

## Latent Variable

*Foundations • The Valuation Engineer*

Bert Craytor • May 29, 2026

Version 0.1.0-draft

---

**Formal definition.** A latent variable is an unobserved attribute that influences observed outcomes, whose value must be inferred from its effects on observable variables rather than measured directly.

**Intuitive framing.** The first five entries treated the characteristics vector  $\mathbf{z}$  as if it could be fully measured: GLA, lot, view, condition, and so on, all available as columns of data. In practice, no characteristics vector is ever complete. Some attributes are not in the data because they are difficult to measure (interior finish quality, layout flow, micro-locational micro-orientation), because they are absent from the standard data feeds (MLS records do not capture lighting quality, soundproofing, or recent unpermitted improvements), or because they have not yet been recognized as relevant (neighborhood-level features that turn out to predict price only after analysis reveals them).

These unobserved-but-relevant attributes are *latent variables*. They exist, they influence price, but they are not in the columns of our regression. Their effects do not vanish because they are unmeasured; they accumulate in the model’s residuals, where they appear as “noise” from the perspective of the included variables but are in fact structured signal from the perspective of the underlying data-generating process.

Latent variables are the bridge between the idealized hedonic framework and the messy reality of appraisal data. Recognizing them explicitly is a precondition for honest reasoning about model uncertainty, residual structure, and the limits of what observable characteristics can explain.

**Where appraisers encounter it.** Appraisers encounter latent variables in three recurring contexts.

**The quality and condition adjustments.** “Quality” and “condition” as appraisal categories are themselves attempts to collapse a multi-dimensional latent construct into a single ordinal scale. A C3 condition rating in UAD 3.6 is meant to summarize a property’s deferred maintenance, finish wear, mechanical-systems age, and presentation — four to ten underlying attributes condensed into one number. The single-number summary is a proxy for a latent construct; the appraiser’s craft includes judgment about how well the proxy captures the underlying truth in each case.

**The unexplained residual in the grid.** When three adjusted comps indicate value but disagree by more than expected, the disagreement is informative. It is rarely all sampling noise; more often it reflects an attribute the appraiser has not yet identified or adjusted for. Investigating the residual disagreement — “why does this comp indicate \$50,000 above the others?” — is the appraiser’s informal version of latent-variable inference.

**The recent improvement that nobody can see.** A property may have undergone a kitchen remodel, a roof replacement, or a permitted ADU buildout that is not yet reflected in MLS or assessor data. From the appraiser’s perspective, the improvement is observable in person but not in the data. From the regression’s perspective, the improvement is latent. The appraiser must reconcile the in-person observation with the data-driven indication, and both should converge on the latent attribute’s true contribution.

**Why it matters for defensibility.** Latent variables introduce several defensibility considerations that the additive linear model in entry 005 does not address directly.

**Omitted-variable bias.** If a latent variable is correlated with an observed variable, its effect leaks into the observed variable’s coefficient. If recently sold Pacifica homes tend to have remodeled

kitchens (a latent attribute) and the contract date of the home sale is in the data, then the age of the sale (aka `sale_age` or "days since sold") coefficient will absorb some of the kitchen-remodel effect. The estimated implicit price contribution of `sale_age` becomes biased — it reflects `sale_age` plus a portion of the unmeasured kitchen-remodel status. Defensible appraisal acknowledges that observed-variable coefficients are conditional on what is missing from the model.

**Residual interpretation.** Standard regression theory treats residuals as Gaussian noise with zero mean. The latent-variable view treats residuals as a mixture of noise and signal: part is sampling variation, part is the cumulative effect of unobserved attributes. Identifying which is which is methodologically substantive. A residual that correlates with location, with sale date, with appraiser identity, or with any other observable suggests that the corresponding latent variable can be partially recovered.

**Model selection.** Adding more observed variables is the obvious remedy for omitted-variable bias, but the latent variables that matter most are often the ones that are hardest to add to the model (condition fine-grain, quality of recent improvements, neighborhood micro-effects). The defensible response is sometimes to recognize that the model will never capture them and to allocate the unexplained variance to the residual rather than to force-fit additional variables.

**Worked appraisal example.** The regression in entry 005 produced an unusual residual for Comp H: the model under-predicted Comp H's price by \$46,455 while the other seven residuals fell within a  $\pm$ \$18,000 band. The observed characteristics of Comp H — GLA 1,600, lot 7,300, no view, condition "good" — place it firmly in the middle of the sample; nothing about its observable profile predicts an unusually high price. And yet it sold for \$46,455 more than the model says it should have.

The natural hypothesis is that Comp H has a latent attribute that contributes to its price but does not appear in the data. In the construction of the dataset for this issue, this is exactly the case: Comp H underwent a recent kitchen remodel that adds approximately \$60,000 to its bundle value. The remodel was not in the regression's variable list, so the contribution accumulated in the residual rather than in a coefficient. The estimated residual (\$46,455) is slightly smaller than the true latent contribution (\$60,000) because regression spreads small portions of the latent effect across the other coefficients, biasing them upward in correlated dimensions.

Three operational consequences follow:

**Treat the outlier residual as evidence, not error.** The first appropriate response to Comp H's residual is not to drop it as an outlier but to investigate what about the property explains the under-prediction. In this case, an in-person inspection or a permit records search would reveal the remodel. The model's failure has diagnostic value.

**Recognize that the other coefficients are slightly biased.** The view, condition, and lot coefficients have absorbed small fractions of the remodel effect via correlations with Comp H's observed characteristics. The bias is small at  $n = 8$  and would shrink further with more comps and more covariates, but it is not zero.

**The residuals as a whole carry recoverable structure.** This is the substantive claim of the Residual Constraint Approach (RCA) methodology developed by the author (Craytor, 2025): in residential appraisal data, residuals are systematically not iid Gaussian noise. They reflect the structured contribution of unobserved attributes — condition fine-grain, quality, functional utility, and other latent constructs — that can be partially recovered by treating the residual itself as a latent-variable signal to be modeled rather than discarded.

The single-comp illustration here is the simplest case of the RCA intuition. With more data and richer modeling, the structure in the residual can be partitioned into recognizable latent factors, each contributing to the bundle's market price in ways that the standard hedonic specification

cannot reach. Latent variables are not a nuisance to be controlled away; they are the next frontier of defensible valuation modeling.

*Reference.* Craytor, B. (2025). Residual Constraint Approach for Real Estate Valuation. Zenodo. doi:10.5281/zenodo.14787917

---

*Cross-references:* *characteristics space; hedonic price function; implicit price; hedonic regression.*