

Koşullu beklenti (Conditional Expectation)

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

DOI: [10.67330/99hhbc69](https://doi.org/10.67330/99hhbc69)

Biçimsel tanım. Bir dizi öngörücü değişken (ölçülen özellikler) göz önüne alındığında, bir sonucun (burada satış fiyatı) koşullu beklentisi, öngörücü değişkenlerin her bir sabit değeri için sonucun dağılımının ortalamasıdır. Öngörücülerin bir fonksiyonu olarak ele alındığında buna koşullu beklenen değer fonksiyonu (CEF) denir ve regresyonun yaklaştırmayı amaçladığı fonksiyon budur. Sıradan en küçük kareler (OLS) regresyonu, CEF'e en iyi doğrusal yaklaştırmayı sağlar ve CEF'in kendisi doğrusal olduğunda CEF'i tam olarak yeniden üretir.

Sezgisel çerçeve. Regresyondan elde edilen bir düzeltme, bir oranla başlar: Piyasanın fit kare başına, banyo başına ve yaş başına ödediği tutar. Bu yazı, bu oranın nereden geldiği ve neye ilişkin olduğu hakkında. Kısaca cevap şudur: Oran, bir özellik değişirken diğerlerinin sabit tutulduğu durumda piyasanın ortalama fiyatının eğimidir ve regresyon, satış verilerinden bu eğimi ölçmenin bir yoludur. OLS regresyonu, özelliğin tüm aralığı için tek bir eğimi ölçer; MARS (Çok Değişkenli Uyarlanabilir Regresyon Spline'ları, Friedman 1991) ise piyasanın farklı fiyatlar ödediği her aralık için bir eğim ölçer ve aralık başına bir oran rapor eder. Her iki durumda da oran, aynı şeyin ölçümüdür: ortalama fiyatın o özellik ile birlikte nasıl değiştiği. Bu yazının geri kalanı da bu konu hakkındadır: bunun ne olduğu, satış karşılaştırma tablosunun da bunu neden ölçtüğü ve aynı satışlara ilişkin iki regresyonun iki farklı oran vermesi ne anlama geldiği.

Mekanizmayı öğrenmek isteyen okuyucular için: “Özellikler değiştikçe ortalama fiyat”ın bir adı vardır: koşullu beklenti fonksiyonu; ve bu yazıda bundan sonra bu terim kullanılacaktır. Bu, özellikler uzayı üzerinde bir yüzeydir; bir oran, bu yüzeyin bir yöndeki eğimidir; OLS, yüzeye en iyi şekilde yaklaşan düzlemi uyarlar ve yalnızca yüzey bir düzlem olduğunda onu tam olarak geri kazanır; MARS ise, bulduğu kırılma noktalarında bükülen düz parçalardan oluşan bir yüzeyi uyarlar ve bu sayede gerçek yüzey büküldüğünde ona daha çok yaklaşır. 1. Sayı'daki sekiz Pacifica emsal, bu konuyu açıklayan canlı bir örnektir.

Bu yazı hukuktan ziyade istatistikle ilgilidir: regresyonun neyi tahmin edebileceğini ele almaktadır ve regresyon analizleri yapan ya da aynı şeyi tahmin eden satış karşılaştırma tablosunu kullanan değerlendirme uzmanlarına yöneliktir.

1. Sayı'da Pacifica emsalleri için gerçekleştirilen regresyon, katsayılarından (regresyonun özelliklere atadığı oranlar) örtük fiyatları türetmek amacıyla kullanıldı; örtük fiyat, piyasanın bir özelliğin bir birimi için ödediği dolar tutarıdır. Ancak, regresyonun neyi tahmin ettiği açıkça belirtilmemiştir ve bunu belirtmekte fayda vardır; zira bu sayıdaki takip eden soruların neredeyse tamamı — önyargı, içsellik ve tanımlama ile ilgili olanlar — tahmin edilen ile istenen arasındaki farkla ilgili sorulardır.

Regresyon genellikle, artıkların karelerinin toplamını en aza indirerek bir doğruya uyum sağlama (artık, satış fiyatından uydurulan denklemin verdiği fiyatın çıkarılmasıyla elde edilir) ve ardından eğimi belirleme şeklinde açıklanır. Bu açıklama, işin aritmetik yönü açısından doğru olsa da, amaç söz konusu olduğunda yanıltıcıdır. Aritmetik işlemler yalnızca bir araçtır; asıl amaç, sadece eldeki satışları değil, tüm pazarı (bir *popülasyon* büyüklüğü) tanımlayan bir büyüklüktür.

Notasyonu okumak.

P satış fiyatı.

$\mathbf{Z}$	bir taşınmazın ölçülen özelliklerinin listesi; bu tür bir listeye <i>vektör</i> denir.
$\mathbf{z}$	bu liste için belirli bir değer grubu; örneğin 1.650 fit kare, 7.200 fit kare büyüklüğünde arsa, manzara yok, durum 7,00.
$E[\]$	<i>beklenen değer</i> : her birinin olasılığıyla ağırlıklandırılmış tüm olası değerlerin ortalaması.
$ $	“verilmiş”.
$E[P \mid \mathbf{Z} = \mathbf{z}]$	$\mathbf{z}$ adresindeki taşınmazlar arasında fiyatın beklenen değeri.
ε	bir satışın bu beklenen değerden sapması.
u	Bir satışın yapısal hatası: satış fiyatı ile nedensel ilişkinin belirlediği olması gereken fiyat arasındaki fark.
$\hat{p}$	uyum denkleminin verdiği değer.
$z_{\text{GLA}}, \dots$	bir özelliğin değeri: GLA ve arsa alanı (fit kare cinsinden); manzara ve tadilat: evet ise 1, hayır ise 0; puan aralığı 0,00–10,00.

$\mathbf{z}$ nitelik vektörünü sabitleyin — yani 1.650 fit kare, 7.200 fit kare arsa, manzara yok ve “iyi” durum (Journal’ın durum ölçeğinde 7,00; ayrıntılı örnekte açıklanmıştır). Piyasada tam olarak bu nitelik vektörüne sahip taşınmazlar arasında fiyatlar farklılık gösterir. Fiyatlar aynı değildir çünkü taşınmazlar vektörün yansıtamadığı yönlerden farklılık gösterir, alıcılar ve satıcılar farklıdır ve pazarlık her zaman aynı fiyatla sonuçlanmaz. Bu piyasada, $\mathbf{z}$ değerine karşılık gelen tam bir fiyat dağılımı mevcuttur. Bu dağılımın ortalaması, $\mathbf{z}$ koşulunda fiyatın koşullu beklentisi olup, $E[P \mid \mathbf{Z} = \mathbf{z}]$ şeklinde yazılır; genel tanım Kolmogorov’a (1933) aittir. Elimizdeki satışların ortalaması, bu beklenen değer kendisi değil, onun bir tahminidir.

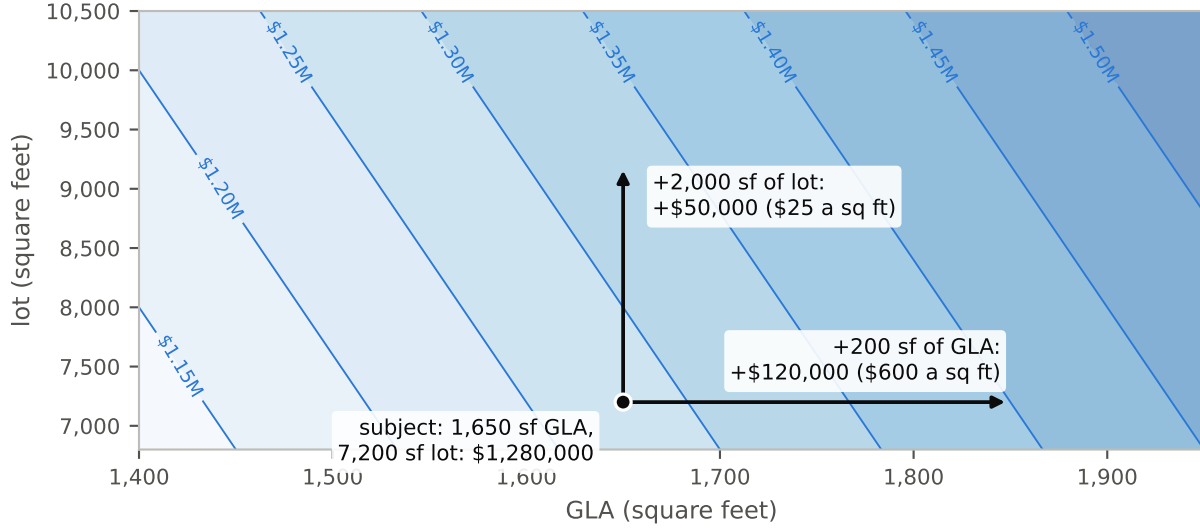
Taşınmazların nitelikleri tüm nitelikler uzayını kapsadığından, koşullu beklenti bir yüzey oluşur: her noktada bir sayı bulunur; bu sayı, söz konusu niteliklere sahip taşınmazların ortalama fiyatını ifade eder. Bu yüzey, *koşullu beklenti fonksiyonu* (veya CEF) olarak bilinir. Bu, piyasanın bir özelliğidir ve belirli bir veri kümesine ait bir şey değildir — hiçbir taşınmaz değerlemeye tabi tutulmamış olsa bile yine de var olacaktı. Şekil1, örnek üzerinde bu yüzeyi göstermektedir.

İki gerçek, CEF’yi tahmin için doğal bir hedef haline getirir. Birincisi, $\mathbf{z}$ adresindeki tüm fonksiyonlar arasında CEF, tahmin hatasının karesinin beklenen değerini en aza indiren fonksiyondur: Yalnızca $\mathbf{z}$ adresindeki verilere dayanarak bir fiyat tahmini yapmanız gerektiğinde ve performansınız hata karesi üzerinden değerlendiriliyorsa, $E[P \mid \mathbf{Z} = \mathbf{z}]$ adresindeki tahmin, yapabileceğiniz en iyi tahmindir. İkincisi, herhangi bir sonuç şu şekilde ifade edilebilir:

$$P = E[P \mid \mathbf{Z}] + \varepsilon, \quad E[\varepsilon \mid \mathbf{Z}] = 0,$$

. İkinci denklem bir varsayım değil, bir özdeşliktir; zira koşullu ortalama sapmanın koşullu ortalaması, tanımı gereği sıfırdır (Angrist& Pischke 2009, bkz. 3. bölüm).

Bu madde, 008 (*Atlanan Değişken Önyargısı*) ile 010 (*Parametre Tanımlama*) numaralı maddelerden önce yer almaktadır; zira bu maddelerin her biri, regresyonun ortaya çıkardığı sonucun ilgilenilen CEF olup olmadığını sorgulamaktadır. $E[\varepsilon \mid \mathbf{Z}] = 0$ kimliği her zaman geçerlidir; çünkü ε , koşullu ortalama uzaklık olarak tanımlanmıştır. Başarısız olabilecek olan ise, farklı bir hata için aynı koşul, yani sıfır koşullu ortalama: 009 numaralı maddede (*Endojenlik*) u şeklinde ifade edilen yapısal hata — fiyat ile değerlendirme uzmanının ölçmek istediği nedensel ilişki arasındaki fark. Bu hata, özelliklerle birlikte değişebilir. Bu durum, örneğin, fiyatı etkileyen bir özelliğin dışarıda bırakılması durumunda ortaya çıkar; fiyatın bir özelliğe geri etki etmesi durumunda — piyasada kalma süresi (DOM) genellikle fiyat arttıkça artar, bu nedenle fiyatı açıklamak için DOM’u kullanan bir regresyon, fiyatın DOM üzerindeki kendi etkisinin bir kısmını da yakalar; bu yüzden çok az sayıda analist, net satış fiyatının regresyonuna DOM’u dahil eder; ya da bir özellik hatalı bir şekilde ölçüldüğünde. *Atlanan Değişken Önyargısı* (giriş 008) ve *Endojenlik* (giriş 009) bu durumlarla ilgilidir; *Parametre Tanımlama* (giriş 010) ise daha öncelikli bir soruyu ele alır: Satış verileri bu oranı gösterebilir mi?



Şekil 1: Oluşturulan Pacifica piyasasının (örnek hesaplama) koşullu beklenti fonksiyonu; manzarasız, durum puanı 7,00 olan ve tadilatsız konutlar için, topografik harita şeklinde çizilmiştir: her bir çizgi, ortalama fiyatı aynı olan brüt kiralanabilir alan (GLA) ve arsa kombinasyonlarını birleştirir; ortalama fiyat arttıkça gölgelendirme koyulaşır. İşaretili nokta, bu girdinin evidir: \$1,280,000 adresinde, 7,200 fit kare arsa üzerinde 1,650 fit kare GLA'ya sahip ev. İki ok, oranları gösteren eğimdir: 200 fit kare daha fazla GLA, ortalama fiyatı \$120,000 artırır; yani fit kare başına \$600 ; 2,000 fit kare daha fazla arsa, fiyatı \$50,000 artırır; yani fit kare başına \$25 .

Değerleme uzmanı nerede karşılaşır. “Bu bölgede, 1,650 fit kare alana sahip ve manzarası olmayan bir evin yaklaşık \$1.28 million’a satıldığını” belirten değerleme uzmanları, koşullu bir beklenti ortaya koymaktadır. Burada kullanılan kelime “yaklaşık”tır. Bu ifade, bu tür her evin \$1.28 million’a satıldığını iddia etmez; aksine, \$1.28 million’ın bu tür evlerin satış fiyatlarının ortalaması olduğunu ifade eder. Değerleme uzmanı, “yaklaşık” ifadesini kullanarak, bireysel fiyatların bu değer etrafında dağıldığını kabul eder.

Satış karşılaştırma tablosu, adresinde de belirtildiği gibi, incelenen nesnenin karakteristik vektöründeki CEF değerini tahmin etmeye yarayan bir yöntemdir; her bir emsal, o nesnenin karakteristik vektöründeki fiyat dağılımından alınan bir gözlemdir; her bir düzeltme, söz konusu gözlemi CEF yüzeyi üzerinde incelenen nesnenin vektörünün yönünde kaydırır; düzeltilmiş gösterge daha sonra konu noktasında CEF’in yüksekliği hakkında bir tahmin verir. Yapısal olarak, düzeltmeleri satış verilerinden çıkarıldığında, tablo koşullu bir ortalamanın basit bir parametrik olmayan tahmincisi (yüzey için herhangi bir formül varsaymayan) olarak görülebilir — bu tahminci, OLS yerine yerel ortalamalama ve manuel olarak belirlenen eğimleri kullanır; maliyetten veya yargıdan alınan düzeltmeler, bu tür tahminler değil, yapılandırılmış görüşlerdir.

Tabloyu bu şekilde çerçevelemek, değerleme uzmanının dile getirilmemiş varsayımlarını gün yüzüne çıkarır. Emsaller kullanıldığında, CEF’in, yakın noktaların incelenen nesnenin noktasındaki değerle ilişkili olabileceği kadar düzgün olduğu varsayılır. Doğrusal bir düzeltme, söz konusu nitelik için CEF’in eğiminin kapsanan aralık boyunca yaklaşık olarak sabit olduğunu varsayar. Üç göstergeye eşit ağırlık verilmesi, her birinin aynı koşullu ortalamayı tahmin ettiğini, hiçbirinin sistematik olarak sapma göstermediğini ve hepsinin benzer bir hassasiyete sahip olduğunu varsayar. Bu varsayımların hiçbirisi olağandışı değildir; hepsi bir yüzeyin şekliyle ilgilidir ve söz konusu yüzey

tanımlandığında hepsi tartışma konusu haline gelir.

Bu çerçeveleme, aralıklar ile ilgili yaygın bir karışıklığı da ortadan kaldırmaktadır. İki tür belirsizlik vardır. Bunlardan biri, CEF'nin nerede olduğu ile ilgilidir: uyumlanmış yüzeyin gerçek ortalamadan ne kadar uzak olabileceği. Diğeri ise, tek tek fiyatların bu ortalamanın etrafında ne kadar dağıldığıdır. Bir güven aralığı yalnızca ilkinin kapsar; bir tahmin aralığı ise her ikisini de kapsar ve bu nedenle her zaman ikisi arasında daha geniş olanıdır. Derginin temel regresyon yöntemi olan MARS ile, earth paketi (Milborrow 2026a), bir konu için ikinci bir modelden, yani fiyat dağılımının piyasa genelinde nasıl değiştiğini tahmin eden varyans modelinden (Milborrow 2026b) bir tahmin aralığı verir. Koşullar, yeterli sayıda satışın olması ve hataların kabaca normal dağılım göstermesi ve birbirinden bağımsız olmasıdır; paket, satışların aralığın içine ne sıklıkla girdiğini rapor eder ve bu da doğrulama işlevi görür. RCA temelli değerlemede, satış karşılaştırma yaklaşımı uyarınca nesnenin piyasa değeri, modelin tahmin ettiği satış fiyatı artı değerleme uzmanı tarafından tahmin edilen kalıntıdır. Varyans modeli, söz konusu koşulların sağlanması durumunda, nesnenin satışının

Tahmin edilen değişkeni —yani, tahmin edilen popülasyonun nitelikleri— isimlendirmek, bir savunulabilirlik eylemidir; zira bu olmadan bir sonuca ilişkin yargıda bulunulamaz. “Düzeltilme değeri \$100,000'dir” ifadesi, henüz doğru ya da yanlış olarak değerlendirilebilecek bir ifade değildir. Ancak, “Brüt yaşam alanı (GLA), arsa ve durum sabit tutulduğunda, manzaralı ve manzarasız taşınmazlar arasındaki koşullu ortalama satış fiyatı farkı, bu pazar segmentinde \$100,000'dir” dediğimizde, sabit tutulan unsurlar, söz konusu nüfus ve ilgili birimler temelinde tartışmaya açık bir iddiada bulunmuş oluruz.

Savunulabilirlik bakımından önemi. Bundan üç savunulabilirlik sonucu ortaya çıkmaktadır.

'daki koşullandırma kümesinin seçimi (sabit tutulan özellikler) metodolojik bir seçimdir: bu seçim, tahmin edilecek büyükeyi tanımlar. Verilerin bu tahmin edilecek büyükeyi tahmin edip edemeyeceği ise ayrı bir sorudur ve bu konu, *Parametre Tanımlama* başlığı altında ele alınır. $E[P \mid \text{GLA, view}]$ ve $E[P \mid \text{GLA, view, condition}]$ ifadeleri, farklı değerlere ve farklı eğimlere sahip olabilecek farklı fonksiyonları temsil eder. Bir değerleme uzmanı, tabloya bir değişken eklediğinde, sadece tahmini iyileştirmekle kalmaz; aynı zamanda tahmin nesnesini de değiştirir. Makul yaklaşım, değerleme görüşünün dayandığı koşullandırma kümesini belirtmektir.

“Diğer faktörler sabit tutulduğunda” ifadesi, değerleme uzmanının analiz sırasında nasıl bir mantık izlediğini açıklar ve kesin bir anlamı vardır. CEF çerçevesi içinde bu ifade, yalnızca tek bir özellik bakımından farklılık gösteren iki nitelik vektöründeki koşullu ortalamaların karşılaştırılmasını ifade eder. Bu, genel popülasyona ilişkin bir ifadedir ve yeterli piyasa verisi mevcut olduğunda bu verilerle karşılaştırılarak doğrulanabilir. Bunun arkasındaki analiz genellikle rapora değil, USPAP'ın Kayıt Tutma Kuralı'nın değerleme uzmanının tutmasını gerektirdiği görev kaydı olan çalışma dosyasına eklenir.

Artık dağılım bir bilgidir, utanç kaynağı değildir. Koşullu beklenti fonksiyonu etrafındaki fiyat dağılımı, piyasanın bir nitelikidir; fiyatların bu fonksiyonun etrafında geniş bir aralıkta dağıldığı bir piyasada, hangi yöntem kullanılırsa kullanılsın, tek bir rakamdan oluşan herhangi bir tahmin büyük bir belirsizlikle birlikte olacaktır. Bu dağılıma atıfta bulunmadan bir değer belirtmek, sonucu piyasa değeri kadar kesinmiş gibi gösterir.

Çözümlü değerleme örneği. 'in 1. sayısında yer alan sekiz Pacifica emsal fiyatı, gürültüsüz bir örneklemdir: bunlar, tek bir doğrusal denklemi tam olarak karşılayacak şekilde oluşturulmuştur. Bu sekiz fiyat, brüt kiralanabilir alan (GLA), arsa, manzara, durum ve tadilat göstergesi üzerinde regresyon analizine tabi tutulduğunda, her fiyat dolar bazında tam olarak yeniden üretilir — hiçbir fark kalmaz — ve regresyon, fiyatların oluşturulduğu denklemi oran oran geri verir:

$$E[P \mid \mathbf{Z}] = -30,000 + 600 \cdot z_{\text{GLA}} + 25 \cdot z_{\text{lot}} + 100,000 \cdot z_{\text{view}} + 20,000 \cdot z_{\text{cond}} + 60,000 \cdot z_{\text{remodel}}.$$

Buradaki “durum”, bu Dergi’nin durum puanıdır: Tanımlanmış piyasa bölgesindeki çevredeki evlerin durumuna göre söz konusu evin durumuna ilişkin bir tahmin. Bu puan, değerlendirme uzmanının söz konusu piyasa bölgesine ilişkin anlayışına dayanır ve aynı piyasa bölgesinde faaliyet gösteren diğer değerlendirme uzmanlarının anlayışlarıyla önemli ölçüde gelişmemelidir. Puan 0,00 ile 10,00 arasında değişir: piyasa bölgesindeki daha kötü durumda olan evlerin yüzdesinin on ile bölünmesiyle elde edilir. 5,00 puanı, piyasa bölgesindeki evlerin yarısının daha kötü durumda olduğu anlamına gelir; 9,00 puanı ise on evden dokuzunun daha kötü durumda olduğu anlamına gelir. 1. Soru, durumu kelimelerle tanımlamış ve regresyon için 1–3 arasında kodlamıştır; bu Dergi’nin puanlamasında kelimeler ortalama 5,00, iyi 7,00, mükemmel 9,00 olarak eşleştirilmiştir; dolayısıyla 1. Soru’daki bir basamak iki puandır ve piyasa puan başına \$20,000 öder. Son terim, oluşturulan verilerin bir tadilatla atfettiği fiyat farkıdır: Aynı brüt kiralananabilir alan (GLA), arsa, manzara ve durum puanına sahip bir eve göre \$60,000 daha fazla. Bu, tadilat eyleminin kendisi için bir fiyat değeridir; diğer özelliklerin kaydetmediği, tadilatın getirdiği her ne varsa onu temsil eder (giriş 008, *İhmal Edilen Değişken Önyargısı*).

Bu, tam olarak oluşturulmuş Pacifica piyasasına ait CEF’dir. Az önce ne tür bir nesnenin tanımlandığını ve bunun neden bu kadar kesin bir şekilde tanımlanabildiğini netleştirmek önemlidir. Veri kümesi oluşturulmuştur ve oluşturulmuş bir veri kümesi aslında gürültüsüz olabilir; gerçek bir veri kümesi ise asla gürültüsüz değildir. Veriyi oluşturma avantajı, fiyatlardaki rastgele gürültüyü tamamen ortadan kaldırmasıdır — her fiyat tam olarak formüle uymaktadır — bu nedenle 008 (*Atlanan Değişken Önyargısı*) ile 010 (*Parametre Tanımlama*) arasındaki maddelerde, bir tahmin ile hedefi arasındaki fark gürültüden kaynaklanamaz: her biri önyargıdır ya da, *Parametre Tanımlamasında*, tanımlama hatasıdır.

Bu örnekte, geleneksel değerlendirme yöntemlerinde ve 1. Sayı’da olduğu gibi, durum, dolar kuru içeren bir özellik olarak ele alınmaktadır; 008 (*Atlanan Değişken Önyargısı*), 009 (*Endojenlik*) ve 010 (*Parametre Tanımlama*) maddeleri ise bu yaklaşımın hangi sorunlara yol açtığını göstermektedir. Geleneksel değerlendirme metodolojisi, Dergi’nin kullandığı RCA tabanlı değerlendirme yöntemi değildir. Artık Kısıt Yaklaşımı’nda regresyon yalnızca ölçülebilen özellikleri dikkate alır; durum ve kalite kasıtlı olarak dışarıda bırakılır. Regresyonun satış fiyatının açıklayamadığı kısım, artık değerdir. Editörün Kaliforniya pazarlarında, ölçülen özellikler genellikle satış fiyatlarındaki değişkenliğin yaklaşık yüzde 80’ini açıklar; geri kalan beşte biri ise artık değere kalır. Değerleme uzmanı, tek bir kısıtlama altında, kendi puanlarını kılavuz olarak kullanarak artık değeri durum, kalite ve verilerin ölçemediği diğer özellikler arasında böler: artık bileşen değer katkılarının toplamı, emsal gayrimenkulün artık değerine eşit olmalıdır. Değer sonucunun, değerlendirme uzmanının bunu nasıl böldüğüne bağlı olmadığı — RCA makalesinin İkinci Kanıtı (Craytor 2026, s. 36): tüm artık bileşen değer katkılarının toplamının emsal gayrimenkulün artık değerine eşit olması kısıtlaması altında artık dağılımını değiştirmek, nihai değer sonucu üzerinde hiçbir etkiye sahip değildir. Bu nedenle, durum oranının tahmini asla yapılmaz ve hiçbir zaman iddia edilmez; değerlendirme uzmanının duruma atfettiği değer, yaş da dahil olmak üzere ölçülen özelliklerin zaten hesaba kattığı değerden ötesindeki durumdur.

1. Sayı’daki regresyonu bu CEF ile karşılaştırsak, regresyon hakkında bir yargıya varılabilir — ancak öncelikle bir ayırım yapmak gerekir. Genel olarak konuşursak, bir OLS regresyonu hedeflediği koşullu beklenti fonksiyonunu yeniden oluşturmaz; bunun yerine, ona en iyi doğrusal yaklaşımı, yani doğrusal izdüşümünü bulur ve bu ikisi yalnızca CEF’in doğrusal olduğu durumda birbirleriyle örtüşür. MARS gibi parametrik olmayan bir yöntem, CEF’in doğrusal olmasına gerek duymaz: CEF’yi parça parça yaklaştırır; yüzeyin büküldüğü bir menteşe noktası (oranın değiştiği nokta) ve kıvrıldığı bir etkileşim terimi (başka bir özelliğe bağlı bir oran) ile, doğrusal bir izdüşümün düzleştirileceği doğrusal olmayan bir CEF’yi takip eder. Bu yine de aynı nesneye yönelik bir yaklaşımdır ve sonraki maddelerde ele alınacak olan ihmal edilen değişken, içsellik ve tanımlama sorunları, bu

yaklaşımına da tam olarak aynı şekilde uygulanır. Yukarıda bahsedilen beş değişkenli CEF doğrusaldır; bu nedenle, yeniden modelleme göstergesini içeren regresyon, bu göstergelyi tam olarak geri kazanabilmektedir. Remodel göstergesini hariç tutan ve dört nitelikte bağı olan Sayı 1 regresyonu, $E[P \mid \text{GLA, lot, view, cond}]$ için en iyi doğrusal yaklaşımı tahmin eder ve koşul derginin puanına göre yeniden ifade edildiğinde (Sayı 1, 1-3 kodlamasının her adımı için 80,777 ve 49,824 değerinde bir kesme noktası vermiştir; bu aynı uyumdur),

$$\hat{p}(\mathbf{z}) = 6,041 + 575 \cdot z_{\text{GLA}} + 22.70 \cdot z_{\text{lot}} + 86,179 \cdot z_{\text{view}} + 24,912 \cdot z_{\text{cond}}.$$

Her katsayı, kendinden bir üstteki katsayıdan farklıdır. Regresyon, doğruya uydurma açısından açıklandıktan sonra, ek bir açıklama yapmaya gerek kalmaz: elimizde iki model ve iki sayı kümesi vardır. Ancak, CEF yaklaşımı kapsamında iki kesin ifade ortaya çıkmaktadır. İlk olarak, bu iki regresyon farklı değişken kümelerine bağlıdır ve dolayısıyla farklı koşullu beklenti fonksiyonlarını yaklaşık olarak ifade eder:

$$E[P \mid \text{GLA, lot, view, cond}] \quad \text{and} \quad E[P \mid \text{GLA, lot, view, cond, remodel}].$$

Bazı yenilenmiş ve yenilenmemiş taşınmazların aynı dört nitelik paylaştığı bir piyasada, bu iki fonksiyon söz konusu noktalarda farklı değerlere sahiptir. Bu küçük yapay pazarda, hiçbir iki taşınmazın dört nitelik aynı değildir; dolayısıyla iki fonksiyon her bir taşınmazda birbiriyle örtüşür ancak biçim olarak farklılık gösterir: ikincisi nitelikleri açısından doğrusal bir formüldür, oysa ilki değildir. Her durumda, regresyon hedefine yalnızca en iyi doğrusal yaklaşımı sağlar. Her iki hedef de geçerli nesnelerdir. İkincisi, yalnızca beş değişkenli olanın, örtük fiyatlar olarak okunabilecek katsayıları vardır — tadilat katsayısı, aynı brüt kiralanabilir alan (GLA), arsa, manzara ve durum puanı olan tadil edilmiş ve tadil edilmemiş bir ev arasındaki fiyat farkıdır — ve 008 numaralı madde (*Atlanan Değişken Önyargısı*) bunun nedenini açıklamaktadır.

1. Sayının Comp H için hesaplanan artık değeri, bu ayrımın ne kadar önemli olduğunu açıkça göstermektedir. 1. Sayıdaki sekiz nitelik tüm pazarı temsil eden unsurlar olarak kabul edersek, her dört değişkenli nitelik vektörü sekiz nitelikten tam olarak biriyle eşleşir ve dolayısıyla dört değişkenli CEF, sekiz fiyatın tümünden tam olarak geçer — Comp H de bunlardan biridir. Comp H, dört değişkenli CEF’den sıfır sapmaya sahiptir; 1. Sayı’daki artırı \$46,455, uyumlu dört değişkenli doğrusal regresyondan — yani o sekiz taşınmazlı pazarın doğrusal izdüşümünden — ne kadar farklı olduğunu gösterir ve yeniden modellemenin etkisi tam olarak bu izdüşüm ile CEF arasındaki boşlukta gizlidir. Beş değişkenli spesifikasyonla, projeksiyon ve CEF özdeş hale gelir ve sapma ortadan kalkar. Artık, Comp H’nin kendisinden değil, spesifikasyondan kaynaklanmaktadır.

References.

- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Craytor, W. B. (2026). Residual Constraint Approach (RCA): Framework & Protocol. *The Valuation Engineer*, 1(1). <https://doi.org/10.67330/kgebx46>
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Julius Springer.
- Milborrow, S. (2026a). *earth: Multivariate Adaptive Regression Splines* (R package version 5.3.6; derived from mda:mars by T. Hastie and R. Tibshirani). <https://doi.org/10.32614/CRAN.package.earth>
- Milborrow, S. (2026b, August 7). *Variance Models in earth*. <https://www.milbo.org/doc/earth-varmod.pdf>

Addendum: Prolog Formalization

The following machine-readable companion (007-conditional-expectation.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>.
```

```
%% 007-conditional-expectation.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 2, Entry 007
%% Title: Conditional Expectation
%% Author: Bert Craytor
%% Concept: conditional_expectation
%% Version: 1.2.0
%%
%% v1.2.0: entry text revised for readers who are not statisticians;
%% no Prolog clause changes.
%%
%% v1.1.0: condition on the Journal's 0.00-10.00 score (average 5.00, good
%% 7.00, excellent 9.00; 20,000 a point). The generating function lives in
%% kb/pacifica_comps_extended.pl, which is restated in those units with this
%% release; nothing here changes but the units it reads.
%%
%% This file is the Prolog galley for the entry: it encodes the entry's
%% main claims, its worked example, and the generating function of the
%% running dataset in a form other Prolog code in the journal can
%% consult, query, and verify. Read alongside entry.tex / entry.pdf.
```

```
% =====
% DICTIONARY IMPORT
% =====
```

```
% The dictionary (Issues/tools/dictionary) is not loaded at run time, so
% that the file runs unchanged in Trealla, SWI-Prolog and Scyer. The
% declarations below are plain facts, read by the dictionary linter.
dictionary_release_pinned('dictionary-2026.1').
```

```
% --- From the dictionary (quoted for reference) -----
% term(conditional_expectation, "Conditional Expectation",
%     econometric_method,
%     "The conditional expectation of an outcome given a set of
%     predictors is the mean of the outcome's distribution at each
%     fixed value of the predictors. Viewed as a function of the
%     predictors, it is called the conditional expectation function
%     (CEF), and it is the function that regression aims to
%     approximate. Ordinary least squares (OLS) regression gives the
%     best linear approximation to the CEF, and it reproduces the CEF
```

```

%      exactly when the CEF itself is linear.", published,
%      'dictionary-2026.1').
%
% depends_on(conditional_expectation, hedonic_regression).
% -----

uses_terms([ conditional_expectation, hedonic_regression,
              implicit_price, hedonic_price_function,
              characteristics_space, residual,
              ordinary_least_squares,
              sales_comparison_approach, adjustment,
              comp_selection, reconciliation ]).
uses_relations([depends_on/2, related/2]).

% =====
% ENTRY SUBJECT
% =====

article_concept(conditional_expectation).
article_issue(volume(1), number(2), year(2026)).
article_id('foundations.2026.007').

% =====
% ENTRY-LOCAL VOCABULARY
% =====
%
% Field names for the worked-example records and for the two properties
% of the CEF the entry relies on. Too narrow to promote to the
% dictionary.

local_term(cef).                % conditional expectation function
local_term(estimand).
local_term(conditioning_set).
local_term(prediction_interval).
local_term(confidence_interval).
local_term(linear_projection).  % best linear approximation to the CEF

% =====
% THE TWO PROPERTIES OF THE CEF THE ENTRY USES
% =====

%! cef_property(?Name, ?Statement) is nondet.

cef_property(best_predictor,
  "Among all functions of the conditioning variables, the CEF minimizes expected squared
   $\hookrightarrow$  prediction error.").
cef_property(orthogonal_deviation,
  "P = E[P|Z] + eps with E[eps|Z] = 0 holds by construction, not by assumption.").

% The second property is the hinge for entries 008-010: each asks when
% the recovered object is the CEF of interest.
hinge_for(orthogonal_deviation, omitted_variable_bias).
hinge_for(orthogonal_deviation, endogeneity).
hinge_for(orthogonal_deviation, identification).

% =====
% CENTRAL CLAIM: CONDITIONING SET DETERMINES THE ESTIMAND
% =====
%

```

```

% Two regressions over the same data with different conditioning sets
% target DIFFERENT population objects, and each recovers only the best
% linear approximation to its target (the linear projection); OLS
% recovers the CEF itself exactly when the CEF is linear. Neither is
% wrong; they answer different questions.

conditioning_set(issue1_model,
    [gla_sqft, lot_sqft, view_numeric, cond_score]).
conditioning_set(issue2_model,
    [gla_sqft, lot_sqft, view_numeric, cond_score, remodel_numeric]).

%! different_estimand(?ModelA, ?ModelB) is nondet.
%
% True when two models condition on different variable sets and
% therefore approximate different conditional expectation functions.

different_estimand(A, B) :-
    conditioning_set(A, SA),
    conditioning_set(B, SB),
    A \== B,
    sort(SA, X), sort(SB, Y),
    X \== Y.

% What OLS recovers, relative to the CEF it conditions on.
ols_recovers(best_linear_approximation_to_cef).
ols_recovers_cef_exactly_when(cef_is_linear).

% =====
% WORKED EXAMPLE - THE GENERATING FUNCTION
% =====
%
% The running dataset is defined in ../../kb/pacifica_comps_extended.pl.
% Its generating function IS the conditional expectation function for
% the constructed market, exactly: the data carry no noise.

:- initialization(consult('../../kb/pacifica_comps_extended')).
% Scyer Prolog loads forall/2 and the list predicates on request;
% SWI-Prolog and Trealla have them built in. The guarded loads below
% let the file run unchanged in all three.
:- initialization(catch(use_module(library(lists)), _, true)).
:- initialization(catch(use_module(library(iso_ext)), _, true)).

%! cef_exact(+Property, -Price) is det.
%
% The CEF evaluated at a property's characteristic vector. Identical to
% true_price/2 in the dataset file; named here to make the entry's claim
% ("the generating function IS the CEF") explicit in the galley.

cef_exact(P, Price) :- true_price(P, Price).

%! noiseless_dataset is semidet.
%
% Verifies the entry's central factual claim about the data: every
% observed sale price equals the CEF at that property, so the residual
% is zero everywhere.

noiseless_dataset :-
    forall(property(P),
        ( cef_exact(P, Price), sale_price(P, Price) ) ).

```

```

% =====
% THE COMP H POINT
% =====
%
% Issue 1 reported a $46,455 residual for Comp H under the four-variable
% model. Taking issue 1's eight properties as the market, every
% four-variable characteristic vector belongs to exactly one property, so
% the four-variable CEF interpolates all eight prices and Comp H deviates
% from it by ZERO. The 46,455 is Comp H's deviation from the fitted
% four-variable LINEAR regression --- the linear projection --- and the
% projection-CEF gap is where the remodel effect hides. Under the
% five-variable specification, projection and CEF coincide and the
% deviation vanishes.

issue1_residual(comp_h, 46455).

% deviation_from(Property, Object, Dollars).
deviation_from(comp_h, four_variable_projection, 46455).
deviation_from(comp_h, four_variable_cef, 0).
deviation_from(comp_h, five_variable_cef, 0).

residual_under(issue2_model, comp_h, 0) :-
    cef_exact(comp_h, Price),
    sale_price(comp_h, Price).

%! residual_is_artifact_of_specification(?Property) is nondet.
%
% The issue-1 residual is a property of the specification (a linear
% projection under a smaller conditioning set), not of the property.

residual_is_artifact_of_specification(P) :-
    issue1_residual(P, R), R =\= 0,
    deviation_from(P, four_variable_cef, 0),
    residual_under(issue2_model, P, 0).

% =====
% APPRAISER ENCOUNTERS
% =====

grid_assumption(smoothness,
    "Using comparables at all assumes the CEF is smooth enough that nearby points inform the
    ↪ subject's point.").
grid_assumption(local_linearity,
    "Adjusting linearly assumes the CEF's slope in that characteristic is roughly constant over the
    ↪ range crossed.").
grid_assumption(unbiased_indications,
    "Averaging indications assumes each is an unbiased draw around the same conditional mean.").

% The grid is structurally a nonparametric estimator of a conditional
% mean, using local averaging and hand-specified slopes.
estimator_role(sales_comparison_approach, conditional_mean_estimator).

% =====
% DEFENSIBILITY
% =====

defensibility_consequence(conditioning_set,
    "Adding a variable to a grid changes the estimand, not merely the precision of an estimate.").

```

```

defensibility_consequence(ceteris_paribus,
    "'Holding other things equal' means comparing conditional means at vectors differing in one
    ↪ coordinate.>").
defensibility_consequence(residual_spread,
    "Conditional dispersion is a market property; a point estimate reported without it overstates
    ↪ precision.>").

% The confidence/prediction interval confusion the entry names.
interval_answers(confidence_interval, "uncertainty about where the CEF is").
interval_answers(prediction_interval, "spread of individual prices around the CEF").
commonly_confused(confidence_interval, prediction_interval).

```