

內生性 (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)

形式定義。內生性是指當自變項——即迴歸分析中用來解釋價格的特徵——與結構誤差——即價格減去因果關係所預期的價格值（而非擬合模型的殘差）——同步變動時，其估計係數便無法反映所關注的因果效應。其常見原因包括：遺漏某個特徵、價格對該特徵產生反作用、銷售量屬於選樣集，以及該特徵的測量存在誤差。

直觀說明。某個比率之所以可能出現偏差，是因為房產特徵與價格之間存在相互影響。房屋在市場上待售的天數取決於開價：房產特徵在某種程度上是價格的結果，或是其他同樣推高價格的因素所致，例如市場上漲；而根據這些銷售數據推算出的比率，則將這兩種影響方向混為一談。無論是將單一比率套用至某項特徵的整個範圍，還是像 MARS（多變量自適應迴歸樣條，Friedman 1991）那樣，針對特徵的不同區間設定不同比率——這都需要仰賴外部變異來源，或是基於銷售數據無法驗證的假設——因為這種糾纏存在於市場本身，而非算術運算之中。估價師所能做的，是了解該特徵可能隨模型未納入因素而變動的四種方式（詳見下文），在資料集中辨識這些變動模式，並在將該資料集的比率當作「買方為該特徵所支付的價格」來使用之前，先釐清當前究竟是哪一種情況在起作用。

給想了解其運作機制之讀者的參考：當某個特徵與模型中未納入的變項同步變動時，統計學家稱其為「內生變項」；該條目闡明了此條件的精確形式，並將其四大來源區分開來。統計文獻所提供的解決方案（主要是工具變數法），雖適用於一般迴歸分析，但在評估數據中卻極少能運用；該條目中亦有說明。與解決方案不同，此診斷方法適用於任何擬合方法。

此條目具有統計學背景，同時也涉及因果推論，因為它探討了迴歸係數未能捕捉到估價師所期望之因果效應的情況。本條目旨在供那些將係數或配對銷售調整視為因果陳述的估價師參考。與條目 008（遺漏變數偏差）相同，本條目採用建構的「帕西菲卡」市場，該模型如同傳統估價方法般將房屋狀況納入迴歸分析；而《期刊》自身的方法——基於 RCA 的估價法——則未將房屋狀況納入（參見條目 007，條件期望），因此此範例僅作為對照，而非應遵循的範本。

「遺漏變數偏差」（參見第 008 篇〈遺漏變數偏差〉）不應被視為一種具有獨立公式的獨立失誤類型，而應被視為一種更普遍情況的例子：內生性。

回顧條目 007（條件期望）的分解： $P = E[P | \mathbf{Z}] + \varepsilon$ ，其中 $E[P | \mathbf{Z}]$ 是具有特徵 $\mathbf{Z}$ 的房屋平均價格，而 ε 表示某棟房屋的價格相對於該平均價格的高於或低於程度，因此根據定義， $E[\varepsilon | \mathbf{Z}] = 0$ 。無論 $\mathbf{Z}$ 包含哪些特徵，此關係對條件期望函數（即該條目中的 CEF）均成立。若估價師希望獲得因果效應——在房屋其他所有條件確實保持不變的情況下，擁有景觀會使價格提高多少——則需再增加一項條件：自變項不得與任何其他影響價格的因素同步變動。當其中一項自變項與其他影響價格的因素同步變動時，該自變項即為內生變項。將因果關係寫為 $P = \beta_0 + \beta_1 z_1 + \dots + \beta_k z_k + u$ ，其中每個 β_j 代表特徵 z_j 每變動一個單位對價格的影響。其誤差項為結構誤差，寫作 u ：即價格減去因果關係所預測的應有價格。這與 ε 不同，後者是與條件平均值的距離，根據其定義，在 $\mathbf{Z}$ 的條件下，其均值为零； u 可能會隨自變項而變動，而內生性條件正是指此項。

這個術語源自聯立方組建模，在該領域中，若某個變數是在系統內部決定的，而非由外部提供，則稱為「內生變數」。在當前的用法中，它具有更機械化的含義： $\text{Cov}(z_j, u) \neq 0$ （Cov，即共變異數，用以衡量兩個量如何共同變動）——因變數與結構誤差呈同步變動，因此擬合結果（OLS——普通最小二乘回歸——或 MARS）無法區分兩者的貢獻，而將誤差變動的一部分歸因於因變數。

造成這種現象的原因有四種，必須將它們區分開來，因為每種情況所需的應對措施各不相同。

遺漏變數。某個影響價格的特徵被遺漏，且其變動趨勢與某個已納入的特徵同步；此特徵便成為 u 的一部分，進而產生 $\text{Cov}(z_j, u) \neq 0$ 。這正是條目 008（遺漏變數偏差）的情況，其中「翻新指標」被遺漏。解決方法是測量並納入該特徵，或是推算出該偏誤在合理範圍內可能出現的最大值與最小值。

同時性。自變項不僅與價格相關，而且部分是由價格所引起，或與價格一同被解釋；因果關係是雙向的，而由於擬合結果（OLS 或 MARS）並未說明方向，因此會呈現兩者的混合結果。解決之道在於使用工具變數——即一個能驅動該自變數、但與價格無其他關聯（此處指計量經濟學意義上的「變數」，而非法律意義上的正式文件）的變數——或是採用不同的分析設計。

抽樣。該樣本並非從結論所涉及的房屋中隨機抽取，且房屋能否被納入樣本取決於其價格。解決方法是建模說明樣本的抽取方式，或是重新定義結論所針對的群體。

測量誤差。該特徵的記錄不準確——例如兩位檢驗員的評分結果不同，或是居住面積是從紀錄中取得而非實際測量得出。記錄誤差會成為 u 的一部分，並隨記錄值一同變化，因此該係數會產生偏誤。在教科書中的典型案例——即經典測量誤差，此處的誤差與真實值或模型中的其他任何因素皆無關聯——被誤測特徵的係數會被拉向零。房況評分可能並非此類情況：其誤差可能取決於房屋的實際品質，以及評分者觀察到的其他因素。對於此類誤差，偏誤並無明確的方向。解決之道在於改進測量方式，或是針對同一棟房屋的同一項特徵取得第二項獨立測量結果，然後由專為測量誤差設計的方法利用該結果來校正估計值。

這四種機制有一個共同點：增加同類交易的數量，並不能消除它們所造成的誤差。這正是內生性與不精確性（因交易數量有限而導致的估計值散布）之間最明確的實務區別：隨著樣本規模擴大，不精確性會隨之縮小，但內生性卻不會。

在選擇變數之前，請先為該效應命名。某個特徵是否應納入模型，取決於要估計的對象為何，而有兩類特徵很容易被混淆。第一種是同時影響自變數和價格的因素：社區內的優質地段（例如安靜的死胡同、視野開闊的轉角）會吸引昂貴的翻新工程，並憑藉自身優勢就能賣出高價。統計學家將此類特徵稱為「混雜因子」。必須將其控制為常數，否則迴歸變量的係數便會包含其影響。第二種是迴歸變量作用於價格的傳導管道：翻新工程透過改善房況與品質來提升價格。此類特徵稱為中介變數，將其固定後，便能從迴歸變數的係數中精確剔除其透過該途徑產生的影響部分。因此，條目 008（遺漏變數偏差）的翻新案例提出了兩個不同的問題。在傳統估價中，若問題在於市場對房屋狀況的定價，則房屋狀況即為自變量（基於 RCA 的估價從不探討此問題；參見條目 007「條件期望」以了解原因），而翻新僅是部分房屋達到當前狀況的途徑；一旦房屋的狀況與品質已被完整記錄，翻新本身便不具備超越其所產生的狀況與品質之外的價值；而狀況與品質相同的翻新房屋與維護良好的房屋，其價值應是相等的。若問題在於翻新工程對特定房屋價格的影響，則房屋狀況與品質即為中介變項；而將其固定不變的模型所回答的則是另一個問題——翻新工程在產生其所帶來的房屋狀況與品質之外，本身具有多少價值，這部分價值可能微乎其微，且無論如何都與翻新工程的總體影響是不同的量。「納入該變數」對於混雜因子而言是正確的規則，但對於中介變數而言則是錯誤的；在明確界定所需探討的效果之前，兩者皆無法被選定。

估價實務中何處遇到。反向因果關係取決於地點。因同時性所導致的內生性，可透過學校品質這個例子來清楚說明——這是一個真實存在的問題，而非假設性的問題。若以學校評級為自變數進行價格迴歸分析，係數看似顯示買家為學校付出了多少代價。但因果關係亦可能反向作用：房價較高的社區擁有更龐大的稅基，家長參與度更高且資源更充裕，因此更能吸引並留住教師。學校品質可推高房價，而較高的房價亦能提升學校品質。該係數並未捕捉到這兩種效應中的任何一種；它反映的是一種混合效應，取決於該市場中這兩條傳導管道的相對強度。同樣的雙向關係也可能出現在犯罪率、零售設施密度以及樹冠覆蓋率等變數上。在每種情況下，該因素都可能同時是社區價格水準的成因與結果。

裝修是一種「處理」，但並非隨機分配的。在因果推論的術語中，「處理」是指任何其效果被探究的干預措施，而廚房裝修便是其中之一。難點在於誰會接受這項處理。許多屋主在預期回報將超過成本時才會進行翻新，而這種預期是基於他們對自家房屋的了解——這些資訊並未記錄在任何資料集中——例如良好的格局、社區內的優越位置，以及能支撐這項投資的地塊。這些特質會

直接推高房價：它們是翻新工程的混雜因子，且是數據集無法測量的因素。因此，包含翻新指標的模型僅解決了該指標的遺漏問題，卻未消除混雜效應。該指標的係數，既包含翻新對房價的影響，也包含那些選擇翻新的屋主所擁有的房屋在已記錄特徵所能解釋範圍之外的額外價格。納入該變數雖消除了指標的遺漏，卻未能使該自變數成為外生變數——也就是說，仍與結構性誤差存在關聯。

已售房產的樣本並非隨機抽樣。有些人會在時機成熟或價格誘人時出售房產，另一些人則是迫於無奈才出售；因此，已售出的房產與未售出的房產之間，可能存在模型無法察覺的差異；當這種差異存在時，已售房產價格中的特殊因素便非隨機：前者會將價格推向一個方向，而困境銷售與被迫銷售則會將價格推向另一個方向。在已售物件中，公平交易與困境銷售各自構成進一步的非隨機子群組。每個網格中的每項可比物件，皆取自這些經篩選的已售交易。任何估值結論若要延伸至未售物件，必須基於對已售物件篩選方式的假設，而報告應明確說明此假設。

房屋狀況評分可能會受到價格的影響。該評分旨在衡量房屋的狀況，而非其他因素；估價師會在得出任何價值結論之前，根據實地勘查結果對標的物進行評分。可比物件則是另一回事：它們的評分可能是在已知售價後，根據掛牌照片和說明進行評定。若已知的價格影響了評分——例如售價較高的房屋被解讀為維護狀況稍佳——那麼回歸變數便在某種程度上成為結果的函數。這並非房屋狀況本身的特性，而是測量過程的缺陷：即「結果依賴型測量誤差」，也就是價格對評分產生影響。這屬於前述的非經典類型，但有一點不同。對於此類會將價格偏差的一部分帶入評分的微調，其方向是已知的：評分會朝向售價所暗示的分數移動，而狀況係數也會被推高（評分中的其他誤差仍可能產生相反的影響）。防護措施在於程序上：在查看可比物件的成交價格之前，先根據其狀況證據為其評分，並將每項評分的依據記錄在工作檔案中。

對結論可辯護性的意義。一個迴歸變數是否為內生變數，取決於論證，而非一般迴歸分析所能顯示的任何結果。常規的模型擬合檢定方法——無論是殘差圖、解釋價格變動的比例（ R^2 ）、對未納入擬合的銷售數據進行模型驗證（交叉驗證），或是檢定係數是否可能為零（顯著性檢定）——均無法證實自變項為外生變項，因為結構性誤差是不可觀測的，其與自變項之間的關聯性亦然。雖然存在正式的檢定方法——例如豪斯曼檢定（Hausman test）會將常規估計值與透過其他方式獲得的估計值進行比較——但這些方法需要工具變數或第二估計量，也就是說，需要額外的假設，而這些假設無法僅憑數據本身來驗證。但我們總能提供對該機制的合理解釋：為何該自變項可能與誤差相關、相關方向為何，以及相關程度為何。一份合理的報告應闡明此論點，而非透過刻意不提來暗示問題並不存在。

應說明係數的實際數值，而非人們希望它呈現的數值。內生自變量的係數是一個真實且定義明確的數值：在將測量到的特徵值固定不變的情況下，該特徵每單位的擬合價格差異。這並非因果效應。在十六筆未納入翻新指標的帕西非卡（Pacifica）房產交易中（條目 008，遺漏變數偏差），正確的說法應是：「在測量特徵保持不變的情況下，這些擁有景觀的房屋售價高出 \$79,178。」若說「景觀為房產價值貢獻了 \$79,178」，則是在主張一種該估計值無法支持的因果關係。前者經得起質詢；後者則不然。

設計勝於估算。當資料的組織方式能夠打破功能與價格之間的雙向影響時——例如比較視線被遮擋前後的房產特性；利用學區劃分變更但房源結構未變的邊界；或是將因火災或洪水等外部事件所迫進行的翻新，與屋主主動選擇的翻新區分開來——每種設計仍基於必須加以辯護的假設（即同時期無其他變因改變，且該分界線僅區分學區劃分），但若其假設能經得起檢證，其論證力將遠勝於任何基於觀察資料的估計方法。這正是工具變數文獻在實務層面所闡述的觀點，而其與房產估價的關聯在於：一套精心挑選的可比物件本身即構成一種設計。

估價計算示例。Pacifica 資料集雖能說明內生性，卻無法解決此問題，而該資料集之所以無法解決此問題，正是這篇貼文的真正啟示。

首先，問題如下。假設一位估價師被問及，某棟房屋擬進行的 \$45,000 廚房翻新工程，其費用是否能在出售時收回。這是一個關於一棟房屋及其無法同時兼具的兩個價格的問題：翻新後的價格與現狀價格， $P_i(1) - P_i(0)$ ，其中 $P_i(1)$ 代表房屋 i 翻新後的價格， $P_i(0)$ 則代表其未翻新時的價格。這並非關於市場對房屋狀況的定價問題，若在回答時將房屋狀況視為恆定不變，便會犯下

上述所述的「中介人謬誤」。

第 008 筆記錄（省略變數偏誤）的十六項屬性附帶一個「翻新」指標，其係數為 \$60,000：這是建模數據賦予翻新工程的價格差異——相較於擁有相同總可租面積（GLA）、地塊、景觀及狀況評分的房屋，翻新後的房價高出 \$60,000。該 \$60,000 並非翻新行為本身的價格；它代表翻新所帶來的、而其他特徵未記錄的任何價值。在該建構資料集中，哪些房屋經過翻新是由作者決定的。翻新指標在設計上屬於外生變數，與誤差項無關——因為誤差項中已無其他因素——因此，在該資料集中，\$60,000 代表在固定已記錄狀況評分的前提下，翻新為房價帶來的增值；若翻新同時提升了狀況評分，則增值額度會更高。

現在假設這十六項房產資料來自實際市場。屋主選擇進行翻新，且假設——這相當合理——那些具備顯著未觀測特質的房屋（將該價格成分稱為 u ）的屋主，相較於其他條件相似的房屋屋主，更傾向於進行翻新。隨著銷售量累積，包含翻新指標的迴歸式現在收斂的並非 \$60,000，而是

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

其中 $\tilde{r}$ 代表包含的特徵無法預測的翻新指標部分，Var（方差）衡量其變動程度，而標有 p 的箭頭意指「隨著銷售量累積，逐漸趨近於」。若翻新過的房屋平均而言，\$25,000 比可觀察上相似的未翻新房屋擁有更多 u ，則係數將收斂至約 \$85,000，且無論從同一市場取得多少額外數據，都無法減少此過剩值。翻新指標——其省略項 008（省略變數偏差）曾被研究過——現已測量並納入分析，但該係數仍非因果效應：目前仍被忽略的是 u ，而沒有任何資料集記錄此項。

在實務上，這項區別具有實質意義。該係數描述的是平均而言，那些經過翻修的房屋——也就是業主認為其投資回報值得的房屋。本案所指的是一棟特定房屋，其屋主尚未做出決定，而該係數僅在該房屋與那些屋主所知情況相似時，才適用於本案——而這正是模型無法預見之處。若估價師根據 \$85,000 來回答投資回收期問題，便會高估回報率；若評估師依據某個模型來回答此問題，而該模型同時測量並固定了翻新所產生的所有狀況與品質，則會低估回報率，使其僅剩翻新在排除這些因素後所增加的部分；根據第 008 條原則（遺漏變數偏誤），這部分幾乎為零或根本不存在。兩者皆將與翻新相關的條件性價格差異，誤讀為翻新工程本身所產生的效果。

因此，對於任何打算從因果角度解讀的係數，應提出的診斷性問題並非「模型是否指定得當？」，而是「期望產生什麼效果，以及這個變數為何會呈現這些數值？」如果答案涉及某人所作的選擇——而此人能洞察模型無法察覺之事——那麼該係數所描述的，不僅是該特徵本身，更是那項選擇。

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.>").

% Stryer 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)).

```