

Omitted Variable Bias

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

DOI: [10.67330/460smg10](https://doi.org/10.67330/460smg10)

Formal definition. Omitted variable bias is the systematic distortion introduced into an estimated coefficient when a feature that affects price is left out of the model *and* the left-out feature moves with one of the features that is included — moves with it after the other included features have been allowed for, which is what the auxiliary regression (a second regression, of the left-out feature on the included ones) measures. The left-out feature’s effect then leaks into the included features’ coefficients (the dollars per unit the model assigns to each feature).

Intuitive framing. A rate can be wrong because it is carrying another feature’s value. Issue 1’s Comp H sold for \$46,455 more than the regression on its four recorded features predicted. The reason, disclosed in Issue 1’s *Latent Variable*, was a recent kitchen remodel, which that entry says adds about \$60,000 to the house’s value. Condition was in the feature list, but nothing recorded the remodel, and whatever the remodel brought was missing from the data. That value did not disappear. It was paid out through the rates of the features that were included, because those features happened to go together with it in the sales. Two things have to be true for this to happen: the missing value must matter to price, and it must travel with a feature that is in the list. When it matters but travels with nothing, the rates are right and only the scatter is wider. This is true whether a feature gets one rate or several: MARS (Multivariate Adaptive Regression Splines, Friedman 1991), the Journal’s principal regression method, gives a feature different rates over different stretches of its values, and there the borrowed value is spread across those stretches instead of lodged in one number, but it is there, for the same reason.

For the reader who wants the mechanism. In ordinary regression a rate is a coefficient — the dollars the model assigns to one more unit of a feature — and the borrowed value has an exact size: the left-out feature’s own coefficient times how strongly it moves with the included feature, that second factor being the coefficient of an auxiliary regression of the left-out feature on the included ones. The entry works through that formula, which is the linear case, and then shows the same leakage in the grid, where the appraiser’s paired sales carry it just as the coefficients do. It is composed for appraisers who obtain adjustments from market data.

What this issue’s example is, and is not. The constructed Pacifica market used in entries 007 (*Conditional Expectation*) to 010 (*Parameter Identification*) treats condition as a feature with a dollar rate, as traditional appraisal does and as Issue 1 did, and the example shows what goes wrong with it. Traditional appraisal methodology is not the Journal’s method, RCA-based appraisal — the Residual Constraint Approach of Issue 1 (Craytor 2026) — which keeps condition and quality out of the regression; entry 007 (*Conditional Expectation*) says how. The reader should carry the example’s lesson, not its specification.

Condition, the condition score, and a remodel. Three things are easily run together. *Condition* is a state of the house: the wear of its finishes, the age of its systems, its deferred maintenance. The *condition score* is an estimate of that state, 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. In this Journal it 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. The score is an estimate, and two houses with the same score can differ. A *remodel* is neither a state nor a score but an event: work done, within some recent period, that changed the house. Its value is the value of what it changed — condition, and usually quality, the materials and finishes — at the rates the market pays for those. Most of a remodel’s effect is usually on condition and less on quality, largely because of labor costs, though in some remodels the value added through quality is the higher of the two. A remodeled house and a never-remodeled house that are alike in every present feature that bears on value should be worth the same; the market pays for the condition and quality the remodel produced, not for the fact that a remodel occurred. So a remodel is not a feature with a price of its own. When a dataset carries a remodel indicator beside a condition score, the indicator stands for whatever the remodel brought that the other features in the model (GLA, lot, view and the condition score) do not record: condition beyond what the score shows, and quality, which the dataset does not record. That is how Issue 1’s statement in *Latent Variable*, that Comp H’s remodel “adds approximately \$60,000 to its bundle value,” is to be read, and it is what this entry’s example turns on.

For bias to occur, two things must both be true: the feature that is left out must have an effect on price, and it must move with one of the features that is included. If either is not true, there is no bias.

The two-part requirement matters because the common version of the rule — that “a model which omits an important variable gives wrong answers” — is false as stated. If a feature influences price but does not move with any of the included features, it widens the scatter of prices around the model and so widens the standard errors (the measure of how far an estimated coefficient might sit from the figure a very large sample would give), but it does not shift the coefficients themselves; the intercept (the model’s constant dollar amount) takes up the left-out feature’s average contribution. In this case, precision is lost, not accuracy. On the other hand, a feature that moves with the included ones but has no effect on price costs nothing at all. It is only when both are true that bias occurs.

The way this works is by substitution. If a feature that affects price is missing from the regressors (the features a regression uses to explain price), OLS regression (ordinary least squares) does not simply leave its contribution unexplained. The included features that move with the missing one serve as partial stand-ins for it, and their coefficients take in whatever part of the missing feature’s effect they are able to predict. The coefficient of an included feature is no longer just that feature’s effect; it is that effect together with a share of the missing feature’s effect that it has borrowed.

Bias is always relative to a question. The coefficients of the model without the remodel indicator are the right answer to one question — how price moves with the condition score across these sales, with everything that goes with the score riding along — and the wrong answer when they are read as the market’s dollars per point of the condition score with everything else held fixed, which is how an adjustment reads them. The model is not wrong because another feature exists; its coefficients are wrong for the use they are put to.

The borrowed share has an exact size, given by a formula. For a left-out feature w with true coefficient β_w , the bias in the coefficient of an included feature z_j is

$$\text{bias}(\hat{\beta}_j) = \beta_w \cdot \delta_j,$$

where $\hat{\beta}_j$ — the hat marks an estimate — is the coefficient the regression estimates for z_j , and δ_j is the coefficient of z_j in the auxiliary regression of w on all the included features. This is Theil’s

formula,¹ Computed from the fitted regressions, with the left-out feature’s estimated coefficient, it holds exactly in every sample; with the true coefficient it holds on average.

The product form is the linear case. In a MARS model the left-out feature’s effect is absorbed by whichever of those stretches go together with it, so the bias is spread across them rather than lodged in one coefficient, but it is there, and for the same reason.

The formula is more useful than it appears, since both factors can often be given a sign from market knowledge even if neither can be measured. When the left-out feature raises price ($\beta_w > 0$) and goes with more of an included feature, the other included features being held constant ($\delta_j > 0$), that feature’s coefficient is too high. An appraiser who knows just the direction of both relationships knows the direction of the error. One qualification: δ_j is a relationship with the other included features held constant, and with more than one included feature the simple correlation (a number from -1 to $+1$ measuring how two features rise and fall together, ignoring all others) can go the other way. In the example below, living area goes with the remodel indicator (correlation $+0.09$), yet its δ is negative and its coefficient is therefore too low.

Where appraisers encounter it. The sales comparison grid can be written as an additive model with the same shape as a regression — a handful of features chosen by hand, a dollar adjustment for each — and when its adjustments are extracted from market data it is subject to omitted variable bias in the same way as any regression is. The lines of the grid are the included features. Any feature of a house that influences its price but is not a line is left out.

Features commonly left out include the part of condition that the condition score does not capture, and quality where no quality score is recorded, the functionality of the floor plan, natural light, noise exposure, the specific location within the block, the seller’s motivation, and the marketing effort put into the listing. Each can affect price, and many go with something the grid does take into account.

Remodeling is the instructive case, because it is recorded and still leaks. The URAR (Uniform Residential Appraisal Report) records whether the kitchen and the bathrooms have been updated or remodeled, and within what period, and the UAD (Uniform Appraisal Dataset) condition definitions themselves turn on the extent of updating. The grid, however, has lines for condition and for quality and none for remodel — and rightly so when the question is the market’s rate for condition and the condition and quality lines capture what the remodel brought, since a third line would then count it twice. When the question is instead what a proposed remodel would do to this house’s price, holding condition and quality fixed removes the very channel the remodel acts through; entry 009 (*Endogeneity*) takes that question up. The leakage enters through what the scores miss. In a market where the top condition scores belong mostly to recent remodels, and the remodels bring value that the scores do not record, a condition adjustment extracted from those sales in traditional appraisal — by paired sales, or by a regression that includes condition — carries that value along with it. By Theil’s formula that adjustment is then too high: it includes not only condition’s own effect but a portion of what the scores did not record. An appraiser who takes it and applies it to a top-scored house that earned its score by maintenance will adjust by too much. Under the Journal’s Residual Constraint Approach no condition rate is extracted. The appraiser divides each comparable’s residual among condition, quality and the other features the data do not measure, as the appraiser judges, under one constraint: the sum of the residual component value contributions must equal the comparable’s residual. A remodel’s value is then split where the appraiser sees it, not carried by a single line.

The leakage just described is a more accurate version of the objection appraisers have about

¹Theil, H. (1957). Specification Errors and the Estimation of Economic Relationships. *Revue de l’Institut International de Statistique*, 25(1/3), 41–51.

market-based adjustments, namely that adjustments derived from one dataset do not transfer well. The reason is that an extracted adjustment consists of both the named effect and the unnamed effects that go with it, and how it is composed depends on what went with what in the dataset it was taken from.

Another example is the situation where “paired sales” in traditional appraisal — two sales alike in everything but the feature in question, the price difference being read as that feature’s value — yield an implausibly high figure. The cause may be not measurement error but leakage: the two sales differ in the named feature *and* in something that goes with it, and the price difference is due to both.

Why it matters for defensibility. A report that does not discuss omitted variable bias leaves its hedonic and grid adjustments open to that objection.

The bias cannot be picked up from the fit statistics. A model can have a high R^2 (the share of the variation in prices that the model reproduces), small residuals (the differences between the sale prices and the model’s prices), and coefficients that pass the usual statistical test (a t -test, which asks whether a coefficient could plausibly be zero), and still be badly biased, because bias comes from what moves with what among the features, not from the size of the residuals. Using R^2 as proof that the model is reliable does not answer the question. In the example below, R^2 is 0.97 and the condition coefficient is wrong by 74 percent.

More data won’t solve the problem. Bias does not shrink as the sample grows. A thousand sales from the same market with the same feature left out give an estimate centred on the same wrong figure, estimated more tightly: the standard error shrinks around the biased estimate, which is worse than a wide interval around it, since the report will appear more confident. The same holds for MARS: more sales make the fitted surface steadier, but around the same biased answer; earth’s `varmod` interval measures the spread of prices, not bias. This corrects Issue 1’s *Latent Variable*, which said the bias would shrink with more comps and that the other coefficients were biased upward. Bias does not shrink with more sales, and in Issue 1’s own eight sales, as in the sixteen below, the view, GLA and lot coefficients are biased downward; only the condition coefficient is biased upward.

It is often possible to justify the direction of a bias even when its size cannot be known. In traditional appraisal, where a condition adjustment is extracted from sales — by paired sales or by a regression that includes condition — the practical approach is not to assert that the adjustment is unbiased, but to name the plausible left-out features, use market knowledge to give each a sign for β_w and for δ_j , and state the resulting direction of the bias. “The condition adjustment is probably overstated. Most of the top-scored sales are recent remodels, and the remodels brought improvements the data do not record, such as new kitchens and finishes. Those improvements raise price and occur mostly in the high-scored sales, so part of their value has been counted as condition” is a reasonable statement. The two requirements are both met, and both point the same way: the left-out improvements raise price (β_w is positive) and go with high scores (δ_j is positive), so the bias is upward. Such a statement turns an unmeasured problem into one that is acknowledged and given a direction, which is exactly what a reader needs in order to weigh the conclusion. Under the Residual Constraint Approach no condition rate is extracted, and the question does not arise for condition.

Worked appraisal example. The sixteen-property extension of the Pacifica dataset records, for each property, gross living area (GLA), lot size, view, the condition score, an observed remodel indicator, and the sale price. Because every price was computed from a known formula with no random part (the data are noiseless), the regression that uses the right features recovers that formula exactly, and any deviation from it is bias with no noise component at all.

Comp	GLA (sf)	Lot (sf)	View	Cond	Remodeled	Price (\$)
A	1,650	9,000	Yes	7.00	No	1,425,000
B	1,650	7,200	No	7.00	No	1,280,000
C	1,850	7,500	No	7.00	No	1,407,500
D	1,450	7,000	No	7.00	No	1,155,000
E	1,700	8,000	Yes	9.00	No	1,470,000
F	1,550	7,100	No	5.00	No	1,177,500
G	1,700	10,500	No	7.00	No	1,392,500
H	1,600	7,300	No	7.00	Yes	1,312,500
I	1,750	8,200	No	9.00	Yes	1,465,000
J	1,600	7,600	No	7.00	Yes	1,320,000
K	1,900	9,500	Yes	9.00	Yes	1,687,500
L	1,500	6,800	No	5.00	No	1,140,000
M	1,800	8,800	No	7.00	No	1,410,000
N	1,650	7,400	Yes	7.00	No	1,385,000
O	1,550	7,900	No	9.00	Yes	1,337,500
P	1,720	9,200	No	5.00	No	1,332,000

Comps A–H are Issue 1’s, with their published attributes and prices unchanged; 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. Comp H’s remodel — disclosed in Issue 1 as the source of its residual, but not recorded as a variable — is now recorded. The market pays \$20,000 per point of condition. In the constructed data a remodeled house sells for \$60,000 more than a house with the same GLA, lot, view and condition score. That \$60,000 is not a price for the act of remodeling; it stands for whatever the remodel brought that the other features do not record. Remodeled houses also have higher scores on average (8.20 against 6.64), so the remodel indicator moves with the included condition score, which is the second thing bias requires: $\text{corr}(\text{remodel}, \text{cond}) = 0.55$.

Two regressions on the same sixteen properties:

	Remodel omitted	Remodel included	Bias
Intercept	−83,923.57	−30,000	−53,923.57
GLA	586.64	600	−13.36
Lot	24.32	25	−0.68
View	79,178.46	100,000	−20,821.54
Condition (per point)	34,815.56	20,000	+14,815.56
Remodel	—	60,000	
R^2	0.9733	1.0000	

The \$60,000 on the remodel line is not a price for remodeling. In this constructed dataset it is the price difference attached to the remodel indicator at equal GLA, lot, view and score; that is what “recovers the generating function” recovers.

The condition coefficient is too high by \$14,816 a point — it reports \$34,816 per point of condition where the market’s figure is \$20,000, an overstatement of 74 percent. The view coefficient is too low by \$20,822. Both errors are invisible in the fit: the model without the remodel indicator has $R^2 = 0.9733$ and residuals that, apart from three sales near \$35,000, look ordinary. Nothing in R^2 or in the tests on the coefficients signals trouble (had these residuals been noise, every coefficient would have passed that test at the usual 5 percent level).

Theil's formula accounts for the numbers exactly. Regressing the remodel indicator on the four included features gives $\delta_{\text{cond}} = 0.24692602$ and $\delta_{\text{view}} = -0.34702564$. Then

$$\beta_w \cdot \delta_{\text{cond}} = 60,000 \times 0.24692602 = 14,815.56, \quad \beta_w \cdot \delta_{\text{view}} = 60,000 \times (-0.34702564) = -20,821.54,$$

matching the observed shifts to the cent. The bias is not approximately explained by the formula; it *is* the formula. Noise would change less than one might expect: with noisy data the formula still holds exactly between the two fitted regressions (without and with the remodel indicator), using the indicator's estimated coefficient $\hat{\beta}_w$ in place of β_w ; against the true coefficient it holds on average over many samples.

The view coefficient repays a second look, because its bias runs the other way and the reason is instructive. With the other features held constant, view properties in this sample are *less* likely to be remodeled ($\delta_{\text{view}} < 0$). So the view coefficient is pushed down rather than up: it is \$20,822 too low, while the condition coefficient is \$14,816 too high. The two errors are separate effects of the same omission, not a transfer between them. An appraiser reasoning only about condition would not have anticipated an error in the view adjustment at all — yet a view adjustment understated by 21 percent is the direct consequence of the same omission.

Practically, the damage is concentrated where appraisers are least likely to look for it. Consider first what the biased model does as a *predictor*. At an unremodeled subject identical to Comp N — 1,650 square feet of GLA on a 7,400 square foot lot with a view and a condition score of 7.00 — it indicates \$1,386,914 against a true average price of \$1,385,000 for houses with exactly these features — from the five-feature formula of entry 007 (*Conditional Expectation*), with remodel = 0 — an error of \$1,914, or 0.14 percent. Across all sixteen sales the typical error is about \$21,000, so the model is badly biased and still predicts tolerably, which is not a coincidence: OLS fits the sample it is given, and the biased coefficients are precisely the ones that come closest to the sample's prices. Prediction accuracy is no evidence against omitted variable bias, and cross-validation — testing the model on sales held back from the fit — does not rule it out either: good cross-validation results do not show that the coefficients are free of bias, because a misspecified model can predict well wherever the features go together as they did in the fit. Leaving each property out in turn, the two largest errors — about \$56,000 at K and \$48,000 at G — fall on properties that follow the pattern of this sample, in which the high condition scores mostly belong to remodeled houses; E, which breaks it, is third at about \$45,000, and H and J, which also break it, rank fourth and fifth at about \$41,000. The data here carry no noise, so every one of these errors comes from the left-out remodel indicator, none from noise; in real sales, where noise is present too, nothing in an error's size says whether it comes from a left-out feature or from ordinary residual spread.

The damage appears the moment the coefficients are used as *adjustments* rather than as a prediction rule. An appraiser extracting a condition adjustment from this model applies \$34,816 per point where the market's figure is \$20,000 — \$29,631 too much on every comp adjusted across the two points from good to excellent — and a view adjustment of \$79,178 where the market's is \$100,000. The grid is where a regression's coefficients get read one at a time, and that is exactly the reading that bias destroys.

Prediction also fails, and badly, at the properties that break the pattern the bias is built on. A well-maintained home like the subject (1,650 square feet of GLA, 7,400 square foot lot) but with no view, scored excellent, 9.00, and never remodeled, is indicated at \$1,377,367 against a truth of \$1,325,000, overvalued by \$52,367 (3.95 percent), because, in these sales, a 9.00 score mostly came with a remodel, and the model's condition rate carries the remodel's \$60,000 along with the score. A remodeled home otherwise identical to the subject (a view, score 7.00) is indicated at \$1,386,914 against \$1,445,000, undervalued by \$58,086 (4.02 percent). These are the atypical properties —

the ones an appraiser is most often specifically retained to value, and the ones for which the report will be most closely read.

References.

- 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>
- Theil, H. (1957). Specification Errors and the Estimation of Economic Relationships. *Revue de l'Institut International de Statistique*, 25(1/3), 41–51.

Cross-references: conditional expectation; hedonic regression; implicit price; latent variable; endogeneity.

Addendum: Prolog Formalization

The following machine-readable companion (008-omitted-variable-bias.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>.

%% 008-omitted-variable-bias.pl -- companion to Foundations Vol. 1, Issue 2,
%% Entry 008, Omitted Variable Bias (omitted_variable_bias), v1.2.0.
%% Checks: theil_formula_holds/0 (each bias = beta_omitted * delta_j to the
%% cent); remodel_reading_holds/0 (the remodel indicator's 60,000).

% The dictionary (Issues/tools/dictionary) is not loaded at run time, so
% that the file runs unchanged in Trealla, SWI-Prolog and Scryer. The
% declarations below are plain facts, read by the dictionary linter.
dictionary_release_pinned('dictionary-2026.1').

uses_terms([ omitted_variable_bias, conditional_expectation,
             hedonic_regression, implicit_price, latent_variable,
             multicollinearity, ordinary_least_squares, residual,
             adjustment, comp_selection, paired_sales_analysis,
             sales_comparison_approach, uspap ]).
uses_relations([depends_on/2, related/2, introduced_by/2]).

article_concept(omitted_variable_bias).
article_issue(volume(1), number(2), year(2026)).
article_id('foundations.2026.008').

:- initialization(consult('../..kb/pacifica_comps_extended')).
% 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)).

kind(condition,          state).
kind(condition_score,    measurement_of(condition)).
kind(quality,           state).
kind(remodel,           event).
changes(remodel, condition).
changes(remodel, quality).

stand_in(remodel_numeric, unrecorded_condition).
stand_in(remodel_numeric, unrecorded_quality).

ovb_condition(affects_outcome,
  "The omitted variable must have a nonzero effect on the outcome.").
ovb_condition(correlated_with_included,
  "The omitted variable must have a nonzero coefficient on at least one included regressor in the
  ↪ auxiliary regression (partial, not merely marginal, correlation).").

biased(BetaOmitted, Delta) :-
  BetaOmitted =\= 0,
  Delta =\= 0.

consequence(affects_but_uncorrelated,
  "Inflates residual variance and widens standard errors; slope coefficients remain unbiased (the
  ↪ intercept absorbs the omitted variable's mean).").
consequence(correlated_but_no_effect,
  "No cost at all.").

theil_bias(BetaOmitted, Delta, Bias) :-
  Bias is BetaOmitted * Delta.

omitted_variable(remodel_numeric).
omitted_variable_effect(remodel_numeric, 60000).

condition_scale(0.0, 10.0).
condition_score(average, 5.0).
condition_score(good, 7.0).
condition_score(excellent, 9.0).
condition_rate(20000).

coefficient(remodel_omitted, intercept, -83923.571).
coefficient(remodel_omitted, gla_sqft, 586.645).
coefficient(remodel_omitted, lot_sqft, 24.323).
coefficient(remodel_omitted, view_numeric, 79178.461).
coefficient(remodel_omitted, cond_score, 34815.561).

coefficient(remodel_included, intercept, -30000).
coefficient(remodel_included, gla_sqft, 600).
coefficient(remodel_included, lot_sqft, 25).
coefficient(remodel_included, view_numeric, 100000).
coefficient(remodel_included, cond_score, 20000).
coefficient(remodel_included, remodel_numeric, 60000).

aux_slope(intercept, -0.89872625).
aux_slope(gla_sqft, -0.00022259).
aux_slope(lot_sqft, -0.00001129).
aux_slope(view_numeric, -0.34702564).
aux_slope(cond_score, 0.24692602).

```

```

model_fit(remodel_omitted, r_squared, 0.9733).
model_fit(remodel_included, r_squared, 1.0000).

observed_bias(R, Bias) :-
    coefficient(remodel_omitted, R, Omitted),
    coefficient(remodel_included, R, True),
    Bias is Omitted - True.

theil_formula_holds :-
    omitted_variable_effect(_, BetaOmitted),
    forall(( observed_bias(R, Observed), aux_slope(R, Delta) ),
        ( theil_bias(BetaOmitted, Delta, Predicted),
          abs(Observed - Predicted) < 0.01 )).

remodel_reading_holds :-
    omitted_variable_effect(remodel_numeric, Beta),
    coefficient(remodel_included, remodel_numeric, Coef),
    abs(Coef - Beta) < 0.01.

reported_bias(cond_score, 14815.56).
reported_bias(view_numeric, -20821.54).

relative_bias(R, Pct) :-
    observed_bias(R, Bias),
    coefficient(remodel_included, R, True),
    True =\= 0,
    Pct is 100 * Bias / True.

use_of_coefficients(prediction,
    "Evaluate the whole fitted surface at a point; biases in individual coefficients can offset.").
use_of_coefficients(adjustment,
    "Read one coefficient at a time as an implicit price; bias enters directly and does not offset
    ↪ .").

prediction_check(subject_typical, remodel_omitted, 1386914).
prediction_check(subject_typical, remodel_included, 1385000).

prediction_check(excellent_unremodeled, remodel_omitted, 1377367).
prediction_check(excellent_unremodeled, remodel_included, 1325000).
prediction_check(remodeled_with_view, remodel_omitted, 1386914).
prediction_check(remodeled_with_view, remodel_included, 1445000).

atypical_property_fails_worse(S) :-
    prediction_check(S, remodel_omitted, A),
    prediction_check(S, remodel_included, B),
    abs(A - B) > 50000.

defensibility_consequence(fit_statistics,
    "Bias is undetectable from R-squared or residual size; it lives in the correlation structure,
    ↪ not the residuals.").
defensibility_consequence(sample_size,
    "Bias does not shrink with n; more data gives a tightly estimated wrong number.").
defensibility_consequence(direction_disclosure,
    "Signing beta_omitted and delta_j from market knowledge yields a defensible directional
    ↪ statement even when magnitude is unknown.").

practice_encounter(what_the_scores_miss,
    "Traditional appraisal: a condition adjustment extracted where top scores go with remodels

```

```

    ↪ carries their unrecorded value. RCA extracts no condition rate.").
practice_encounter(no_remodel_line,
    "The URAR records kitchen and bathroom updates; the grid has condition and quality lines and no
    ↪ remodel line, rightly: a remodel line would count the same value twice.").
practice_encounter(adjustments_do_not_travel,
    "An extracted adjustment is a composite of the named effect and unnamed effects correlated with
    ↪ it, so it is dataset-specific.").
practice_encounter(implausible_paired_sale,
    "A paired-sales adjustment far larger than cost can reflect leakage from a correlated omitted
    ↪ attribute.").

```