

# 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)

---

**Formal definition.** The conditional expectation of an outcome (here, sale price) given a set of predictors (the measured features) 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.

**Intuitive framing.** An adjustment derived from a regression starts from a rate: what the market pays per square foot, per bath, per year of age. This entry is about where a rate comes from and what it is a rate of. The short answer is that it is the slope of the market’s average price as one feature changes and the others are held where they are, and a regression is a way of measuring that slope from sales. OLS regression measures one slope for the whole range of the feature; MARS (Multivariate Adaptive Regression Splines, Friedman 1991) measures a slope for each range in which the market pays differently, and reports one rate per range. In both cases the rate is a measurement of the same thing, the way average price moves with the feature, and the rest of this entry is about that thing: what it is, why the sales comparison grid measures it too, and what it means when two regressions on the same sales give two different rates.

*For the reader who wants the mechanism.* The “average price as the features vary” has a name, the conditional expectation function, and the entry uses it from here on. It is a surface over the space of features; a rate is its slope in one direction; OLS fits the plane that best approximates the surface and recovers it exactly only when the surface is a plane; MARS fits a surface made of flat pieces that bend at the breaks it finds, and so comes closer when the true surface bends. The eight Pacifica comparables of Issue 1 are the running example.

This entry is about statistics rather than law: it deals with what a regression is able to estimate, and it is intended for appraisers who carry out regressions or who use the sales comparison grid — which also estimates the same thing.

The regression carried out in Issue 1 for the Pacifica comparables was used to derive implicit prices from its coefficients (the rates the regression assigns to the features); an implicit price is the dollars the market pays for one unit of a feature. However, what the regression estimated was not stated clearly, and it is worth mentioning this because almost all of the questions that follow in this issue — concerning bias, endogeneity, and identification — are questions relating to the difference between what was estimated and what was desired.

Regression is typically presented in terms of fitting a line by minimizing the sum of the squared residuals (a residual is a sale’s price minus the price the fitted equation gives) and then giving the slope. While this description is correct regarding the arithmetic involved, it is misleading when it comes to the objective. The arithmetic is only the means; the target is a quantity that describes the whole market (a *population* quantity), not just the sales in hand.

*Reading the notation.*

$P$                       sale price.

|                                     |                                                                                                                              |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| $\mathbf{Z}$                        | the list of a property’s measured features; a list like this is called a <i>vector</i> .                                     |
| $\mathbf{z}$                        | one particular set of values for that list, for example 1,650 square feet, a 7,200 square foot lot, no view, condition 7.00. |
| $E[\ ]$                             | the <i>expected value</i> : the average of all possible values, each weighted by its probability.                            |
| $ $                                 | “given”.                                                                                                                     |
| $E[P \mid \mathbf{Z} = \mathbf{z}]$ | the expected value of price among properties whose features equal $\mathbf{z}$ .                                             |
| $\varepsilon$                       | a sale’s deviation from that expected value.                                                                                 |
| $u$                                 | a sale’s structural error: its price minus what the causal relationship says it should be.                                   |
| $\hat{p}$                           | the price the fitted equation gives.                                                                                         |
| $z_{\text{GLA}}, \dots$             | one feature’s value: GLA and lot in square feet; view and remodel 1 if yes, 0 if no; condition the score 0.00–10.00.         |

Fix the vector of characteristics  $\mathbf{z}$  — that is, 1,650 square feet, a 7,200 square foot lot, no view, and a condition of “good” (7.00 on the Journal’s condition scale, explained in the worked example). Among the properties in the market that have exactly this characteristic vector, the prices differ. The prices are not the same because the properties vary in ways that are not captured by the vector, because buyers and sellers are different, and because negotiation does not always end at the same price. In that market there is a full distribution of prices corresponding to  $\mathbf{z}$ . The average of this distribution is the conditional expectation of price given  $\mathbf{z}$ , written  $E[P \mid \mathbf{Z} = \mathbf{z}]$ ; the general definition is due to Kolmogorov (1933). The average of the sales in hand is an estimate of this expected value, not the expected value itself.

As the characteristics of the properties range over the entire characteristics space, the conditional expectation draws out a surface: at each point there is a number, namely the average price of properties that have those characteristics. The surface is known as the *conditional expectation function*, or CEF. It is a feature of the market and not something belonging to any particular dataset — it would still exist even if no property had ever been appraised. Figure 1 draws it for the running example.

Two facts make the CEF the natural target of estimation. First, of all the functions of  $\mathbf{z}$ , the CEF is the one which minimizes the expected value of the squared prediction error: when you have to guess a price on the basis of only  $\mathbf{z}$  and are judged by squared error,  $E[P \mid \mathbf{Z} = \mathbf{z}]$  is the best guess you can make. Second, any outcome can be expressed in the form

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

the second equation not being an assumption but an identity, since the deviation from the conditional mean has a conditional mean of zero by definition (Angrist & Pischke 2009, ch. 3).

This entry comes before entries 008 (*Omitted Variable Bias*) to 010 (*Parameter Identification*) because each of them asks whether what a regression recovers is the CEF of interest. The identity  $E[\varepsilon \mid \mathbf{Z}] = 0$  itself always holds, because  $\varepsilon$  is defined as the distance from the conditional average. What can fail is the same condition, a conditional mean of zero, for a different error: the structural error, written  $u$  in entry 009 (*Endogeneity*) — the difference between the price and the causal relationship the appraiser wants to measure. That error can move with the features. This happens, for example, when a feature that affects price is left out; when price acts back on a feature — days on market (DOM) usually rise as the price rises, so a regression that uses DOM to explain price picks up some of the price’s own effect on DOM, which is why few analysts include DOM in a regression of net sale price; or when a feature is measured with error. *Omitted Variable Bias* (entry 008) and *Endogeneity* (entry 009) are about these ways; *Parameter Identification* (entry 010) asks the prior question, whether the sales can show the rate at all.

**Where appraisers encounter it.** An appraiser who states that “a home with an area of 1,650 square feet and no view in this area sells for about \$1.28 million” is giving a conditional expectation.

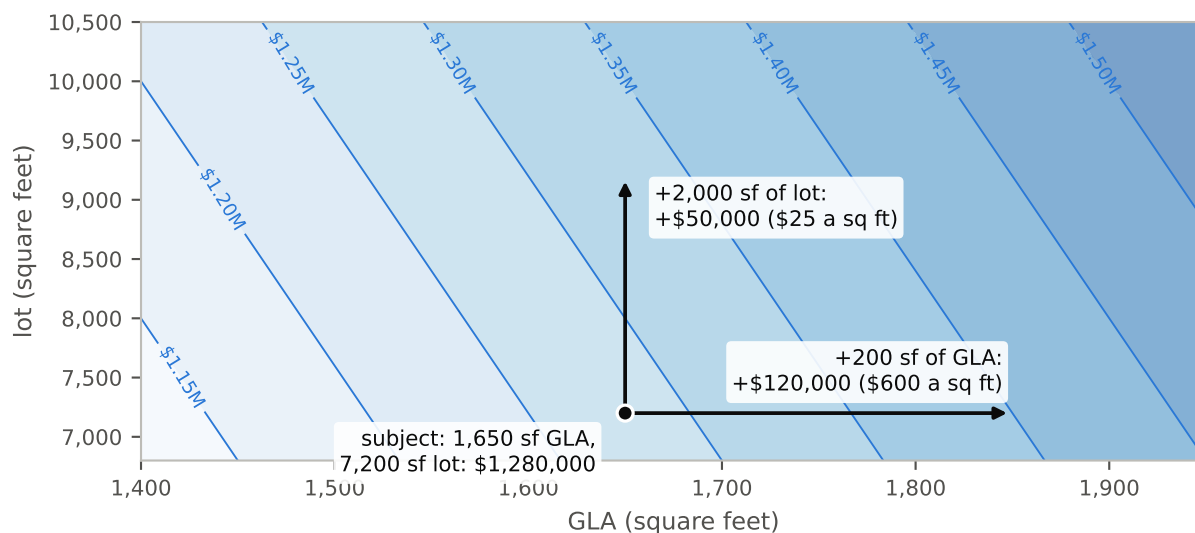


Figure 1: The conditional expectation function of the constructed Pacifica market (worked example), for houses with no view, a condition score of 7.00 and no remodel, drawn like a topographic map: each line joins the GLA and lot combinations whose average price is the same, and the shading darkens as the average price rises. The marked point is the house of this entry, 1,650 square feet of GLA on a 7,200 square foot lot, at \$1,280,000. The two arrows are the two slopes, which are the rates: 200 more square feet of GLA raise the average price by \$120,000, or \$600 a square foot; 2,000 more square feet of lot raise it by \$50,000, or \$25 a square foot.

The word used here is “about”. The statement does not claim that every such house sells for \$1.28 million; rather, it means that \$1.28 million is the average of the prices at which houses of that type sell. By saying “about”, the appraiser acknowledges that individual prices spread around that value.

A sales comparison grid is a method of estimating the value of the CEF at the subject’s characteristic vector; each comparable is an observation taken from the price distribution at that comparable’s characteristic vector; each adjustment shifts that observation along the CEF surface in the direction of the subject’s vector; the reconciled indication then gives an estimate of the height of the CEF at the subject. Structurally, when its adjustments are extracted from the sales, the grid can be viewed as a simple nonparametric estimator (one that assumes no formula for the surface) of a conditional mean — one which employs local averaging and manually specified slopes instead of OLS; adjustments taken from cost or from judgment are structured opinion, not estimates of that kind.

Framing the grid this way brings the appraiser’s unspoken assumptions to light. Whenever comparables are used, it is assumed that the CEF is smooth enough for nearby points to be relevant to the value at the subject’s point. A linear adjustment supposes that the slope of the CEF for that particular characteristic is approximately constant across the range being covered. Giving three indications equal weight assumes that each estimates the same conditional mean, none systematically off, and all with similar precision. None of these assumptions is unusual; they all concern the shape of a surface, and all become topics for discussion once the surface in question has been named.

This framing also ends a common confusion about intervals. There are two kinds of uncertainty.

One is about where the CEF lies: how far the fitted surface may be from the true average. The other is how far individual prices spread around that average. A confidence interval covers only the first; a prediction interval covers both, and so is always the wider of the two. With MARS, the Journal’s principal regression method, the earth package (Milborrow 2026a) gives a prediction interval for a subject from a second model, its variance model (Milborrow 2026b), which estimates how the spread of prices changes across the market. The conditions are that there are enough sales and that the errors are roughly normal and independent of one another; the package reports how often the sales fall inside the interval, and that is the check. In RCA-based appraisal the subject’s market value, per the Sales Comparison Approach, is the model’s estimated sale price plus an appraiser-estimated residual. The variance model gives the interval within which, with probability 95 percent, a sale of the subject would fall, provided those conditions hold. When an appraiser is asked “how sure are you?”, they are usually being asked how far individual prices spread (the prediction interval), and they usually answer how sure they are of the average (the confidence interval).

Naming the estimand — that is, the characteristic of the population being estimated — is a defensibility act, because a conclusion cannot be judged without it. The statement “The adjustment is \$100,000” is not yet one that can be considered right or wrong. It is only when we say “The difference in the conditional mean sale price between properties with a view and those without a view, holding gross living area (GLA), lot, and condition constant, is \$100,000 in this market segment” that we make a claim which can be disputed on the grounds of what is held constant, the population in question, and the units involved.

**Why it matters for defensibility.** Three defensibility consequences follow.

The choice of a conditioning set (the features held constant) is a methodological one: it defines the estimand. Whether the data can estimate that estimand is a separate question, the subject of *Parameter Identification*. The expressions  $E[P \mid \text{GLA, view}]$  and  $E[P \mid \text{GLA, view, condition}]$  represent different functions, which may have different values and different slopes. When an appraiser includes a variable in the grid they are not simply improving the estimate; they are altering the estimand. The reasonable stance is to state the conditioning set on which the value opinion rests.

The expression “holding other things equal” describes how the appraiser reasons in the analysis, and it has a definite meaning. Within the CEF framework it signifies comparing the conditional means at two characteristic vectors which differ in only one feature. This is a statement relating to the population and can be checked against market data when there is enough of it. The analysis behind it usually goes into the workfile — the record of the assignment that USPAP’s Record Keeping Rule requires the appraiser to keep — rather than into the report.

Residual spread is information, not a source of embarrassment. The spread of prices around the conditional expectation function is a characteristic of the market; in a market where prices spread widely around it, any single-figure estimate — no matter what method is used — will be accompanied by great uncertainty. Stating a value without referring to this spread presents the conclusion as more precise than the market warrants.

**Worked appraisal example.** The eight Pacifica comparable prices in Issue 1 are a noiseless sample: they were constructed to satisfy a single linear equation exactly. When these eight prices are regressed on GLA, lot, view, condition and the remodel indicator, every price is reproduced to the dollar — nothing is left over — and the regression returns the equation the prices were built from, rate for rate:

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

Condition here is the condition score of this Journal: an estimate of the house’s condition

in relation to the condition of the surrounding homes in its defined market area. It is based on the appraiser’s understanding of that market area, and should not conflict substantially with the understanding of peer appraisers operating in the same market area. The score runs from 0.00 to 10.00: the percentage of houses in the market area that are in worse condition, divided by ten. A score of 5.00 means half the houses in the market area are in worse condition; 9.00 means nine in ten are. Issue 1 described condition in words and coded it 1–3 for its regression; on this Journal’s score the words map to average 5.00, good 7.00, excellent 9.00, so one Issue 1 step is two points, and the market pays \$20,000 per point. The last term is the price difference the constructed data attach to a remodel: \$60,000 more than a house with the same GLA, lot, view and condition score. It is not a price for the act of remodeling; it stands for whatever the remodel brought that the other features do not record (entry 008, *Omitted Variable Bias*).

This is the CEF for the constructed Pacifica market, precisely. It is important to make clear what kind of object has just been described and why it can be described so exactly. The dataset has been constructed, and a constructed dataset can in fact be noiseless; a real one never is. The advantage of constructing it is that it removes random noise from the prices altogether — every price lies exactly on the formula — so in entries 008 (*Omitted Variable Bias*) to 010 (*Parameter Identification*) no difference between an estimate and its target can come from noise: each is bias or, in *Parameter Identification*, a failure of identification.

The example treats condition as a feature with a dollar rate, as traditional appraisal does and as Issue 1 did, and entries 008 (*Omitted Variable Bias*), 009 (*Endogeneity*) and 010 (*Parameter Identification*) show what goes wrong with it. Traditional appraisal methodology is not the Journal’s method, RCA-based appraisal. In the Residual Constraint Approach the regression takes only the features that can be measured; condition and quality are left out of it on purpose. What the regression leaves unexplained of a sale price is the residual. In the editor’s California markets the measured features commonly explain about 80 percent of the variation in sale prices, leaving about a fifth to the residual. The appraiser divides the residual among condition, quality and the other features the data do not measure, using the appraiser’s scores as a guide, under one constraint: the sum of the residual component value contributions must equal the comparable’s residual. The value conclusion does not depend on how the appraiser divides it — Proof II of the RCA paper (Craytor 2026, p. 36): altering the residual breakdown under the constraint that all residual component value contributions sum to the comparable residual has no impact on the final value conclusion. Condition’s rate is therefore never estimated and never claimed; what the appraiser assigns to condition is condition beyond what the measured features, age among them, already account for.

If we compare the Issue 1 regression with this CEF, the regression can be judged — though one distinction must be made first. Generally speaking, an OLS regression does not reconstruct the conditional expectation function it targets; instead, it finds the best linear approximation to it, that is, the linear projection, and these two agree only in the case where the CEF is linear. A nonparametric method such as MARS does not need the CEF to be linear: it approximates the CEF piecewise, with a hinge (a point where the rate changes) where the surface bends and an interaction term (a rate that depends on another feature) where it twists, and so follows a nonlinear CEF that a linear projection would flatten. It is still an approximation to the same object, and the omitted-variable, endogeneity and identification questions of the next entries attach to it in exactly the same way. The five-variable CEF mentioned above is linear, which is the reason why the regression that includes the remodel indicator is able to recover it exactly. The Issue 1 regression which is conditioned on four characteristics, excluding the remodel indicator, estimates the best linear approximation to  $E[P \mid \text{GLA, lot, view, cond}]$ , and returns, with condition restated on the Journal’s score (Issue 1 printed an intercept of 80,777 and 49,824 per step of its 1–3 coding; it is the same fit),

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

Each coefficient is different from the one above it. Once the regression has been described in terms of line-fitting, there is no further point to make: we have two models and two sets of numbers. However, under the CEF approach two exact statements are available. First, the two regressions condition on different sets of variables and therefore approximate different conditional expectation functions:

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

In a market where some remodeled and unremodeled properties share the same four characteristics, the two functions have different values at those points. In this small artificial market, no two properties have the same four characteristics, so the two functions agree at each property but differ in form: the second is a straight-line formula in its features, while the first is not. In each case, the regression provides only the best linear approximation to its target. Both of the targets are valid objects. Second, only the one with five variables has coefficients that can be read as implicit prices — the remodel coefficient being the price difference between a remodeled and an unremodeled house with the same GLA, lot, view and condition score — and entry 008 (*Omitted Variable Bias*) gives the reason why.

The residual for Comp H from Issue 1 illustrates just how significant the distinction is. If we take the eight properties of Issue 1 as representing the entire market, then each four-variable characteristic vector matches exactly one of the eight properties, and thus the four-variable CEF passes exactly through all eight prices — Comp H being one of them. Comp H has a zero deviation from the four-variable CEF; its Issue 1 residual of \$46,455 is the amount by which it differs from the fitted four-variable linear regression — the linear projection of that eight-property market — and it is precisely in the gap between the projection and the CEF that the remodel’s effect is concealed. With the five-variable specification, the projection and the CEF become identical, and the deviation disappears. The residual did not come from Comp H itself; it came from the specification.

### 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>

---

*Cross-references: hedonic regression; implicit price; omitted variable bias; parameter identification.*

---

### 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
    ↪ 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).
```