine engelle karşılaşır; konunun bulunduğu noktadaki oran tablosu bu engeli aşmanın yoludur. Şekil7, yazılımın yaptığı gibi, yaşam alanı ve satış yaşı üzerinde bir yüzey olarak böyle bir modeli çizer ve ardından iki farklı satış yaşı için yaşam alanı değişken grafiğini buradan okur.

Şekil 8, Şekil 2'deki piyasadan yola çıkarak aynı fikri ters yönde ele almaktadır. Bu pazar için, yaşam alanı ile satış yaşı arasındaki etkileşimi içeren bir model oluşturulmuştur; böylece yaşam alanının ana etkisi tek bir düz çizgi oluşturur (\$500/fit kare) ve dört etkileşim terimi eğrilikleri oluşturur. Yazılımın yaşam alanı için kendi başına çizdiği grafik, düz bir çizgidir. Dört yıllık bir satış verisine göre, yaşam alanı oranları dolarına kadar Şekil 2'dekilerle aynıdır: \$600, \$1,000, \$750

Living area by sale age: the surface the software draws



The living-area variable graph, read at two points



Şekil 7: Yaşam alanı ile satış süresi arasındaki etkileşim: $1,000,000 + 700 h(\text{gla} - 800) - 150 \text{ saleage} - 0.25 h(\text{gla} - 2000) h(\text{saleage} - 1500)$, 800 ila 3.500 fit kare aralığındaki veriler ve en fazla 3.000 gün önce gerçekleşmiş satışlar temel alınmıştır. Yukarıda, modelleme yazılımının bu ikili için çizdiği yüzey gösterilmektedir. Aşağıda, Teorem 3'ün gerektirdiği şekilde iki noktada okunan yaşam alanı değişkeni grafiği yer almaktadır: 1.500 güne kadar olan herhangi bir satışta ortak faktörü $h(\text{saleage} - 1500)$ sıfırdır, etkileşim hiçbir katkı sağlamaz ve grafik \$700 fit kare değerinde tek bir düz çizgidir; 3.000 gün öncesine ait bir satışta ortak değeri 1.500'dür; etkileşim, 2.000 fit kare üzerindeki alanlarda fit kare başına $0.25 \times 1,500 = 375$ değerini çıkarır ve buradaki oran \$325 'dir. Yaşam alanı ana etkisi (ortak olmadan, yalnızca yaşam alanındaki terimler) her ikisinde de aynı düz çizgidir; eğrilik etkileşime aittir ve yalnızca ortak aktif olduğunda ortaya çıkar.

ve \$100; yeni bir satış verisine göre ise düz bir çizgi izlemektedir; iki yıllık bir veriye göre ise bu iki değer arasında yer almaktadır. Bu model, bir noktayı vurgulamak amacıyla oluşturulmuştur; satış verilerine uyumlu değildir ve gerçek bir yenileme çalışması²'da gösterildiği kadar düzgün bir sonuç vermeyecektir; ancak işleyiş, gerçek bir yenileme çalışmasında kullanılan mekanizmayla aynıdır ve bu nedenle, etkileşimler eklendikten sonra yaşam alanı grafiğinin düzleştiğini gören bir değerlendirme uzmanı, eğrilikleri gözden kaçırmış sayılmaz. Bükülmeler etkileşim terimlerine taşınmıştır; burada yalnızca bir noktadaki okuma değeri sayesinde bunlar yeniden elde edilebilir. Etkileşimler içeren bir modelin ana etki grafiği, değişkenin terimlerini tek başına gösterir; bu, her etkileşim teriminin sıfır olduğu bir ortaklık ayarındaki kesiti (burada, yeni bir satış) temsil eder, pazarın şeklini değil; konuyla ilgili değişken grafiğine bakınız.

Son sonuç, oranla değil, düzeltmelerle ilgilidir ve satış tablosuna yansıtılan da budur.

Teorem 4 (bir düzeltmenin bölünmesi sıraya bağlıdır). Durum 2 bağlamında, $y = z_p$ ($p \geq 1$) diğer özelliklerden biri olsun ve w kalanları, z_k ise $k \notin \{0, p\}$ 'yi temsil etsin; argümanları bu şekilde gruplandırılmış f için $f(x, y, w)$ yazılsın. Bir emsal değeri (x_c, y_c, w) ve bir konu (x_s, y_s, w) olsun; her ikisi için de w aynıdır. y adresindeki bir değer y^* için, y adresinde y^* konumunda tutulan x adresine yönelik düzeltmeyi şu şekilde tanımlayalım:

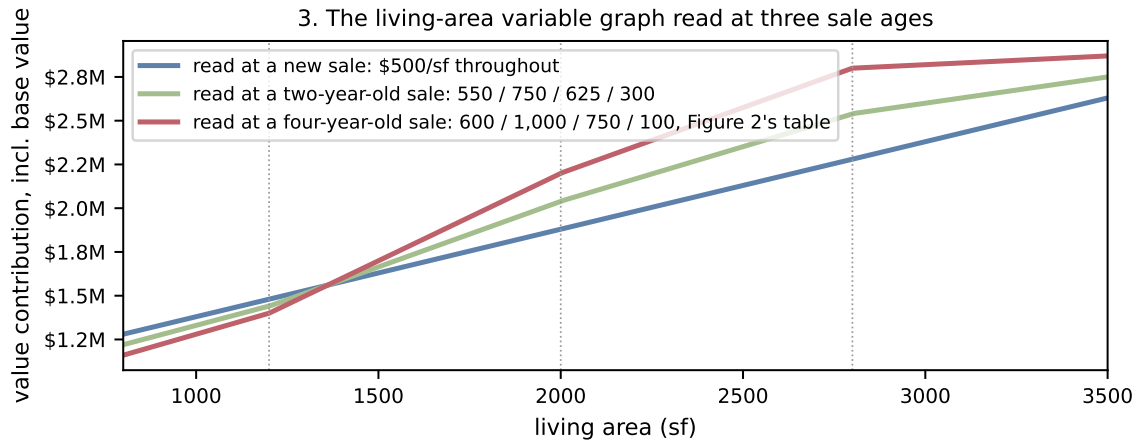
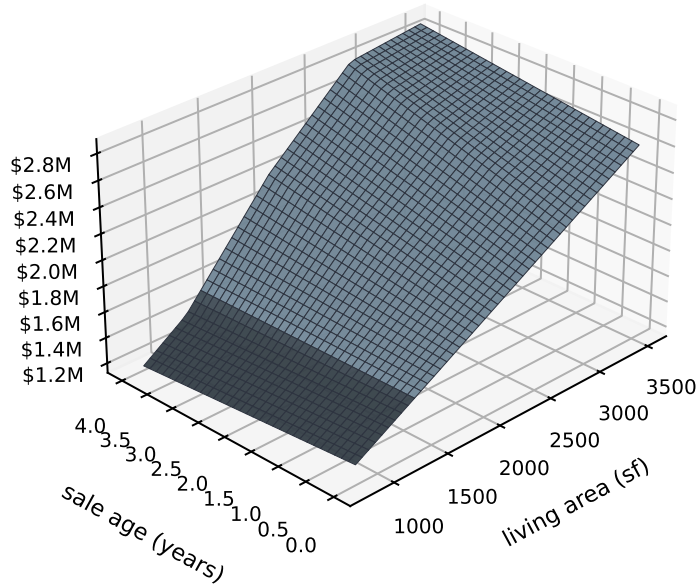
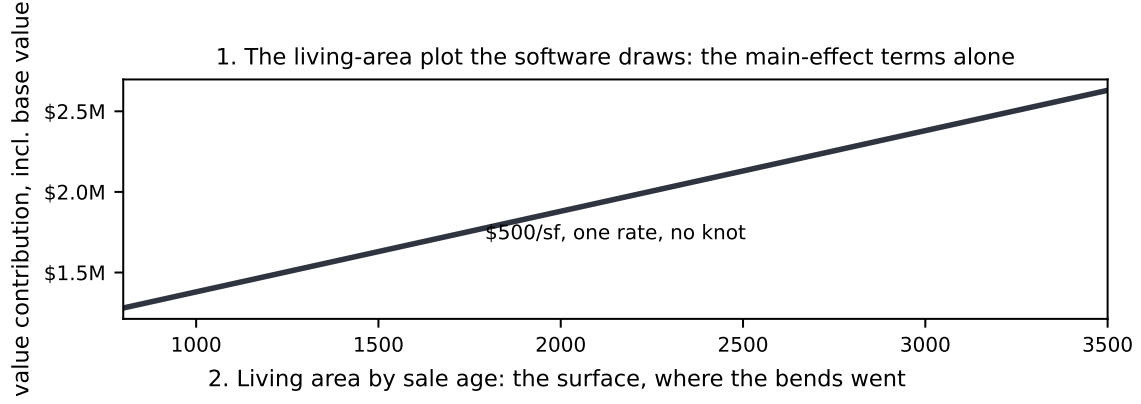
$$\Delta_x(y^*) = f(x_s, y^*, w) - f(x_c, y^*, w).$$

$M_{x,y}$, hem x hem de y 'yi içeren terimler olsun. Bu tür her terim, katsayısının üç parçayla çarpımıdır: $\gamma_m A_m(x) C_m(y) Q_m(w)$: $A_m(x) = B_{m,0}(x)$, x içindeki faktörüdür; bu, x 'in kendisi veya x içindeki bir menteşedir; $C_m(y) = B_{m,p}(y)$, y içindeki faktörüdür; bu, y 'in kendisi veya y içindeki bir menteşedir; ve $Q_m(w)$, kalan özelliklerdeki faktörlerinin çarpımıdır; bu özelliklerin her biri, terimin kendisi veya içindeki bir menteşedir ve terimin başka faktörü olmadığında 1 olur. Aşağıdaki örnekte bu tür bir terim $0.03 h(\text{gla} - 1200) h(\text{lot} - 6000)$ olup, $A_m(x) = h(\text{gla} - 1200)$, $C_m(y) = h(\text{lot} - 6000)$ ve $Q_m(w) = 1$ şeklindedir. " $\Delta_x(y_s)$ " ve " $\Delta_x(y_c)$ " değerlerini, " x " için düzeltme olarak yazınız; burada " y " değeri, incelenen mülk için " y_s " ve emsal mülk için " y_c " değerlerinde tutulmaktadır; yani " $\Delta_x(y^*)$ " değeri, " $y^* = y_s$ " ve " $y^* = y_c$ " değerlerinde tutulmaktadır; örnekte, " x " değeri yaşam alanı ve " y " değeri arsa alanı olup, " $\Delta_x(y_c) = \$240,000$ " değeri emsal mülkün 5.000 fit kare arsa alanı ve " $\Delta_x(y_s) = \$288,000$ " değeri incelenen mülkün 10.000 fit kare arsa alanı için geçerlidir. Ardından

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

Oysa $f(x_s, y_s, w) - f(x_c, y_c, w)$ adresindeki toplam fark, y^* adresindeki herhangi bir seçimi içermez. Bu ifade, iki özellikli bir dilimle ilgilidir; bir ortak y ve diğer her şey her iki özellik için de aynı w değerinde tutulur; ikisi arasında üç veya daha fazla özellik farklılık gösterdiğinde, toplamın bunlar arasında dağılımı, tek özellikli hareketlerden oluşan daha uzun bir zincir boyunca aynı şekilde yola bağlıdır ve teorem, henüz hareket ettirilmemiş özellikler mevcut değerlerinde tutularak zincirin her adımı için geçerlidir. Özellikle, tek bir $\gamma A(x) C(y)$ terimi varken ve w içinde hiçbir faktör bulunmadığında, x için yapılan iki ayarlama $\gamma [A(x_s) - A(x_c)] [C(y_s) - C(y_c)]$ kadar farklılık gösterir; bu fark, x içindeki hareket A 'i değiştirdiğinde ve iki özellik C açısından farklılık gösterdiğinde 0 değildir. $A = h(x - t)$ ve $C = y$ için bu fark $\gamma [h(x_s - t) - h(x_c - t)] (y_s - y_c)$ 'dir; $A = h(x - t)$ ve $C = h(y - s)$ için, aşağıdaki örnekte olduğu gibi, $\gamma [h(x_s - t) - h(x_c - t)] [h(y_s - s) - h(y_c - s)]$ 'dir.

Teorem 4'ün Kanıtı. f terimlerini, içerdikleri x ve y öğelerine göre gruplandırın. Ne x ne de y içeren bir terim, (x_s, y^*, w) ve (x_c, y^*, w) adreslerinde aynı değeri alır ve $\Delta_x(y^*)$ adresinde birbirini ortadan kaldırır. y içeren ancak x içermeyen bir terim de benzer şekilde birbirini ortadan kaldırır. x içeren ancak y içermeyen bir terim, $\Delta_x(y^*)$ adresine y^* 'a bağlı olmayan bir miktar katkıda bulunur; dolayısıyla $\Delta_x(y_s) - \Delta_x(y_c)$ adresinde birbirini ortadan kaldırır. $m \in M_{x,y}$ terimi, $\Delta_x(y^*)$ 'a



Şekil 8: Şekil 2'deki pazar, yaşam alanı ile satış yaşı etkileşimine göre: $1,280,000 + 500 h(gla - 800) - 20,000 \text{ saleyrs} + [-25 h(1200 - gla) + 125 h(gla - 1200) - 62.5 h(gla - 2000) - 162.5 h(gla - 2800)] \times \text{saleyrs}$; satış yaşı 0 ila 4 yıl arasındadır. Paneller başlıklarında açıklanmıştır; dört yıllık dönemde alttaki grafikteki oranlar, Şekil 2, \$120,000 'deki oranların tam olarak aynısıdır, ancak seviye olarak daha düşüktür. Gösterim amacıyla oluşturulmuştur, uyumlandırılmamıştır.

$\gamma_m[A_m(x_s) - A_m(x_c)]C_m(y^*)Q_m(w)$ katkıda bulunur; $y^* = y_s$ ve $y^* = y_c$ değerlerini belirleyip çıkarma işlemi yapıldığında belirtilen toplam elde edilir. Toplam fark, f fonksiyonunun iki değeri arasındaki farktır ve y^* 'dan bahsetmez. ■

Dolayısıyla, etkileşimin değeri özelliklerden herhangi birine değil, özellik çiftine aittir. Emsal nesnenin değerlerine göre her seferinde tek bir özelliği düzelten bir tablo ile, konu değeri olan nesnenin değerlerine göre düzelten bir tablo, toplam değer konusunda hemfikir olurken, dağılım konusunda fikir ayrılığına düşer. Diğer özellikleri incelenen mülkün konu değerlerinde sabit tutmak savunulabilir bir kuraldır ve programın kullandığı yöntem budur; bir raporda hangi kuralın kullanıldığı belirtilmelidir. Uniform Residential Appraisal Report (URAR) gibi standart bir konut tablosunda etkileşim için bir satır bulunmaz. Her seferinde tek bir özellik üzerinde yapıldığında, bu iki işlem eşleri zıt uçlarda tutmalıdır — yaşam alanı düzeltmesi konu nesnesinin arsa değerine göre yapılır, ardından arsa emsal nesnenin yaşam alanına göre yapılır ya da tersi — böylece etkileşim bir kez sayılır ve satırların toplamı modelin toplam farkına eşit olur. Katkılar doğrudan hesaplandığında, bir tablo etkileşimi her bir özelliğe ait satırların yanında, konu ve emsal mülk arasındaki etkileşim terimlerinin farkı olarak ayrı bir satırda da gösterebilir; bu durumda her satır bir terim grubu olur ve herhangi bir sıra seçilmesi gerekmez. Aşağıdaki örnekte üç satır şunlardır: yaşam alanı \$240,000, arsa \$75,000 ve arsa bazında yaşam alanı \$120,000; bunların toplamı aynı değere, yani \$435,000'a eşittir. Etkileşimler söz konusu olduğunda, çiftin sekant oranı iki şekilde değişir: birincisi kesişen aralıklar üzerinden, ikincisi ise çiftin ortaklar içindeki konumuna göre.

Etkileşimli çözümlü bir örnek

Yine bir sahil pazarı; bu sefer iki özellik var: yaşam alanı ve arsa büyüklüğü; ve bunlar arasında tek bir etkileşim. earth prints:

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

Teorem 3'te belirtildiği gibi okuyun. Yaşam alanı, 1.200 ve 2.400 olmak üzere iki düğüme sahiptir ve M_{gla} içinde üç terim bulunur: iki düz menteşe ve eş $h(\text{lot} - 6000)$ olan etkileşim. 1.200 ile 2.400 arasındaki segmentte, 1.200'deki basit menteşe aktiftir ve etkileşimin 1.200'deki menteşesi de aktiftir, bu nedenle buradaki oran $600 + 0.03 h(\text{lot} - 6000)$ şeklindedir: arsa büyüklüğüne göre değişken bir grafik olup, 6.000 fit kareye kadar olan arsalar için 600 sabit kalır ve her bir fit kare artış için 3 sent artar. 2.400'ün üzerinde, 2.400'deki eşik devreye girer ve oran $250 + 0.03 h(\text{lot} - 6000)$ olur. Arsanın kendi oranı, 6.000 fit kare üzerindeki arsa alanları için $15 + 0.03 h(\text{gla} - 1200)$ ve 6.000 fit kare altındaki arsa alanları için 15 şeklindedir; dolayısıyla her bir özelliğin oranı, diğerinin miktarına bağlıdır.

İki nokta, iki tablo. Programın, 800 ila 3.500 fit kare aralığındaki yaşam alanı için hazırladığı ücret tablosu; biri 5.000 fit kare arsa, diğeri ise 10.000 fit kare arsa üzerinden hesaplanmıştır. Her sütun, arsa büyüklüğünün sadece o değerde tutulduğu ve başka hiçbir değerde tutulmadığı durumlarda, modelin yaşam alanı eğimini göstermektedir:

yaşam alanı bölümü (sf)	5.000 lotluk fiyat (\$/sf)	10.000 lotluk fiyat (\$/sf)
800 ila 1.200	0	0
1.200 ila 2.400	600	720
2.400 ila 3.500	250	370

Daha küçük partide ortak faktör $h(5000 - 6000)$ sıfırdır; dolayısıyla etkileşim hiçbir katkı sağlamaz ve tablo, etkileşimin olmadığı bir modelin vereceği tablo ile aynıdır; bu, Sonuç 6'nın "terim sıfırlanabilir" durumudur. Daha büyük partide ortak 4,000 şeklindedir; etkileşim, 1.200'ün üzerindeki her segmente $0.03 \times 4,000 = 120$ ekler ve aynı düğüm (1.200) iki aktif terime beslenir; bu da "düğüm başına bir" sınırının ihlali anlamına gelir.

Şimdi düzeltmeye geçelim. Emsal nesne, 5.000 fit kare büyüklüğündeki bir arsa üzerinde 1.800 fit kare alana sahip olup, belirlenen değeri \$1,335,000'dir; konu ise 10.000 fit kare arsa üzerinde 2.200 fit kare alana sahip ve değerlemesi \$1,770,000 olarak belirlenmiştir. Açıklanması gereken toplam tutar \$435,000'dır. Önce, emsal arsasının arsasına göre yaşam alanını, ardından da söz konusu nesnenin yaşam alanına göre arsayı düzeltin. Her düzeltme, karşılaştırılan iki değer farkıdır; düzeltme tek bir segment içinde kalıyorsa, bu fark genişlik çarpı orandır; tabloda da bu şekilde yazılmıştır:

	önce yaşam alanı, sonra arsa	önce arsa, sonra yaşam alanı
yaşam alanı düzeltmeleri	$400 \times 600 = 240,000$	$400 \times 720 = 288,000$
arsa düzeltmeleri	195,000	147,000
toplam	435,000	435,000

Arsa düzeltmeleri de aynı şekilde katkı paylarındaki farklardır ve oranlar bunları açıklamaktadır; arsanın kendi oranı 6.000 fit kare seviyesinde değişir: söz konusu arsanın 2.200 fit kare alanına göre arsa oranı, eşik değerinin altında \$15 ve üzerinde \$45 olduğundan, 5.000'den 10.000'e geçiş $1,000 \times 15 + 4,000 \times 45 = 195,000$ olur, 5,000 $\times$ 45 değil; emsal mülkün 1.800 fit kare büyüklüğünde ise bu değer $1,000 \times 15 + 4,000 \times 33 = 147,000$ 'dir. Toplamlar, Teorem 4'ün belirttiği gibi birbirleriyle uyumludur ve bölünmeler \$48,000 kadar farklılık gösterir; bu da tam olarak teoremin formülüdür: $0.03 \times [h(2200 - 1200) - h(1800 - 1200)] \times [h(10000 - 6000) - h(5000 - 6000)] = 0.03 \times 400 \times 4,000$. Bu \$48,000 değeri, evin fazladan 400 fit kare ve arsanın fazladan 4.000 fit karesinin bir arada bulunmasının değeridir; bu değer her iki özelliğe de ait değildir ve hangisi ikinci olarak düzeltilirse o, bu değeri alır. Programın kuralına göre, diğer özellikler konunun değerlerinde tutulduğunda, sağdaki sütunda yaşam alanı değeri \$288,000 (\$720) olarak bir fit kare olarak verilir ve raporda oranın hangi noktada okunduğu belirtilir:

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.

Her iki katkı da, söz konusu mülkün 10,000 fit kare büyüklüğündeki arsası üzerinden hesaplanmıştır; emsal nesnesinin 5,000 fit kare büyüklüğündeki kendi arsasında ise yaşam alanı katkısı \$360,000'dir.

Bunu emsal nesnenin park yerinde yeniden çalıştıran bir denetçi, \$240,000 sonucunu alır ve bu yanlış değildir; denetçi ve program, aynı yüzeyi iki farklı noktada okumaktadır ve raporun hangisi olduğunu belirtmesi gerekir.

Kısacası: Etkileşimler olmadan bir MARS modeli, her bir özellik için ayrı bir oran tablosudur ve sıradan regresyon için yapılan oranlarla ilgili tüm argümanlar, segment segment aktarılır. Etkileşimler olduğunda ise bir oran bir yüzeydir; Teorem 3, bu yüzeyin tam olarak nasıl eğildiğini belirtir ve değerlendirme uzmanları, hangi noktada okunduğunu belirtmelidir.

Program

Ek bölümde tam metni yer alan `mars_rates.pl` programı, “earth” adlı basılı tabloyu ya da bir Prolog terimi olarak yazılmış bir modeli okur ve oran tablosunu, düğüm noktalarındaki değerleri, iki katkıyı ve bunların farkını, aralık bazında açıklamalarıyla birlikte düzeltmeyi, çiftin sekant oranını ve aralıkların toplamının düzeltmeye eşit olup olmadığını kontrol eder. Veri aralığı ve satışlardaki en küçük ve en büyük ev göz önüne alındığında, bu program bu aralığın dışındaki oranları raporlamaz: bir menteşenin cebiri bir tarafta sonsuza kadar uzanır, satışlar ise uzanmaz ve bunların ötesindeki bir oran ekstrapolasyondur. Örnek olarak, 800 ila 3.500 fit kare arasındaki veriler üzerinde:

```
Feature: gla
Rates by range (data from 800 to 3500; nothing is reported outside it):
  800 .. 1200   rate 600
  1200 .. 2000   rate 1000
  2000 .. 2800   rate 750
  2800 .. 3500   rate 100
Fitted value at the knots (intercept and partners included):
  1200   1520000
  2000   2320000
  2800   2920000
Comparable at 1800: contribution 600000
Subject at 2400:    contribution 1100000
Adjustment, range by range:
  1800 .. 2000   200 at 1000 = 200000
  2000 .. 2400   400 at 750 = 300000
Adjustment added to the comparable: 500000
Secant rate for the pair (one straight-line rate): 833.3333333333334
Check: segments sum to the adjustment.
```

Bu blokta iki tür sayı yer almaktadır. Katkılar, yalnızca yaşam alanı terimlerinden oluşmaktadır; kesim noktası ve diğer tüm özelliklerin terimleri hariç tutulmuştur: \$600,000 ve \$1,100,000, tam olarak düzeltme farkı kadar birbirinden farklıdır. Düğüm noktalarındaki değerler, kesim noktası dahil olmak üzere uyum içinde olan değerlerdir ve etkileşimlerin bulunduğu bir modelde, o noktadaki ortak terimler de buna dahildir; burada 1.200 fit kare'deki \$1,520,000 değeri, her bir menteşenin orada sıfır olması nedeniyle kesim noktasıdır. Karşılaştırılabilir ve denek noktalarındaki uyum değerleri, \$2,120,000 ve \$2,620,000, aynı şekilde kesme noktasını taşır. Rapor, düğüm değerlerini uyum değerleri olarak etiketler; böylece okuyucu bunları katkı olarak algılamaz.

O bölüm, bir raporun ekinde yer alması gereken kısımdır: oranlar, düğümler ve bunlardan birinden diğerine yapılan hesaplamalar; bir gözden geçirenin elle yeniden hesaplayabileceği bir biçimde. Programın 69 testi, işaret kuralını, etkileşim kuralını, düzeltmelerle segmentlerin tutarlılığını ve

Earth çıktısı için ayrıştırıcıyı doğrular ve Trealla, Journal'ın Prolog'u, SWI-Prolog ve Scyer Prolog altında başarılı olur. R'da çalışan okuyucular için, programdan ayrılmadan aynı içeriğe erişilebilir: Programla birlikte yayınlanan `mars_rates.R` dosyası, bir `earth` nesnesini veya bunun yazdırılmış tablosunu alır ve oranları veri aralığına sabitleyerek aynı bloğu yazdırır; Prolog, doğrulanmış referans olarak kalır ve R dosyası, bu ekin örneklerine göre kendi kendini kontrol eder. Program, testleri ve şekil komut dosyası, bu sayıyla birlikte PDF dosyasının yanında, GNU Affero Genel Kamu Lisansı, sürüm 3 veya üstü kapsamında indirilebilir olarak yayınlanmaktadır.

Neden MARS, ve bu ekin başka maddelere bıraktıkları

Dergi, üç nedenden ötürü MARS ile modelleme yapmaktadır ve bu ek sayının var olabilmemesinin nedeni de budur. Nesnellik: düğümler, satışların yerleştiği yere, belirtilen bir budama kuralına sahip tek bir prosedürle yerleştirilir (yerini hak etmeyen terimleri ortadan kaldıran sonraki aşama) ve belirtilen ayarların (maksimum etkileşim derecesi, budama cezası, düğümler arasındaki minimum aralık) ötesinde değişken bir grafiğin şekliyle ilgili hiçbir şey önceden belirlenmez. Keşfedilebilirlik: İleri geçiş, her özelliği ve her aday düğümünü, belirtilen dereceye kadar dahil edilen etkileşimleri arar ve bulduklarını raporlar; bu, seçilmiş bir şekilden başlayan hiçbir yöntemin yapamayacağı bir şeydir. Yorumlanabilirlik: Verilerin seçtiği aralıklarda sabit oranlar ve aynı bağlantı noktalarından oluşturulan etkileşimler, modelin bir oran tablosu olarak yazılmasını ve bir düzeltmenin “oran çarpı genişlik” şeklinde açıklanmasını sağlar. Diğer iki yöntem de bir özelliğin aralığı boyunca değişen oranlar üretir ve bunlar birbirinden farklıdır. Birincisi, bölümlere ayrılmış (“kırık çubuk”) regresyon, bir değişkeni ve analist tarafından seçilen bir dizi kesme noktasını alır ve kesme noktalarının nereye düştüğünü tahmin eder; her biri için bir standart hata (tahminin diğer satışlarla birlikte ne kadar değişebileceğinin bir ölçüsü) bulunur; kesme noktaları arasında oran, MARS'ta olduğu gibi sabittir. İkincisi olan şekil kısıtlanmalı eklemeli model ise, düz segmentler yerine düzgün bir eğriye uyum sağlar ve analistin dayattığı bir şekle zorlanır — örneğin, yaşam alanı büyüdükçe değerin asla düşmemesi ya da eklenen her bir fit kare'nin bir öncekinden daha az artış sağlaması gibi. Burada ikisi de kullanılmamaktadır. Her ikisi de bu ekte sorulmayan bir soruyu yanıtlar ve her ikisi de üç taşınmazdan en az birini feda eder: kesimli regresyon, değişken ve kesme noktalarının sayısı önceden seçildiği için keşfedilebilirlikten ödün verir; şekil kısıtlanmalı model ise yorumlanabilirlikten ödün verir, çünkü bir eğrinin tabloya dökülebilecek sabit oranları yoktur; ayrıca nesnellikten de ödün verir, çünkü şekli dayatılmıştır. Bu ek, MARS'ın ürettiği modeli inceler; yöntem seçimi ile MARS'ın bir satış kümesinden diğerine bulduğu düğümlerin ve oranların istikrarı ise ayrı başlıklar altında ele alınacak konulardır.

Bu ekin çözmediği konular

Oranlar modele aittir ve dört istatistiksel ilke bunlara aynen uygulanır: bir oran, koşullu beklentiye yalnızca modelin yapısının izin verdiği ölçüde yaklaşıtır (*Koşullu Beklenti*); ihmal edilen bir özelliğin değerini yansıtabilir (*İhmal Edilen Değişken Önyargısı*); bir özellik yapısal hata ile birlikte hareket ettiğinde oran çarpıtılabilir (*Endojenlik*); ve iki özellik her zaman birlikte hareket ettiğinde, satışlar ne kadar çok olursa olsun, oran satışlar tarafından kesin olarak belirlenemeyebilir (*Parametre Tanımlama*). MARS, bir oranın yalnızca satışların bunu gösterecek bir kontrast içerdiği durumlarda değiştiğini tespit eder; bir düğüm, piyasa hakkında bir iddiadır ve yalnızca her iki tarafındaki satışlar kadar geçerlidir. Bu eklemenin kattığı şey daha dar kapsamlıdır: model göz önüne alındığında, modelin ne söylediğini nasıl ifade edeceğimiz budur.

References.

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

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

Addendum: Prolog Formalization

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

```
% SPDX-License-Identifier: AGPL-3.0-or-later
% Copyright (C) 2026 William Bert Craytor. Published by The Valuation Engineer
% Journal (Pacific Vista Net). This program is free software: you can
% redistribute it and/or modify it under the terms of the GNU Affero General
% Public License as published by the Free Software Foundation, either version 3
% of the License, or (at your option) any later version. It is distributed
% WITHOUT ANY WARRANTY, without even the implied warranty of MERCHANTABILITY or
% FITNESS FOR A PARTICULAR PURPOSE; see the COPYING file beside it or
% <https://www.gnu.org/licenses/agpl-3.0.html>.

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

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

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

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

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

Representation. A model is

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

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

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

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

```
linear(Var)           Var
```

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

Entry points:

```
value(+Model, +Point, -Y)           the fitted value at a point
contribution(+Model, +Var, +Point, -C) the part of the value that the
                                     terms involving Var contribute
rate(+Model, +Var, +Point, -R)       slope in Var at the point (other
                                     variables held at the point)
knots(+Model, +Var, -Knots)          the sorted knots on Var
rate_table(+Model, +Var, +Point, -Rows)
                                     rows range(Lo, Hi, Rate), Lo/Hi
                                     numbers or the atoms minus_inf,
                                     plus_inf; other variables fixed
                                     at the point
rate_table(+Model, +Var, +Point, +data(Lo, Hi), -Rows)
                                     the same table clamped to the
                                     data range: the first row starts
                                     at Lo, the last ends at Hi, rows
                                     outside the data are dropped.
                                     Rates outside the data are
                                     extrapolation and are not
                                     reported
knot_values(+Model, +Var, +Point, -Pairs)
                                     Knot-Value at each knot
adjustment(+Model, +Var, +Point, +CompQ, +SubjQ, -A)
                                     amount added to the comparable's
                                     contribution (Var = CompQ) to
                                     equal the subject's (Var = SubjQ)
segments(+Model, +Var, +Point, +CompQ, +SubjQ, -Segs)
                                     the adjustment as the appraiser
                                     explains it: seg(Lo, Hi, Rate,
                                     Width, Amount) per range crossed
pair_rate(+Model, +Var, +Point, +CompQ, +SubjQ, -R)
                                     the pair's secant rate: the one rate a straight line
↳ would
                                     need between the two quantities:
                                     adjustment over width
consistent(+Model, +Var, +Point, +CompQ, +SubjQ)
                                     the segments sum to the adjustment
parse_earth(+Atom, -Model)          read earth's printed coefficient
                                     table (see parse_earth below)
print_report(+Model, +Var, +Point, +CompQ, +SubjQ)
```

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

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

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

```

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

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

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

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

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

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

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

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

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

/* d/dVar of a term = Coef * (slope of the factor in Var) * (product of the
   other factors' values). A hinge is active strictly above its knot, a
   reflected hinge strictly below; at the knot itself the slope is taken from
   the right, which is the convention of the rate table below. */

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

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

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

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

```

```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

/* ===== */

```

```

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

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

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

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

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

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

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

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

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

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

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

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

```

```

dots and underscores, as R prints them.                                */

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

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

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

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

/* the coefficient is the LAST whitespace-separated token; the name is what
   precedes it and may itself contain spaces (earth prints h(x- -3) for a
   negative knot)                                                         */
split_ws(S, Name, Coef) :-
  mr_reverse(S, R), split_ws_(R, [], Coef, NameR),    % consuming the reversed line puts the
  ↪ coefficient back in order
  mr_reverse(NameR, Name0), strip(Name0, Name).
split_ws_([C | Cs], Acc, Coef, Rest) :- ws(C), !, Coef = Acc, Rest = Cs.
split_ws_([C | Cs], Acc, Coef, Rest) :- split_ws_(Cs, [C | Acc], Coef, Rest).
split_ws_([], Acc, Acc, []).

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

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

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

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

```

```

parse_factor(Cs, linear(Var)) :- atom_chars(Var, Cs).

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

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

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

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

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

print_report(Model, Var, Point, data(DLo, DHi), CompQ, SubjQ) :-
    write('Feature: '), write(Var), nl,
    write('Rates by range (data from '), write(DLo), write(' to '), write(DHi), write('; nothing is
    ↪ reported outside it):'), nl,
    rate_table(Model, Var, Point, data(DLo, DHi), Rows),
    print_report_(Model, Var, Point, CompQ, SubjQ, Rows).

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

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

```

```

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

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

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

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

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

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

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

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

```