

參數識別 (Parameter Identification)

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

DOI: [10.67330/xjszh949](https://doi.org/10.67330/xjszh949)

形式定義。當一個參數能根據可觀測資料的聯合分佈（即在所有此類銷售中，各項特徵與價格的每種組合出現的頻率），並在模型假設與規格的條件下被唯一確定時，即被視為可識別。可識別性在邏輯上先於估計，且與估計不同。

直觀說明。有些價格，銷售人員根本無法提供給您，無論採用何種方法都無法改變這一點。倘若市場上每棟經翻修的房屋都處於極佳狀態，而每棟處於極佳狀態的房屋也都經過翻修，那麼銷售數據所反映的僅是兩者結合後的價值，而非任何單一因素的價值；若進行迴歸分析並分別報告兩者的比率——畢竟這兩者幾乎總是相伴而行——所呈現的，不過是市場中極其微小的區分。若每筆可比交易物件的景觀都比標的物更佳，則景觀價格的解讀便已超出數據範圍。若唯一能顯示市場狀況隨時間變化的成交案例，同時也是唯一顯示房貸利率變動的成交案例，則兩者便無法區分。這些是關於銷售資料所包含內容的問題，須在任何擬合之前提出；無論該方法是對特徵的整個範圍擬合單一比率，還是像 MARS（多變量自適應迴歸樣條，Friedman 1991，《期刊》的主要迴歸方法）那樣，對其不同區段擬合不同的比率，答案都是一樣的：MARS 僅能在銷售資料中包含能顯示比率變化的對比時，才能找出比率變化的位置。因此，估價師的首要問題並非「比率是多少」，而是「這些銷售資料能否告訴我」。

給想了解運作機制讀者的參考連結：。此現象在統計學上的術語稱為「識別性」：即若將整個銷售總體進行無誤差測量，是否能藉此確定該參數。普通最小二乘法透過捨棄兩項完全同步變動的特徵之一，使極端情況顯而易見；但同時也掩蓋了危險的情況——即當兩項特徵幾乎同步變動，且計算結果最終僅依賴於同一小部分銷售數據時。這篇文章深入探討了這兩種情況，並涵蓋了外推法以及網格分析中每日都會遇到的混雜時間效應等情境。

本文探討統計學相關內容：它分析在尚未使用任何估計量（例如 OLS 或 MARS 等從銷售數據中計算比率的方法）之前，一組數據在理論上可能呈現的結果，並旨在供正在彙整可比數據集的人士，以及正在擬合迴歸模型的人士參考。

此處的參數是指模型應提供的數值——某個率或係數，據此計算出調整值。估計所提出的問題是：根據樣本，我該將什麼視為該參數的最佳估計值？而識別問題則提出了一個更根本的問題：如果我擁有整個總體——也就是這個市場中所有將要發生的交易，且測量結果毫無誤差——那麼這個參數是否能被確定出來？

如果答案是否定的，那麼該參數即無法辨識，且之後無論採取何種措施都無法糾正此問題。即使採用更精確的估計量也無濟於事，增加同類銷售案例的樣本量亦無助益，而更複雜的模型同樣無濟於事——除非該模型加入了數據無法驗證的假設（如下文所述的因果情況），否則甚至可能掩蓋問題。無法識別參數的問題在於問題本身及所依賴的數據——市場產出的結果與估價師蒐集的資料——而非分析師的方法。

當兩項特徵總是同時出現時，識別失誤最容易被察覺。試想，每間經過翻新的房產都獲得「優秀」（9.00）的評分，而每間獲得「優秀」評分的房產也都經過翻新；在這種情況下，「經過翻新」的房產集合與「評分優秀」的房產集合便完全相同。市場對這類房產的定價更高——相較於未經翻修的「良好」房屋，市場更願意為「評分極佳且經翻修」的房屋支付更高價格。無論我們觀察到多少筆交易，都無法將這價差歸因於「記錄在案的現況」與「翻修所帶來的、但其他特徵未記錄的效益」之間，因為在這個市場中，這兩者永遠是同步變動的。我們知道總體情況；卻無法釐清各個組成部分。

當最小二乘法在完全共線性（即某個特徵是其他特徵的精確線性函數）的情況下能明確指出這一點時，這是因為設計矩陣——也就是由估計器所使用的迴歸變數（解釋變數）的欄所組成的表格——具有秩不足的問題：其其中一欄是其他變數的精確組合。因此，軟體會捨棄其中一個係數，或返回缺失值。這僅僅是一種禮節性的處理。真正危險的情況是「近似多重共線性」——即兩變數幾乎總是同步變動的情況——因為在這種情況下，參數在技術上雖可識別，軟體會提供兩個數值，但這兩者皆基於極少數打破該模式的觀測值。這些參數雖已識別，但由於僅有少數銷售資料打破了模式，其估計值極不精確，以致在實務上無法透過資料將兩者區分開來。這正是識別問題與精確度問題看似相似之處。從識別的角度來看，多重共線性（幾乎完全同步變動的特徵）的本質即在於此。

即使完全不存在共線性，當所討論的參數具有因果關係且資料為觀察性資料時，若未加入資料無法驗證的假設，識別仍可能失敗。第 009 條（內生性）中的例子——即同一批十六筆銷售交易來自實際市場——即屬此類：翻新與價格之間的關聯性可以被識別——因為這是聯合分佈的特徵，且大樣本能準確估計該關聯性——但翻新工程的因果效應卻無法被識別，因為可觀測的分佈中並未包含任何資訊，無法將翻新所增加的價值，與房屋原本就具備且促使屋主進行翻新的額外價值區分開來。要識別因果參數，必須做出一個無法僅憑數據本身進行檢定的假設，因此結論的可靠性取決於該假設的可信度。

估價實務中何處遇到。「區間涵蓋」的要求。該規則要求對標的物進行「區間涵蓋」——也就是說，針對每一項重要特徵，都應將其上、下方的可比物件一併納入考量——此規則被視為一項經驗法則，且它回答了兩個值得區分開來的不同問題。第一個是辨識：這些成交案例在該特徵上是否存在對比，且其變化與其他案例不同？如果每個可比案例的景觀都相同，就無法從中得出景觀的定價基準。第二個問題是支持性：標的物是否位於該費率所涵蓋的範圍內？如果每個可比物件的景觀都優於標的物，雖然可能存在某個費率——無論是來自這些成交案例或更龐大的資料集——但將其套用至標的物，便等於超出了資料範圍進行外推。「區間限定」可確保第二項條件；唯有此類差異才能確保第一項條件。經區間限定的標的物仍可能位於兩項特徵從未出現差異的區域內；而未經區間限定的標的物，仍可採用在其他地方識別出的費率——這正是「可辯護性」章節所允許的「更大資料集」補救措施。

以下是一個在子市場內特徵並非獨立變動的例子：某地塊中，街道景觀側的每棟房屋皆擁有較大的樓面面積；在該地塊中，景觀與面積無法被視為獨立因素。若估價師針對這兩項分別進行調整，實質上等於給出了兩個該地塊數據無法區分的數值，而這兩項調整的總額應如何分配，則取決於估價師的判斷或另一市場。此做法或許恰當，但這與「源自市場」的說法有所不同。

屋齡、屋況及建造年代。這些因素往往密切相關。較舊的房屋通常狀況較差，且建造時的標準也不同。若採用分別針對屋齡與狀況劃設獨立欄位的調整表格，這暗示市場會將這些因素分別定價，且數據能夠區分兩者。現有數據能否確實區分這兩者，是需要驗證而非預設的事項，而數據往往無法通過這項驗證。

時間與市場狀況。此處隱含兩種不同的誤差。若網格中包含一條市場狀況線，卻未針對該期間內其他變動因素（例如利率、庫存、新屋主）設置對應線，則時間係數雖可被識別，但該係數實為複合變量，並非僅由市場狀況單獨所致：整個變動都被歸因於唯一可用的標籤，這正是條目 008（遺漏變數偏差）中的遺漏變數情況。若兩條線均存在，但僅有顯示利率變動的銷售記錄同時顯示市場狀況變動，則兩者之間的區分無法被識別，而本條目即屬此情況。

對結論可辯護性的意義。在提出辨識問題時，必須先於估計問題提出，因為順序至關重要。如果評估者首先思考「我的資料能讓我區分出什麼？」，然後才思考「我該如何最佳地估計它？」，那麼所獲得的調整結果將會是資料所能支持的。反之，若將辨識問題留到最後才提出，結果便是會得出一些無人能實際測量的數量的「可信」數據。

未被識別的參數應如實標示為「未被識別」。當數據能確定某種綜合效應，卻無法釐清其組成部分時，合理的做法並非省略調整，而是如實說明該效應是如何被劃分的：即說明可比組別無法將這兩種效應區分開來、所採用的劃分方式是基於較大的資料集或專業判斷，以及綜合效應的具體數值——因為資料實際上所確定的正是這個綜合效應。如此一來，讀者便能自行判斷該劃分方

式。相較之下，若文中列出兩組數據並聲稱其源自市場，但實際上僅有兩者的總和源自市場，讀者便無法對此進行判斷。

「識別」正是方法論與《統一專業估價實務標準》(USPAP) 交匯之處。USPAP 即《統一專業估價實務標準》。該標準要求估價師必須蒐集、核實並分析所有必要資訊，以確保估價結果具可信度（標準規則 1-4）。貸款機構的指引直接闡明了調整的重點：房利美 (Fannie Mae) 的《銷售指南》要求進行調整，以反映市場對不同物業之間差異的反應。這兩項規則均未規定必須將每條格線與比價物件分別向讀者明確標示；市場反應可能源自格線外的配對成交案例、更廣泛的數據擷取，或是基於已披露的判斷——即上述提出的補救措施。本刊的解讀是，這些規則所做的正是將「識別」問題列為強制性要求：聲稱某項調整有市場依據的估價師，必須能夠指出哪些數據能識別該調整。以此方式界定，該問題便具有可核查性——人們可以檢視所依據的數據，並確認其中是否包含對比——同時又不會將識別的數學測試轉變為合規性測試。

解決之道在於設計，而非計算。若某個參數無法被識別，解決方法是變更所使用的資料——也就是擴大地理範圍、延長時間跨度、找出打破該模式的特徵——或是修改論點。任何估計器都無法從資料中提取其原本不包含的資訊。

估價計算示例。以第 008 項（遺漏變數偏差）中的 16 處房產組成的 Pacifica 資料集為例。其中的「狀況」指標為《期刊》的狀況評分：這是根據該房屋在其定義市場區域內相鄰房屋狀況的比較，對該房屋狀況所作的估計。該評分基於估價師對該市場區域的理解，且不應與在同一市場區域執業的同儕估價師的理解產生重大衝突。評分範圍為 0.00 至 10.00：即該市場區域內狀況較差的房屋所佔百分比，除以十。該資料集記錄了三個等級：平均 (5.00)、良好 (7.00) 和極佳 (9.00)，而建模後的數據每增加一個點，價值即增加 \$20,000。如同條目 008（遺漏變數偏差）與 009（內生性）所述，將房屋狀況納入迴歸分析是傳統估價的做法，而非本期刊所採用的基於 RCA 的估價方法（參見條目 007，條件期望）。將資料集限制在十處位於同一模式的房產——狀況為「良好」且未翻新，或狀況為「極佳」且已翻新：比較物件 A、B、C、D、G、I、K、M、N 及 O。透過兩項篩選條件形成該子樣本，且這兩項篩選發揮不同的作用。剔除 E、H 和 J（那棟從未翻修的「極佳」房產，以及兩棟翻修後評為「良好」的房產）後，便呈現出一個切合實際的模式：當翻修是房產在數據中獲得最高狀況評分（此處為 9.00）的主要原因時，市場會呈現何種樣貌——這種情況確實可能發生。將三棟平均水準的房屋 F、L 和 P 排除在外是人為操作，其目的是讓失敗結果精確發生而非僅是接近；而「接近」的情況——這才是危險的——將在表格之後探討。將這十處房產視為估價師的完整數據世界——一個子市場，在其中，翻新與最高記錄狀態從未分開出現。相對於該樣本群體，這種區分無法被識別；全部十六個案例顯示，錯誤源自該子市場的數據，而非市場本身。

在此子樣本中，記錄狀態與翻新狀態之間存在一個精確的等式： $\text{cond} = 7 + 2 \cdot \text{remodel}$ ，此等式在十處物業的每一處皆成立。在擬合完整規格——總居住面積 (GLA)、地塊、景觀、狀況、翻新——時，軟體必須捨棄這兩欄中的一欄，而捨棄哪一欄將決定輸出結果。以下展示兩種捨棄情況：

	子樣本 (10)		完整範例 (16)
	翻新計畫取消	條件已取消	
攔截	-240,000	110,000	-30,000
GLA	600	600	600
洛特	25	25	25
檢視	100,000	100,000	100,000
條件（每項）	50,000	掉落	20,000
翻新	掉落	100,000	60,000

這張表格中有三點值得注意。

首先，無論是狀態係數還是翻新係數，都不是單純地不精確；無論捨棄哪一個，該項便會缺失，而另一項的數值則僅取決於被捨棄的項。設計矩陣有六列（一列全為 1，分別代表截距、總樓面面積、地塊、景觀、狀況和翻新），但僅有五列攜帶獨立資訊（秩為 5），且最小二乘法無法給出

單一解：直線 $\beta_{\text{cond}} + \frac{1}{2}\beta_{\text{remodel}} = 50,000$ 上的每對點 $(\beta_{\text{cond}}, \beta_{\text{remodel}})$ （每對點皆有其專屬截距），皆能等同且精確地擬合這十筆銷售資料。表格中的兩個子樣本欄位，正是該直線上的兩個點。截距之所以不同，原因相同：每一個截距代表一棟既無居住面積、無地塊、也無景觀的房屋之擬合價格——在其中一個擬合中，其條件評分為零；在另一個擬合中，則為未經翻修的狀態——但實際上沒有任何房屋符合這些條件，這也並非市場事實。

其次，不受跌幅影響的是：一棟經過翻修的優質房屋與一棟未經翻修但狀況良好的房屋之間的價格差異：\$100,000，讀作「\$50,000」，即一欄中兩個點的總和為「\$100,000 表示翻新後的房況，兩者皆精確等同於 $2 \cdot 20,000 + 60,000$ —— 即原始記錄狀態的兩項數值，加上翻新所帶來的所有變動。若有讀者將 \$50,000 與市場的 \$20,000 並列，並稱其為誇大其詞，那是因為該讀者將結果的一種任意呈現方式（即省略翻新部分的版本）當作比率來解讀；而欄位標題已明確說明其性質。這十筆交易精確地（以美元為單位）確定了經翻新且狀況極佳的房屋與未經翻新但狀況良好的房屋之間的價格差異，但並未說明該差異如何分配。請注意，這並非條目 008（遺漏變數偏差）所指的偏差。\$100,000 對於子樣本所能回答的問題而言完全正確；這僅僅是另一個問題罷了。

第三，GLA、地塊面積和視野均恢復為其生成值—— 600、25 和 100,000——與完整樣本中的數值相同。識別是按參數逐一進行的。一個模型可能識別出其大部分係數，卻在三個變數上失敗——此處為「屋況」、「翻新」以及與之相關的截距，其顯示值僅取決於軟體會捨棄哪一組共線性欄位——而這種失敗在成功識別的係數中是無法察覺的。這十項屬性所確定的，是「屋況」與「翻新」這兩項記錄變數的綜合價格差異：\$100,000。

與第 008 篇（遺漏變數偏差）的對比，正是本篇要點所在。在那篇中，所需的資訊其實已存在於資料中，但模型卻未能加以利用：只要加入「重新建模」指標，問題便迎刃而解。在此，即使納入翻新指標也毫無助益，因為該資訊根本不存在於子樣本中。相同的模型規格、相同的估計器、相同的軟體——卻出現了截然不同類型的失誤，唯有透過 $\text{cond} = 7 + 2 \cdot \text{remodel}$ 之外的特性才能解決。這六項預留的房產透過兩種不同方式達成此目的。E（一棟從未翻修的優質房屋），以及 H 和 J（狀況良好的翻修房屋），各自直接區分這兩欄：無論哪種對比，僅憑模型本身的結構，便能從銷售數據本身識別出區分點，且無需對數據範圍外的比率做出任何假設。F、L 和 P 這三棟平均水準的房屋，同樣打破了此等式，但它們僅透過一項假設來區分兩欄：即它們所提供的 5.00 至 7.00 之間的狀況率，同樣適用於 7.00 至 9.00 之間——而所有翻新過的房屋皆位於此區間內。若僅將 F、L 和 P 重新納入，軟體又會返回兩個數值——此處的數值與生成值完全一致，因為數據中不含雜訊——但這兩組數值皆基於三筆銷售資料以及前述假設。這與該條目中稱為「危險」的近似情況極為相似：雖在形式上被識別出來，但仍基於那三筆打破模式的交易，以及「5.00 至 7.00 之間的比率同樣適用於 7.00 至 9.00 之間」的假設。若存在真實的雜訊，這兩個數值將會沿著 $\beta_{\text{cond}} + \frac{1}{2}\beta_{\text{remodel}} = 50,000$ 這條直線一同波動。將 E、H 和 J 重新納入，才是區分其與僅基於銷售數據之差異的關鍵。此修正從未是基於統計學的。

References.

- Fisher, F. M. (1966). *The Identification Problem in Econometrics*. McGraw-Hill.
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Koopmans, T. C. (1949). Identification Problems in Economic Model Construction. *Econometrica*, 17(2), 125–144.
- Manski, C. F. (1995). *Identification Problems in the Social Sciences*. Harvard University Press.

Cross-references: endogeneity; omitted variable bias; conditional expectation.

Addendum: Prolog Formalization

The following machine-readable companion (010-identification.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>.

%% 010-identification.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 2, Entry 010
%% Title: Parameter Identification
%% Author: Bert Craytor
%% Concept: identification
%% Version: 1.2.0
%%
%% 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 knowledge base carries the
%% score and the generating function in those units.
%%
%% The galley reconstructs the entry's identification failure from the
%% dataset itself: not_identified_in_subsample/0 proves the collinearity
%% rather than asserting it.

% 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 -----
% term(identification, "Identification", econometric_method,
% "A parameter is identified when it is uniquely determined by the
% joint distribution of the observable data given the assumptions
% of the model; identification concerns what the data could reveal
% in principle and is logically prior to, and separate from,
% estimation.", published, 'dictionary-2026.1').
%
% depends_on(identification, conditional_expectation).
% depends_on(identification, endogeneity).
% -----

uses_terms([ identification, endogeneity, omitted_variable_bias,
               conditional_expectation, multicollinearity,
               hedonic_regression, comp_selection,
               paired_sales_analysis, adjustment, uspap,
               standards_rule ]).
uses_relations([depends_on/2, related/2, introduced_by/2,
               formalized_by/2, generalized_by/2]).

article_concept(identification).
article_issue(volume(1), number(2), year(2026)).
article_id('foundations.2026.010').

:- 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)).

local_term(rank_deficiency).
local_term(design_matrix).
local_term(bracketing).
local_term(estimand).

% =====
% THE DISTINCTION
% =====

asks(estimation,      "Given this sample, what is the best guess at the parameter?").
asks(identification, "Would the entire population pin the parameter down at all?").

prior_to(identification, estimation).

% What does NOT repair an identification failure.
does_not_repair(better_estimator).
does_not_repair(larger_sample).
does_not_repair(more_sophisticated_model).

% =====
% THE SUBSAMPLE ON WHICH REMODEL IS NOT IDENTIFIED
% =====
%
% Condition score from the dataset's three levels: average 5.00, good 7.00,
% excellent 9.00.
cond_score(P, S) :- attribute_value(P, cond_score, S).

% The generating function, from the knowledge base.
true_coefficient_score(K, V) :- true_coefficient(K, V).

% Restrict to good (7.00) or excellent (9.00) properties, in a submarket
% where every excellent house was remodeled and no good one was (the three
% average homes are set aside to keep the illustration exact). Within the
% restriction, cond = 7 + 2*remodel identically, so the two columns are
% perfectly collinear given an intercept.

in_subsample(P) :-
    property(P),
    cond_score(P, C),
    attribute_value(P, remodel_numeric, R),
    ( C == 9.0 -> R == 1 ; C == 7.0, R == 0 ).

subsample(Ps) :- findall(P, in_subsample(P), Ps).

%! collinear_in_subsample is semidet.
%
% Proves cond = 7 + 2*remodel at every property of the subsample. This is
% the identification failure, derived from the data rather than asserted.

collinear_in_subsample :-
    forall(in_subsample(P),
        ( cond_score(P, C),
          attribute_value(P, remodel_numeric, R),
          C == 7 + 2*R ) ).

%! not_identified_in_subsample(?ParamA, ?ParamB) is semidet.

```

```

not_identified_in_subsample(cond_score, remodel_numeric) :-
    collinear_in_subsample.

% The six properties that BREAK the pattern. Two kinds: E, H and J separate
% the columns directly (an excellent house never remodeled; remodeled houses
% at good); F, L and P separate them only under the assumption that the
% condition rate from 5.00 to 7.00 holds from 7.00 to 9.00.
breaks_pattern(P) :- property(P), \+ in_subsample(P).
direct_contrast(P) :- breaks_pattern(P), cond_score(P, C), C >= 7.0.
linearity_contrast(P) :- breaks_pattern(P), cond_score(P, C), C < 7.0.

% =====
% WHAT THE SUBSAMPLE DOES AND DOES NOT DETERMINE
% =====

% Fitted on the subsample (n = 10) with the remodel column dropped; see
% entry.tex. With the condition column dropped instead: intercept 110,000,
% remodel 100,000, the same three rates. Both lie on
%  $\beta_{\text{cond}} + \beta_{\text{remodel}}/2 = 50,000$ .
subsample_coefficient(intercept,      -240000).
subsample_coefficient(gla_sqft,       600).
subsample_coefficient(lot_sqft,       25).
subsample_coefficient(view_numeric,   100000).
subsample_coefficient(cond_score,     50000). % per point
% remodel_numeric: not estimable (rank deficiency).

not_estimable(remodel_numeric, subsample).
alternative_normalization(condition_dropped, intercept, 110000).
alternative_normalization(condition_dropped, remodel_numeric, 100000).

%! normalizations_agree_on_the_difference is semidet.
% 2 points * 50,000 = 100,000 on the remodel = 2*20,000 + 60,000.
normalizations_agree_on_the_difference :-
    subsample_coefficient(cond_score, S),
    alternative_normalization(condition_dropped, remodel_numeric, M),
    2 * S =:= M,
    true_coefficient_score(cond_score, A), true_coefficient_score(remodel_numeric, B),
    M =:= 2 * A + B.

%! identified_sum_is_correct is semidet.
%
% The subsample's condition coefficient, per point, equals the recorded
% condition's rate plus the remodel's 60,000 spread over the two points
% that separate good from excellent: 20,000 + 60,000/2 = 50,000. Across
% those two points, 100,000 = 2*20,000 + 60,000, the price difference the
% data determine and cannot divide.

identified_sum_is_correct :-
    subsample_coefficient(cond_score, S),
    true_coefficient_score(cond_score, A),
    true_coefficient_score(remodel_numeric, B),
    S =:= A + B / 2,
    2 * S =:= 2 * A + B.

%! identification_is_parameter_by_parameter is semidet.
%
% GLA, lot and view are recovered at their generating values in the
% subsample even though two other parameters are not identified.

```

```

identification_is_parameter_by_parameter :-
    forall(member(R, [gla_sqft, lot_sqft, view_numeric]),
        ( subsample_coefficient(R, X), true_coefficient_score(R, X) )).

% =====
% CONTRAST WITH ENTRY 008
% =====

failure_kind(omitted_variable_bias,
    "The information is present in the data and the model fails to use it. Including the variable
    ↪ fixes it.").
failure_kind(identification,
    "The information is absent from the data relied upon. Including the variable fixes nothing.").

remedy(omitted_variable_bias, include_the_variable).
remedy(identification,      change_the_data_or_the_claim).

% =====
% APPRAISAL ENCOUNTERS
% =====

encounter(bracketing,
    "Bracketing answers two questions: contrast varying apart from the other characteristics (is
    ↪ the rate identified at all?) and support (does the subject lie inside the range where it was
    ↪ found?). Bracketing guards support; only such contrast guards identification.").
encounter(tract_subdivision,
    "Where every view-side home is also the larger plan, view and size cannot be separated within
    ↪ that tract.").
encounter(age_and_condition,
    "Age, condition and era of construction often move together; separate grid lines assert a
    ↪ separation the data may not support.").
encounter(time_adjustment,
    "A market-conditions coefficient is identified but composite: it attributes to time everything
    ↪ else that changed over the period, and is not the effect of market conditions alone.").

% Near-misses are the dangerous case: formally identified, practically not.
dangerous_case(near_collinearity,
    "Two coefficients technically identified but determined by the handful of observations that
    ↪ break the pattern; this is multicollinearity seen from the identification side.").

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

defensibility_consequence(order_of_questions,
    "Ask what the data can distinguish before asking how best to estimate it; the reverse order
    ↪ produces confident numbers for unmeasurable quantities.").
defensibility_consequence(report_the_sum,
    "Where two effects are not separately identified, report the combined effect the data do
    ↪ determine and disclose the basis of any split applied.").
defensibility_consequence(uspap_link,
    "In the Journal's reading, USPAP SR 1-4 (credible results) and the Fannie Mae Selling Guide (
    ↪ market-based adjustments) make the identification question obligatory -- which data identify
    ↪ this adjustment? -- without requiring each grid line to be identified from the comparables in
    ↪ front of the reader.").
defensibility_consequence(remedy_is_design,
    "Widen the geography, extend the period, or find the properties that break the correlation; no
    ↪ estimator recovers absent information.").

```