

参数识别 (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)

形式定义。 当一个参数在给定模型假设和规格的情况下，能够由可观测数据的联合分布（即在此类销售中，特征与价格的每种组合出现的频率）唯一确定时，该参数即被识别。识别在逻辑上先于估计，且与估计不同。

直观说明。 有些数据是销售数据根本无法提供的，无论采用何种方法都无法改变这一点。如果市场上每套翻新过的房屋都状况极佳，而每套状况极佳的房屋也都经过翻新，那么成交数据所报告的只是这两者结合后的价值，而无法体现任何一方单独的价值；一种针对这两者分别给出比率的回归分析——正如人们在两者几乎总是同时出现时所做的那样——实际上是在报告市场几乎未曾发生的分化。如果每套可比实例的景观都优于估价对象，那么景观率的解读就超出了数据的范围。如果唯一能反映市场状况随时间变化的成交案例，恰好也是唯一能反映抵押贷款利率变化的成交案例，那么这两者就无法区分。这些是关于成交数据包含什么内容的问题，需在任何拟合之前提出；无论该方法是拟合特征范围内的单一比率，还是拟合不同区间内的不同比率——如 MARS（多元自适应回归样条，Friedman 1991，《期刊》的主要回归方法）所做的那样——答案都是一样的：MARS 只能在销售数据中包含能体现比率变化的对比项时，才能确定比率发生变化的位置。因此，估价师的首要问题不是“比率是多少”，而是“这些销售数据能否告诉我”。

对于想了解具体机制的读者，请参阅：[识别](#)。这一现象在统计学上的名称是“识别”：即如果对整个销售总体进行无误差测量，是否能够确定该参数。普通最小二乘法通过舍弃两个完全同步变化的特征中的一个，使极端情况显而易见；而当两个特征几乎同步变化，且两个数值最终都落在同一小部分销售数据上时，这种危险情况则会被掩盖。该条目详细探讨了这两种情况，以及网格分析中日常遇到的外推和混杂时间效应等情形。

本文涉及统计学内容：它探讨了一组数据在尚未使用任何估计量（即从销售数据中计算比率的方法，如 OLS 或 MARS）之前，原则上可能呈现出什么样的情况，主要面向正在整理可比数据集以及拟合回归模型的研究人员。

此处的 参数是模型应提供的数值——一个比率或系数，据此计算出调整值。估计所提出的问题是：基于样本，我应该将什么作为该参数的最佳估计值？识别则提出了一个先决问题：如果我拥有整个总体——即该市场中将来会发生的每一笔交易，且测量结果毫无误差——那么该参数是否能够被确定？

如果答案是否定的，那么该参数就无法确定，此后采取任何措施都无法纠正这一问题。改进的估计量无济于事，扩大同类销售样本的规模也无济于事，更复杂的模型同样无济于事——除非该模型添加了数据无法检验的假设（如下文所述的因果情况），否则甚至可能掩盖问题。参数无法识别的问题源于问题本身及所依据的数据——即市场产生的数据和估价师收集的数据——而非分析师的方法。

当两个特征总是同时出现时，识别失误就最容易被察觉。试想，每套经过翻新的房产都获得了“优秀”（9.00）的评分，而每套获得“优秀”评分的房产也都经过了翻新；在这种情况下，“经过翻新”的房产集合与“获得优秀评分”的房产集合完全一致。市场对这一类房产的出价更高——即相较于未翻新的“良好”房屋，人们更愿意为“优秀”且经过翻新的房屋支付更高价格。无论我们观察到多少笔交易，都无法将价格差异归因于记录在案的房屋状况与翻新带来的、但其他特征未记录的价值之间的区别，因为在这个市场上，这两者总是相伴而生，绝不会单独变化。我们知道总和，却无法知晓各部分的具体构成。

当最小二乘法在精确共线性（即一个特征与其他特征的精确线性函数）的情况下能清晰地揭示这一点时，这是因为设计矩阵——即由估计量所使用的自变量（解释变量）列组成的表格——存在秩不足的问题：其其中一列是其他列的精确组合。因此，软件会舍弃一个系数或返回缺失值。这仅仅是一种“礼节性”处理。真正危险的情况是“近似多重共线性”——即两个变量几乎总是同步变化的情况——因为在这种情况下，参数在技术上虽可识别，软件会提供两个数值，但这两个数值均基于少数几个打破该模式的观测值。这些参数虽然已被识别，但由于仅有少数销售数据打破了这种模式，导致参数估计极其不精确，以至于在实践中数据无法将它们区分开来。这正是识别问题与精度问题表现相似之处。从识别的角度来看，多重共线性（几乎完全同步变化的特征）本质上就是这种情况。

即使完全不存在多重共线性，当所涉及的参数具有因果关系且数据为观测数据时，识别仍可能失败，除非添加一个数据无法检验的假设。第 009 条（内生性）中的例子——即同一批十六笔销售交易来自实际市场——就属于这种情况：翻新与价格之间的关联可以被识别——因为这是联合分布的一个特征，且大样本能够准确估计该关联——但翻新工程的因果效应却无法识别，因为可观测分布中不包含任何信息，无法将翻新带来的增值与房屋原本就具备的、促使业主进行翻新的额外价值区分开来。要识别因果参数，必须做出一个无法基于数据本身进行检验的假设，因此结论的可靠性取决于该假设的可信度。

估价实务中何处遇到。“区间比较”的要求。要求对估价对象进行“区间比较”——即针对每一项重要特征，同时纳入其上、下方的可比实例——这一规则被作为一条经验法则提出，它回答了两个值得区分的问题。第一个问题是识别：这些成交案例在该特征上是否存在差异，与其他案例不同？如果每个可比实例的景观视野都相同，就无法从中得出关于景观视野的单价。第二个是支持性：估价对象是否位于该费率所依据的范围之内？如果每个可比实例的景观都优于估价对象，那么可能确实存在一个费率——无论是来自这些成交案例还是更大的数据集——但将其应用于估价对象则属于超出数据范围的外推。“区间限定”可保障第二点；而只有这种差异性才能保障第一点。被区间限定的估价对象仍可能位于两个特征从未出现差异的区域内，而未被区间限定的估价对象仍可采用在其他地方确定的费率——这正是“辩护性”部分所允许的“更数据集”补救措施。

以下是一个特征在子市场内不独立变动的例子：在一个地块中，街道景观一侧的每栋房屋都拥有更大的建筑面积；在该地块内，景观和面积不能被视为独立的因素。如果估价师对这两项分别进行调整，实际上就给出了两个该地块数据无法区分的数值，而这两项调整金额的合并数额如何在两者之间分配，则取决于估价师的判断或另一个市场。这种做法或许是合适的，但这与“源自市场”的表述并不相同。

房龄、房屋状况及建造年代。这些因素往往密切相关。老房子通常状况较差，且建造时的标准也不同。使用分别针对房龄和房屋状况设置不同行数的调整网格，暗示市场对这些因素进行了独立定价，且数据能够区分它们。手头的数据能否真正区分这些因素，需要通过验证而非直接假设，而数据往往无法通过这一验证。

时间与市场状况。这里隐藏着两种不同的偏差。如果数据集中有一条市场状况线，却没有针对该期间内发生变化的其他因素（如利率、库存、新雇主）的相应线，那么时间系数虽然被确定，但它是复合的，并非仅由市场状况所致：整个变化都被归因于唯一可用的标签，这正是条目 008（遗漏变量偏差）所描述的遗漏变量情况。如果两条线都存在，但唯一显示利率变化的销售额恰好是显示市场条件变化的销售额，则二者之间的区分无法确定，而本条目正是这种情况。

对结论可辩护性的意义。在提出识别问题时，必须先于估计问题提出，因为顺序至关重要。如果估价师首先考虑“我的数据能让我区分出什么？”，然后才考虑“我应该如何最好地进行估计？”，那么所得出的调整结果将是数据所能支持的。反之，如果将识别问题放在最后，结果将是针对某些根本无法测量的量得出的“确凿”数值。

未被识别的参数应如实报告。当数据能确定一个综合效应但无法确定其构成时，合理的做法不是省略调整，而是如实说明该效应是如何划分的：即说明可比实例无法将这两种效应区分开来，所采用的划分依据是更大的数据集或主观判断，以及该综合效应的具体数值——因为数据实际上确定的正是这一综合效应。这样，读者便能自行判断这种划分。相比之下，如果给出的两个数据被标注

为“源自市场”，而实际上只有它们的总和才源自市场，读者便无法对此作出判断。

识别是方法论与《统一专业评估实践标准》(USPAP)的交汇点。USPAP即《统一专业评估实践标准》。该标准要求估价师收集、核实并分析所有对得出可信评估结果必不可少的信息(标准规则 1-4)。贷款机构的指导方针直接指出了调整的要点：房利美(Fannie Mae)的《销售指南》要求进行调整，以反映市场对不同房产之间差异的反应。这两项规则均未规定必须将网格中的每一行与可比实例分别向读者明确标示；市场的反应可能源自网格外的配对成交案例、更广泛的样本提取，或是上述补救措施中所述的已披露判断。本刊认为，这些规则的作用是使识别问题成为强制性要求：声称某项调整有市场依据的估价师，必须能够指出是哪项数据支持该调整。以此方式界定，该问题便可核查——人们可以审查所依据的数据，并查看其中是否包含对比信息——同时又不会将识别的数学检验转化为合规性检验。

解决之道在于设计，而在于计算。如果某个参数无法确定，解决办法是改变所用的数据——即扩大地理范围、延长时间跨度、寻找打破该模式的特征——或者修改论断。没有任何估计量能够从数据中提取其本身不包含的信息。

估价计算示例。以第 008 条(遗漏变量偏差)中包含 16 处房产的 Pacifica 数据集为例。其中的“状况”指标是《期刊》的状况评分：该评分是对房屋状况相对于其所处定义市场区域内周边房屋状况的评估。该残差基于估价师对该市场区域的理解，且不应与在同一市场区域执业的同行估价师的理解存在重大冲突。残差范围为 0.00 至 10.00：即该市场区域内状况更差的房屋所占比例除以十。该数据集记录了三个等级：一般(5.00)、良好(7.00)和优秀(9.00)，构建的数据按每点 \$20,000 进行计价。与条目 008(遗漏变量偏差)和 009(内生性)一样，将房屋状况纳入回归分析是传统评估的做法，而非《期刊》采用的基于 RCA 的评估方法(条目 007，条件期望)。将数据集限制为位于同一模式下的十处房产——即“良好”且未翻新，或“优秀”且已翻新：可比实例 A、B、C、D、G、I、K、M、N 和 O。通过两个筛选条件形成该子样本，且它们起着不同的作用。剔除 E、H 和 J(即那套从未翻修过且状况为“良好”的“极佳”房产，以及两套状况为“良好”且已翻修的房产)后，便形成了一个现实的模式：当翻修是房产达到数据中最高状况评分(此处为 9.00)的主要原因时，市场会呈现何种面貌——这种情况确实可能发生。将三套普通住宅 F、L 和 P 排除在外是一种人为操作，其目的是让模型失败的结果是精确的而非近似；而近似的情况——即危险的情况——将在表格之后进行讨论。将这十处房产视为估价师的全部数据世界——一个子市场，在该子市场中，翻新与记录中的最高状况评分从未单独出现。相对于该总体而言，这种划分无法确定；全部十六处房产表明，错误源于该子市场的数据，而非市场本身。

在此子样本中，记录的状况与翻新情况之间存在一个精确的恒等关系：在十处房产中的每一处， $cond = 7 + 2 \cdot remodel$ 。在拟合完整规格(即总居住面积(GLA)、地块、景观、状况、翻新)时，软件必须舍弃两个共线列中的一个，而舍弃哪一个将决定打印结果。以下展示了两种舍弃情况：

	子样本 (10)		完整样本 (16)
	翻新计划取消	条件已取消	
截获	-240,000	110,000	-30,000
GLA	600	600	600
Lot	25	25	25
查看	100,000	100,000	100,000
条件 (每项)	50,000	掉落	20,000
翻新	掉落	100,000	60,000

该表中有三点值得注意。

首先，无论是状态系数还是翻新系数，都不是单纯的不精确；无论舍弃哪一个，该项都会缺失，而另一项的取值则完全取决于被舍弃的项。设计矩阵有六列(截距、GLA、地块、景观、状况和翻新各占一列)，但只有五列承载独立信息(秩为 5)，且最小二乘法无法给出唯一解：直线 $\beta_{cond} + \frac{1}{2} \beta_{remodel} = 50,000$ 上的每一对 $(\beta_{cond}, \beta_{remodel})$ (每对都有各自的截距)都能等好且精确地拟合这十笔销售数据。表格中的两个子样本列正是该直线上的一对点。截距之所以不同，原因相

同：每个截距都代表一栋既无居住面积、也无地块、更无景观的房屋的价格——在一种拟合中其条件评分（condition score）为零，在另一种拟合中则未经过翻新——而实际上没有任何房屋符合这些条件，这也不属于市场事实。

其次，不取决于价格跌幅的是：一栋经过翻新的优质房屋与一栋未翻新的良好房屋之间的价格差异：\$100,000，其中一列中两个数据点之间的差值为 \$50,000，另一列中翻新部分的差值为 \$100,000，无论哪种情况，其差值都精确等于 $2 \cdot 20,000 + 60,000$ ——即记录中房屋原有状况的两个数据点之和，加上翻新带来的全部价值。某位读者将 \$50,000 与市场的 \$20,000 并列，并称其为夸大其词，这其实是将结果的一种任意呈现方式（即省略了翻新部分的那种）误读为比率；而该栏目的标题已明确说明了其含义。这十笔成交记录精确到美元，反映了经过翻新的优质房屋与未翻新的良好房屋之间的价格差异，但并未说明该差异如何分布。请注意，这并非条目 008（遗漏变量偏差）所指的偏差。\$100,000 关于子样本能解答的问题的论述完全正确；这仅仅是一个不同的问题。

第三，GLA、地块面积和视野以它们的生成值——600、25 和 100,000——被恢复，与完整样本中的数值一致。识别是按参数逐一进行的。一个模型可以识别出其大部分系数，却可能在三个系数上识别失败——此处指“房况”、“翻新”以及与之相关的截距，其显示值仅取决于软件从两个共线列中舍弃了哪一列——而这种识别失败在成功识别的系数中是无法察觉的。这十项房产属性所确定的，是“房屋状况”和“翻新”这两个记录变量的两点组合所产生的价格差异：\$100,000。

与第 008 条（遗漏变量偏差）的对比正是本条要阐述的重点。在该例中，所需的信息原本就存在于数据中，但模型未能加以利用：加入 remodel 指示变量后，问题便迎刃而解。而在这里，加入翻新指标却毫无作用，因为该信息在子样本中根本不存在。相同的模型规格、相同的估计量、相同的软件——却出现了完全不同类型的错误，只能通过脱离 $\text{cond} = 7 + 2 \cdot \text{remodel}$ 这条线的属性来解决。预留的六个房产通过两种不同方式实现了这一点。E（一栋从未翻修过的优质房屋）以及 H 和 J（两栋状况良好的翻修房屋）分别直接将两列数据区分开来：仅根据模型本身的结构，这两种对比都能从销售数据本身中识别出分界点，且无需对数据范围之外的比率做出任何假设。F、L 和 P 这三套普通住宅虽然也打破了这种等同关系，但它们仅通过一个假设来区分两列：即它们所提供的 5.00 至 7.00 之间的状况率，同样适用于 7.00 至 9.00 之间——而所有翻新过的房屋都位于这一区间内。若仅将 F、L 和 P 重新纳入，软件又会返回两个数值——此处恰好是生成值，因为数据中不含噪声——但这两个数值均基于三笔交易以及上述假设。这与该条目中被称为“危险”的近似情况极为相似：虽然在形式上被识别出来，但其依据正是那三笔打破模式的交易，以及“5.00 至 7.00 之间的比率在 7.00 至 9.00 之间同样成立”这一假设。若存在真实的噪声，这两个数值将沿着 $\beta_{\text{cond}} + \frac{1}{2}\beta_{\text{remodel}} = 50,000$ 这条线共同波动。重新加入 E、H 和 J 这三笔交易，才将这一情况与仅基于交易数据的情况区分开来。这一修正从来都不是统计性的。

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.").

```