

İçsellik (Endogeneity)

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

DOI: [10.67330/z9feps66](https://doi.org/10.67330/z9feps66)

Biçimsel tanım. Endojenlik, bir regresörün — regresyonun fiyatı açıklamak için kullandığı bir özelliğin — yapısal hata ile birlikte hareket etmesi durumudur; burada yapısal hata, fiyatın, nedensel ilişkinin fiyatın olması gerektiğini belirttiği değerden çıkarılmasıyla elde edilen farktır (uyumlanmış modelin artıkları değildir); bu durum, söz konusu regresörün tahmin edilen katsayısının, ilgilenilen nedensel etkiyi yansıtmamasına yol açar. Bunun yaygın nedenleri arasında, bir özelliğin dışarıda bırakılması, fiyatın o özelliğe geri etki etmesi, satışların seçilmiş bir küme olması ve özelliğin hatalı bir şekilde ölçülmesi sayılabilir.

Sezgisel çerçeve. Bir oran, özellik ile fiyatın birbirini etkilemesi nedeniyle hatalı olabilir. Bir evin piyasada kalma süresi, istenen fiyata bağlıdır: özellik kısmen fiyatın bir sonucudur ya da yükselen bir piyasa gibi fiyatı artıran başka bir faktörün sonucudur; bu satışlardan ölçülen bir oran ise bu iki yönü birbirine karıştırır. Hiçbir regresyon, bunları yalnızca satışlardan ayırarak çözemez; ister bir özelliğin tüm aralığı boyunca tek bir oranın uyumu sağlansın, ister MARS (Çok Değişkenli Uyarlanabilir Regresyon Spline’ları, Friedman 1991) modelinde olduğu gibi, özelliğin farklı aralıkları için farklı oranlar uygulansa bile — bu, dışsal bir varyasyon kaynağına ya da satış verilerinin doğrulayamadığı bir varsayma ihtiyaç duyar — çünkü bu karmaşıklık piyasada yatmaktadır, aritmetikte değil. Değerleme uzmanının yapabileceği şey, bir özelliğin modelin dışarıda bıraktığı unsurlarla birlikte hareket edebileceği dört yolu (aşağıda belirtilmiştir) bilmek, bunları bir veri kümesinde tanımak ve o veri kümesinden elde edilen bir oran, sanki alıcının o özellik için ödediği tutarı ölçüyormuş gibi kullanılmadan önce, hangisinin söz konusu olduğunu belirlemektir.

Mekanizmayı öğrenmek isteyen okuyucular için. İstatistikçiler, bir özelliğin modelin dışarıda bıraktığı bir unsurla birlikte hareket etmesi durumunda bu özelliğe “endojen” adını verir; bu madde, söz konusu koşulun kesin biçimini ortaya koyar ve dört kaynağını ayrı ayrı ele alır. İstatistik literatürünün sunduğu çözüm yolları, başlıca araçsal değişkenler olmak üzere, sıradan regresyon için belirtilmiştir ve değerlendirme verilerinde nadiren mevcuttur; maddede bu durum belirtilmiştir. Teşhis ise, çözüm yolunun aksine, herhangi bir uyum yöntemine uygulanabilir.

Bu madde istatistiksel bir bağlam içermekte ve aynı zamanda nedensel akıl yürütmeyi de kapsamaktadır; zira bir regresyon katsayısının, değerlendirme uzmanının aradığı nedensel etkiyi tam olarak yansıtmadığı durumları ele almaktadır. Bu madde, katsayıları veya eşleştirilmiş satış düzeltmelerini nedensel ifadeler olarak yorumlayan değerlendirme uzmanlarına yöneliktir. 008 numaralı madde (*Atlanmış Değişken Önyargısı*) gibi, bu madde de geleneksel değerlendirme yönteminde olduğu gibi regresyona koşulu dahil eden, yapay olarak oluşturulmuş Pacifica pazarını kullanmaktadır; derginin kendi yöntemi olan RCA tabanlı değerlendirme ise bunu yapmamaktadır (007 numaralı madde, *Koşullu Beklenti*) ve bu örnek, izlenecek bir model değil, karşılaştırma amaçlı bir örnektir.

008 numaralı konuda (*Atlanan Değişken Önyargısı*) ele alınan atlanan değişken önyargısı, kendine özgü bir formülü olan bağımsız bir hata türü olarak değil, daha genel bir durumun örneği olarak görülmelidir: içsellik.

007 numaralı maddeyi (*Koşullu Beklenti*): $P = E[P | \mathbf{Z}] + \varepsilon$ ’daki ayrıştırmayı hatırlayalım; burada $E[P | \mathbf{Z}]$, $\mathbf{Z}$ özelliklerine sahip evlerin ortalama fiyatını, ε ise bir evin fiyatının bu ortalama-
madan ne kadar yukarıda veya aşağıda olduğunu ifade eder; dolayısıyla tanım gereği $E[\varepsilon | \mathbf{Z}] = 0$

olur. Bu, $\mathbf{Z}$ 'in hangi özellikleri içerdiğine bakılmaksızın koşullu beklenti fonksiyonu (o maddede geçen CEF) için geçerlidir. Değerleme uzmanı bunun yerine nedensel bir etki istiyorsa — evle ilgili diğer tüm faktörler gerçekten sabit tutulduğunda, manzara olması durumunda fiyatın ne kadar daha yüksek olacağı — bir koşul daha gereklidir: regresörler, fiyatı etkileyen başka hiçbir faktörle birlikte değişmemelidir. Bunlardan biri değişirse, o regresör *endojen*dir. Nedensel ilişkiyi $P = \beta_0 + \beta_1 z_1 + \dots + \beta_k z_k + u$ şeklinde yazalım; burada her bir β_j , z_j özelliğindeki bir birimlik değişimin fiyata yapacağı etkiyi gösterir. Hata terimi, olarak ifade edilen *yapısal* hatadır ve u şeklinde yazılır: fiyatın, nedensel ilişkinin öngördüğü değerden arındırılmış hali. Bu, koşullu ortalama uzaklık olan ε 'ten farklı bir kavramdır; , yapısı gereği $\mathbf{Z}$ verildiğinde ortalama değeri sıfırdır; u ise regresörlerle birlikte değişebilir ve içsellik koşulu buna atıfta bulunur.

Bu terim, eşzamanlı denklem modellemesinden kaynaklanmaktadır; bu modellemede, bir değişken dışarıdan verilmek yerine sistem içinde belirleniyorsa “endojen” olarak adlandırılır. Güncel kullanımda ise terimin daha mekanik bir anlamı vardır: $\text{Cov}(z_j, u) \neq 0$ (Cov, yani kovaryans, iki büyüklüğün birbiriyle nasıl birlikte hareket ettiğini ölçer) — regresör ve yapısal hata birlikte hareket eder; bu nedenle uyum (OLS — sıradan en küçük kareler regresyonu — veya MARS), bunların katkılarını birbirinden ayıramaz ve hatanın hareketinin bir kısmını regresöre atfeder.

Buna dört mekanizma neden olmaktadır ve her biri farklı tepkiler gerektirdiği için bunları birbirinden ayırmak önemlidir.

Atlanan değişkenler. Fiyatı etkileyen bir özellik dışarıda bırakılır ve dahil edilen bir özellik ile birlikte hareket eder; bu durum “ u ” (Atlanan Değişken Yanılgısı) haline gelir ve ardından “ $\text{Cov}(z_j, u) \neq 0$ ” (Atlanan Değişken Yanılgısı) ortaya çıkar. Bu durum, “remodel” göstergesinin dışarıda bırakıldığı 008 numaralı girişte (Atlanan Değişken Yanılgısı) görülmektedir. Bunu düzeltmek için, söz konusu özelliği ölçüp modele dahil etmek ya da önyargının makul olarak alabileceği en büyük ve en küçük değerleri belirlemek gerekir.

Eşzamanlılık. Regresör, fiyatla sadece ilişkili olmakla kalmaz, kısmen fiyat tarafından da tetiklenir ya da fiyatla birlikte belirlenir; nedensellik her iki yönde de işler ve uyum (OLS veya MARS) yön hakkında bir şey söylemediğinden, bu iki durumun bir karışımını ortaya koyar. Bunun çözümü, bir *araç* değişkeni — söz konusu regresör değişkenini etkileyen ancak fiyatla başka hiçbir bağlantısı olmayan (kelimenin ekonometrik anlamında, resmi bir belgenin hukuki anlamında değil) bir değişken — ya da farklı bir tasarımdır.

Örnekleme. Örneklem, sonucun ilgili olduğu konutlardan rastgele seçilmemiştir ve bir konutun örnekleme dahil edilip edilmemesi, konutun fiyatıyla bağlantılıdır. Bunun çözümü, örneklemin nasıl seçildiğini modellemek ya da sonucun ilgili olduğu grubu yeniden tanımlamaktır.

Ölçüm hatası. Özellik yanlış kaydedilmiştir — iki denetçi bu özelliğe farklı puanlar verecektir ya da yaşam alanı ölçülmek yerine bir kayıttan alınmıştır. Kayıt hatası, u adresindeki verinin bir parçası haline gelir ve kaydedilen değerle birlikte değişir; dolayısıyla katsayı sapma gösterir. Klasik bir örnek olarak — hatanın gerçek değerle veya modeldeki başka herhangi bir şeyle ilgisi olmadığı *klasik* ölçüm hatası durumunda — yanlış ölçülen özelliğin katsayısı sıfıra doğru çekilir. Koşul puanı bu durumdan farklı olabilir: hatası, evin gerçekte ne kadar iyi olduğuna ve puanlayıcının gördüğü diğer unsurlara bağlı olabilir. Bu tür hatalarda sapmanın genel bir yönü yoktur. Bunun çözümü, daha iyi ölçüm yapmak ya da aynı evin aynı özelliğine ilişkin ikinci ve bağımsız bir ölçüm elde etmektir; ölçüm hatası için tasarlanmış bir yöntem, bu ikinci ölçümü kullanarak tahmini düzeltebilir.

Bu dört mekanizmanın ortak bir özelliği vardır: Aynı türden satışların sayısını artırmak, bunların yol açtığı hatayı ortadan kaldırmaz. Bu, içsellik ile belirsizlik (sadece sınırlı sayıda satıştan kaynaklanan bir tahmin değerindeki dağılım) arasındaki en net pratik ayırmadır. Örneklem büyüdükçe belirsizlik azalır, içsellik ise azalmaz.

Değişkenleri seçmeden önce, etkinin adını belirleyin. Bir özelliğin modele dahil edilip edilmeyeceği, neyin tahmin edildiğine bağlıdır ve iki tür özellik kolayca karıştırılabilir. Birincisi,

hem regresörün hem de fiyatın nedenidir: bir mahalledeki cazip yerler (sessiz bir çıkmaz sokak, manzaralı bir köşe) pahalı tadilatları çeker ve kendi başlarına yüksek fiyatlara neden olur. İstatistikçiler bu tür bir özelliğe *karıştırıcı faktör (confounder)* derler. Bu faktör sabit tutulmalıdır, aksi takdirde regresörün katsayısı bu özelliğin etkisini de taşır. İkincisi, regresörün fiyat üzerinde etki ettiği kanaldır: bir tadilat, durumu ve kaliteyi iyileştirerek fiyatı yükseltir. Böyle bir özellik bir *aracı* değişkendir ve bunu sabit tutmak, regresör katsayısından tam olarak bu kanal üzerinden geçen etkisinin kısmını ortadan kaldırır. Dolayısıyla, 008 numaralı madde (*Atlanmış Değişken Önyargısı*) iki farklı soru ortaya koyar. Geleneksel değerlemede, eğer soru piyasanın durum için belirlediği oran ise, durum regresör olur (RCA temelli değerlendirme bu soruyu asla sormaz; 007 numaralı giriş, *Koşullu Beklenti*, bunun nedenini açıklar) ve tadilat, yalnızca bazı evlerin mevcut durumlarına nasıl ulaştıklarını gösterir; bu, durum ve kalite tam olarak kaydedildikten sonra, bunların ötesinde hiçbir değer kazandırmaz; durum ve kalite açısından benzer olan yenilenmiş bir ev ile bakımlı bir evin değeri aynıdır. Soru, bir tadilatın belirli bir evin fiyatına ne gibi bir etki yaptıysa, o zaman durum ve kalite aracı değişkenlerdir ve bunları sabit tutan bir model farklı bir soruyu yanıtlar — tadilatın, ortaya çıkardığı durum ve kaliteden *ayrı olarak* ne kadar değer kattığı sorusunu; bu değer küçük olabilir ve her halükarda tadilatın toplam etkisinden farklı bir miktardır. “Değişkeni dahil et” kuralı, karıştırıcı faktörler için doğrudur ancak aracı faktörler için yanlıştır; istenen etki belirlenene kadar ikisinden biri de seçilemez.

Değerleme uzmanı nerede karşılaşır. Ters nedensellik, konumla ilgili bir konudur. Eş-zamanlılıktan kaynaklanan içsellik, varsayımsal değil, gerçek bir sorun olan okul kalitesi örneğiyle iyi bir şekilde açıklanabilir. Fiyatı okul derecelendirmesine göre regresyon analizine tabi tutarsanız, katsayı alıcıların okul için ne kadar ödediğini gösteriyor gibi görünür. Ancak nedensellik tersi yönde de işler: daha pahalı mahalleler daha geniş bir vergi tabanına ve daha ilgili, daha iyi kaynaklara sahip ebeveynlere sahiptir ve öğretmenleri daha iyi çekip elde tutabilir. Okul kalitesi fiyatları artırabilir ve daha yüksek fiyatlar okul kalitesini iyileştirebilir. Katsayı, bu iki etkiden hiçbirini tek başına yansıtmaz; o pazardaki iki kanalın göreceli gücüne bağlı bir karışımı yansıtır. Aynı iki yönlü ilişki, suç oranları, perakende tesis yoğunluğu ve ağaç örtüsü ile de ortaya çıkabilir. Her durumda bu faktör, mahalle fiyat seviyelerinin hem nedeni hem de sonucu olabilir.

Bir tadilat bir müdahaledir, ancak rastgele atanmaz. Nedensel çıkarım terminolojisinde *müdahale*, etkisi incelenen herhangi bir müdahaledir ve mutfak tadilatı da buna dahildir. Zorluk, bu müdahalenin kime uygulandığıdır. Birçok ev sahibi, getirinin maliyeti aşacağını beklediklerinde tadilat yaptırır ve bu beklenti, veri kümesinde kayıtlı olmayan, evleri hakkında bildikleri şeylere dayanır — iyi bir kat planı, mahalle içinde iyi bir konum, yatırımı destekleyen bir arsa. Aynı nitelikler fiyatı doğrudan artırır: bunlar tadilatın karıştırıcı faktörleridir ve veri kümesi tarafından ölçülmeyen faktörlerdir. Dolayısıyla, tadilat göstergesini içeren bir model, göstergenin eksikliğini gidermiştir; karıştırıcı faktörleri ortadan kaldırmamıştır. Göstergenin katsayısı, tadilatın fiyat üzerindeki etkisini, sahipleri tadilat yapmayı seçen evlerin kayıtlı özelliklerinin açıkladığı değer ötesindeki ek fiyatla birleştirir. Değişkenin dahil edilmesi, eksikliği gidermiştir; ancak regresör değişkeni dışsal hale getirmez — yani yapısal hatayla bağlantısız hale getirmez.

Satılan taşınmazlar örneklemi rastgele değildir. İnsanlar koşullar uygun olduğunda veya fiyatlar cazip olduğunda satış yaparlar; diğerleri ise mecbur oldukları için satış yaparlar. Dolayısıyla, satılan taşınmazlar, modelin öngöremediği şekillerde satılmayanlardan farklılık gösterebilir; bu durumda, satılan bir taşınmazın fiyatındaki kendine özgü unsur rastgele değildir: ilk grup fiyatı bir yöne, zor durumdaki ve zorunlu satışlar ise diğer yöne itmektedir. Satışlar içinde, piyasa koşullarına uygun işlemler ve zor durum satışları, her biri rastgele olmayan alt gruplardır. Her tablodaki her emsal değer, bu seçilmiş satışlardan alınır. Herhangi bir değerlendirme sonucu, satılan taşınmazların nasıl seçildiğine dair varsayımlara dayalı olarak satılmamış taşınmazlara genişletilebilir; bu

varsayımlar raporda belirtilmelidir.

Durum puanı, fiyat nedeniyle etkilenebilir. Bir puan, yalnızca evin durumunu ölçmek için tasarlanmıştır; değerlendirme uzmanı, herhangi bir değer belirlemeden önce inceleme sırasında söz konusu evi puanlar. Emsaller ise başka bir konudur: bunlar, satış fiyatları öğrenildikten sonra, ilan fotoğrafları ve açıklamalarına göre puanlanabilir. Bilinen fiyat puanı etkiliyor ise —yüksek fiyata satılan bir evin bakımı biraz daha iyi olarak değerlendirilmesi gibi— o zaman regresyon değişkeni kısmen sonucun bir fonksiyonu haline gelmiştir. Bu, durumun kendisinin bir nitelik değil, ölçüm sürecinin bir kusurudur: *sonuca bağlı ölçüm hatası*; fiyatın puan üzerinde etkisi olmasıdır. Bu, az önce açıklanan klasik olmayan türden bir hatadır, ancak bir farkı vardır. Fiyat sapmasının bir kısmını puana yansıtan bu tür bir etkide, yön bellidir: puan, satış fiyatının işaret ettiği puana doğru kayar ve durum katsayısı yukarı doğru itilir (puanı etkileyen diğer hatalar yine de ters yönde etki edebilir). Koruma, usule dayalıdır: emsalleri, satış fiyatlarına bakmadan önce durumlarına ilişkin kanıtlara göre puanlayın ve her puanın dayanağını çalışma dosyasına kaydedin.

Savunulabilirlik bakımından önemi. Bir regresörün içsel olup olmadığı, argümanlarla belirlenir; sıradan bir regresyon analizinin ortaya koyabileceği hiçbir şeyle değil. Uyumun olağan kontrol yöntemlerinden hiçbirisi — artıkların grafiği, açıklanan fiyat değişiminin oranı (R^2), uyum dışında tutulan satışlar üzerinde modelin test edilmesi (çapraz doğrulama) veya bir katsayının sıfır olup olamayacağının testi (anlamlılık testi) — regresörün dışsal olduğunu kanıtlayamaz; çünkü yapısal hata gözlemlenemez ve regresörlerle olan bağlantısı da öyle. Resmi testler mevcuttur — Hausman testi, olağan tahmini başka bir yolla elde edilen tahminle karşılaştırır — ancak bunlar bir araç değişkeni veya ikinci bir tahminciye, yani verilerin kendilerinin doğrulayamayacağı ek varsayımlara ihtiyaç duyar. Her zaman verilebilecek olan şey, mekanizmaya ilişkin mantıklı bir açıklamadır: bu regresörün hatayla neden, hangi yönde ve ne ölçüde bağlantılı olabileceği. Makul bir rapor, sorunu bahsetmeyerek sorunun var olmadığını ima etmek yerine bu argümanı ortaya koyar.

Katsayının ne olduğunu söyleyin, ne olmasını istediğinizi değil. Endojen bir regresörün katsayısı, gerçek ve net bir rakamdır: ölçülen özellikler sabit tutulduğunda, söz konusu özelliğin birim başına uyumda bulunan fiyatındaki fark. Bu, nedensel etki değildir. Yenileme göstergesinin dışarıda bırakıldığı on altı Pacifica satışında (giriş 008, *Çıkarılan Değişken Önyargısı*), “ölçülen özellikler sabit tutulduğunda, manzaralı bu evler \$79,178 daha fazla fiyata satıldı” demek doğrudur. “Manzara, değere \$79,178 katkıda bulunur” demek, tahminin desteklemediği bir nedenselliği öne sürmektir. İlk ifade çapraz sorgulamaya dayanır; ikincisi ise dayanmaz.

Tasarım, tahminlerden üstündür. Bir özellik ile fiyat arasındaki karşılıklı etkileşimin, verilerin düzenlenme şekliyle ortadan kaldırılabilmesi — bir manzaranın engellenmesinden önceki ve sonraki taşınmazları karşılaştırarak, okul atamalarının değiştiği ancak konut stokunun değişmediği bir sınır kullanarak ya da mal sahibi tarafından seçilen bir yenileme yerine yangın veya sel gibi dışsal bir olayın zorladığı bir yenilemeyi dikkate alarak — her tasarım yine de savunulması gereken varsayımlara dayanır (aynı anda başka hiçbir şeyin değişmediği, sınırın yalnızca okul atamasını ayırdığı gibi), ancak varsayımları savunulan bir tasarım, gözlemsel verilerle çalışan herhangi bir tahminciden daha fazla savunulabilirlik sağlar. Araç değişkenler literatürü pratik açıdan bunu ifade etmektedir ve bunun değerlendirme ile ilgisi, iyi seçilmiş bir emsal grubunun başlı başına bir tasarım olmasıdır.

Çözümlü değerlendirme örneği. adresindeki Pacifica veri kümesi, içselliği örnekleyebilir ancak bu sorunu ortadan kaldıramaz; bu veri kümesinin bunu neden başaramadığının açıklanması, bu yazının asıl dersidir.

Öncelikle, soruya gelelim. Bir değerlendirme uzmanına, belirli bir ev için önerilen \$45,000 mutfak tadilatının satış sırasında geri kazanılıp kazanılmayacağı sorulduğuna varsayalım. Bu, tek bir evle ilgili ve aynı anda sahip olamayacağı iki fiyat hakkında bir sorudur: yenilenmiş haliyle fiyatı ve

mevcut haliyle fiyatı, $P_i(1) - P_i(0)$; burada $P_i(1)$, i adresindeki evin yenilenmiş haliyle fiyatı, $P_i(0)$ ise yenilenmemiş haliyle fiyatıdır. Bu, piyasanın duruma göre belirlediği fiyatla ilgili bir soru değildir ve soruyu yanıtlarken durumu sabit tutmak, yukarıda açıklanan aracı hatası olur.

008 numaralı girdinin (*Atlanmış Değişken Önyargısı*) on altı taşınmazı, katsayısı \$60,000 olan bir tadilat göstergesi içerir: Oluşturulan verilerin bir tadilata atfettiği fiyat farkı — aynı Brüt Kiralanabilir Alan (GLA), arsa, manzara ve durum puanı sahip bir eve kıyasla \$60,000 daha fazladır. Bu \$60,000 rakamı, tadilat işleminin bedeli değildir; diğer özelliklerin kaydetmediği, tadilattan kaynaklanan her ne varsa onu temsil eder. Bu oluşturulmuş veri kümesinde, hangi evlerin tadilata tabi tutulduğu yazar tarafından belirlenmiştir. Yenileme göstergesi, yapısı gereği dışsaldır ve hata teriminde geriye hiçbir şey kalmadığı için hata terimiyle bağlantılı değildir; dolayısıyla bu veri kümesinde \$60,000, kaydedilen durum puanı sabit tutulduğunda yenilemenin fiyata kattığı değerdir; puanı da yükselten bir yenileme ise daha fazla değer katacaktır.

Şimdi, aynı on altı taşınmazın gerçek bir piyasadan geldiğini varsayalım. Ev sahipleri evlerini yenilemeyi tercih ediyorlar ve — ki bu makul bir varsayımdır — gözlemlenemeyen nitelikleri güçlü olan taşınmazların sahiplerinin (fiyatın bu bileşenine u diyelim) diğer yönlerden benzer taşınmazların sahiplerine kıyasla evlerini yenileme olasılığının daha yüksek olduğunu varsayalım. Yenileme göstergesini içeren regresyon, satışlar arttıkça artık \$60,000’a değil,

$$\hat{\beta}_{\text{remodel}} \xrightarrow{p} \underbrace{60,000}_{\text{what the remodel brought that the other features do not record}} + \underbrace{\frac{\text{Cov}(\tilde{r}, u)}{\text{Var}(\tilde{r})}}_{\text{extra price of the houses selected into remodeling}},$$

değerine yakınsamaktadır; burada $\tilde{r}$, dahil edilen özelliklerin tahmin edemediği tadilat göstergesinin kısmıdır; Var (varyans), ne kadar değiştiğini ölçer ve p ile işaretlenmiş ok, “satışlar arttıkça giderek daha da yaklaşır” anlamına gelir. Yenilenmiş evler, gözlemlenebilir şekilde benzer olan yenilenmemiş evlere kıyasla ortalama olarak \$25,000 daha fazla u değerine sahipse, katsayı kabaca \$85,000 değerine yakınsar ve aynı pazardan elde edilen hiçbir ek veri bu fazlalığı azaltmaz. Daha önce 008 (*Çıkarılan Değişken Önyargısı*) girişinde incelenen tadilat göstergesi artık ölçülüp dahil edilmiştir, ancak katsayı hâlâ nedensel etkiyi yansıtmamaktadır: hâlâ dışarıda bırakılan unsur, hiçbir veri kümesinde kayıtlı olmayan u ’dir.

Bu ayrımın pratikte önemli bir etkisi vardır. Katsayı, ortalama olarak, yenilenmiş evleri tanımlar; yani sahipleri bu yatırımın kârlı olduğunu değerlendiren evleri. Konu, sahibi henüz karar vermemiş belirli bir evdir ve katsayı, sadece söz konusu ev, sahiplerinin bildikleri bilgiler açısından o evlere benziyorsa bu ev için de geçerlidir — ki bu tam da modelin göremeyeceği bir durumdur. Geri ödeme sorusuna \$85,000’dan yola çıkarak yanıt veren bir değerlendirme uzmanı, getiriyi abartır; yenilemenin ortaya çıkardığı tüm koşul ve kaliteyi de ölçen ve sabit tutan bir modelden yanıt veren bir değerlendirme uzmanı ise, geri dönüşü, bunların dışında yenilemenin kattığı değere indirger; bu da 008 numaralı ilke (*Atlanan Değişken Önyargısı*) uyarınca çok az ya da hiç bir şey değildir. Her ikisi de yenilemeyle ilişkili koşullu fiyat farkını, yenilemenin kendisinin etkisi olarak yorumlamıştır.

Dolayısıyla, nedensel bir şekilde yorumlamak istediğimiz herhangi bir katsayı için sorulması gereken tam sorusu, “model iyi tanımlanmış mı?” değil, “hangi etki aranıyor ve bu değişken aldığı değerlere nasıl ulaştı?” sorusudur. Eğer cevap, modelin göremediği şeyleri görebilen birinin yaptığı bir seçimi içeriyorsa, katsayı bu seçimi, özelliğin kendisi kadar iyi tanımlar.

References.

- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Haavelmo, T. (1943). The Statistical Implications of a System of Simultaneous Equations. *Econometrica*, 11(1), 1–12.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251–1271.
 Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
 Wright, P. G. (1928). *The Tariff on Animal and Vegetable Oils*. Macmillan.

Cross-references: omitted variable bias; conditional expectation; parameter identification.

Addendum: Prolog Formalization

The following machine-readable companion (009-endogeneity.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>.

%% 009-endogeneity.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 2, Entry 009
%% Title: Endogeneity
%% Author: Bert Craytor
%% Concept: endogeneity
%% Version: 1.2.0
%%
%% v1.1.0: the estimand is named before the variables are chosen
%% (confounder vs. mediator); a remodel is a treatment that is not randomly
%% assigned; its 60,000 is the price difference the constructed data
%% attach to a remodel; measurement error attenuates only in the classical case; a rating
%% contaminated by the known price is outcome-dependent measurement error.

% 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').

uses_terms([ endogeneity, omitted_variable_bias,
             conditional_expectation, hedonic_regression,
             latent_variable, residual, comp_selection,
             paired_sales_analysis, adjustment,
             sales_comparison_approach ]).
uses_relations([depends_on/2, related/2, introduced_by/2, formalized_by/2]).

article_concept(endogeneity).
article_issue(volume(1), number(2), year(2026)).
article_id('foundations.2026.009').

local_term(instrument).
local_term(simultaneity).
local_term(selection).
local_term(causal_effect).
local_term(conditional_association).
```

```

local_term(treatment).
local_term(confounder).
local_term(mediator).
local_term(estimand).

defining_condition(endogeneity,
    "The regressor is correlated with the model's structural error term.").

mechanism(omitted_variables,
    "A price-affecting variable is excluded and correlates with an included regressor; it now sits
    ↪ inside the error.").
mechanism(simultaneity,
    "The regressor is partly caused by, or jointly determined with, the outcome. Least squares
    ↪ returns a blend of both directions.").
mechanism(selection,
    "Sample membership is related to the outcome, so the estimated relationship is conditional on
    ↪ being in the sample.").
mechanism(measurement_error,
    "The regressor is recorded with error; the recording error enters the structural error and
    ↪ biases the coefficient. Under classical error (independent of the true value and of everything
    ↪ else) the coefficient is attenuated toward zero; under non-classical error no general
    ↪ direction follows.").

measurement_error_kind(classical,          attenuates_toward_zero).
measurement_error_kind(non_classical,      no_general_direction).
measurement_error_kind(outcome_dependent,  toward_the_outcome).
measurement_error_of(condition_rating, non_classical).
measurement_error_of(condition_rating_after_price, outcome_dependent).

remedy(omitted_variables, "Measure and include the variable, or bound the bias.").
remedy(simultaneity,      "Find an instrument, or a design that breaks the loop.").
remedy(selection,        "Model the selection, or redefine the population the conclusion is about
    ↪ .").
remedy(measurement_error, "Measure better, or obtain a second independent measurement of the same
    ↪ characteristic.").

generalizes(endogeneity, omitted_variable_bias).
special_case_of(omitted_variable_bias, endogeneity).

shrinks_with_sample_size(imprecision).

distinguishable_from_imprecision(endogeneity) :-
    \+ shrinks_with_sample_size(endogeneity).

role(confounder, "A cause of both the regressor and the price.").
role(mediator,   "The channel through which the regressor acts on the price.").
include_rule(confounder, include, "Its effect otherwise rides on the regressor's coefficient.").
include_rule(mediator,   exclude, "Holding it fixed leaves only the direct effect, a different
    ↪ quantity from the total effect.").

treatment(remodel).
confounder_of(unobserved_quality_u, remodel, price).    % floor plan, micro-location, suitability
mediator_of(condition, remodel, price).
mediator_of(quality,   remodel, price).

estimand(condition_rate,
    "Traditional appraisal: the market's rate for condition, with condition as the regressor; the
    ↪ remodel is how some houses came to their condition and earns no value of its own. RCA-based
    ↪ appraisal never asks this.").

```

```

estimand(remodel_total_effect,
  "What a remodel does to a particular house's price: condition and quality are mediators and
  ↳ must not be held fixed.>").
estimand(remodel_direct_effect,
  "The remodel's effect apart from the condition and quality it produced: what a model holding
  ↳ those fixed estimates; may be small; not the total effect.>").

variable_choice(remodel_total_effect, V, exclude) :- mediator_of(V, remodel, price).
variable_choice(remodel_total_effect, V, include) :- confounder_of(V, remodel, price).
variable_choice(condition_rate, condition, regressor).

encounter(school_quality, simultaneity,
  "School quality can raise prices and prices can raise school quality via tax base and parental
  ↳ resources; the coefficient then recovers neither effect.>").
encounter(renovation_choice, omitted_variables,
  "A remodel is a treatment that is not randomly assigned: owners remodel when they expect the
  ↳ return to justify it, on qualities u that also raise price. u is a confounder the dataset does
  ↳ not measure, so including the remodel indicator cures the omission of the indicator, not the
  ↳ confounding.>").
encounter(comparable_set, selection,
  "Sold properties are selected: sellers who choose their moment push the price one way,
  ↳ distressed and forced sellers the other; this matters when the selection is related to what
  ↳ the model does not see.>").
encounter(condition_rating, measurement_error,
  "Comparables may be rated from photographs after their sale prices are known; if the known
  ↳ price nudges the rating, the regressor is partly a function of the outcome. For a nudge that
  ↳ carries part of the price deviation into the score, the condition coefficient is pushed upward
  ↳ .").
protection(condition_rating,
  "Rate the comparables from the evidence of their condition before looking at the price; record
  ↳ the basis of each rating.>").

bidirectional_with_price(school_quality).
bidirectional_with_price(crime_rate).
bidirectional_with_price(retail_amenity_density).
bidirectional_with_price(tree_canopy).

cures(measuring_the_variable, omitted_variables).

omitted_component(renovation_choice, unobserved_quality_u).
measurable(remodel_numeric).

remains_endogenous_after_inclusion(remodel_numeric) :-
  encounter(renovation_choice, omitted_variables, _),
  omitted_component(renovation_choice, U),
  \+ measurable(U).

coefficient_component(remodel_numeric, what_the_work_produced,
  "60,000: the price difference the constructed data attach to a remodel, holding GLA, lot, view
  ↳ and the condition score fixed; it stands for whatever the remodel brought that the other
  ↳ features do not record.>").
coefficient_component(remodel_numeric, value_of_selected_properties,
  "The unobserved quality of properties whose owners chose to remodel; positive under positive
  ↳ selection, negative under negative selection.>").

coefficient_limit(SelectionDifference, Limit) :-
  Limit is 60000 + SelectionDifference.

consequence(renovation_payback_question,

```

```

    "A coefficient from observational data is the conditional association among the properties that
    ↪ were remodeled; it need not equal the causal effect of remodeling the particular subject,  $P_i$ 
    ↪  $(1) - P_i(0)$ ."
consequence(two_ways_wrong,
    "Reading the selected coefficient overstates the return under positive selection; reading a
    ↪ model that holds condition and quality fixed understates it, to the direct effect alone.").

defensibility_consequence(not_from_the_regression_itself,
    "Residual plots, R-squared, cross-validation and significance tests cannot establish exogeneity
    ↪; formal tests (Hausman) need an instrument or a second estimator, i.e. assumptions the data
    ↪ cannot check.").
defensibility_consequence(state_what_it_is,
    "A coefficient on an endogenous regressor is a well-defined conditional association; reporting
    ↪ it as a causal contribution asserts more than the estimate supports.").
defensibility_consequence(design_beats_estimation,
    "Breaking the loop in how data are assembled does more than any estimator applied afterward; a
    ↪ well-chosen comparable set is itself a design.").

design(before_after_view_obstruction).
design(boundary_discontinuity_school_assignment).
design(externally_forced_renovation).

diagnostic_question(
    "What effect is wanted, and how did this variable come to take the values it took? If a chooser
    ↪ could observe what the model cannot, the coefficient describes the choice as much as the
    ↪ characteristic.").

% Scryer 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)).

```