# Variable latente

*Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.*

*Foundations • The Valuation Engineer*

Bert Craytor• 1 de agosto de 2026

Versión 1.0.0
DOI: 10.67330/11n3g168

---

**Definición formal.** Una variable latente es un atributo no observado que influye en los resultados observados, y cuyo valor debe inferirse a partir de sus efectos sobre las variables observables en lugar de medirse directamente.

**Planteamiento intuitivo.** Las primeras cinco entradas trataron el vector de características **z** como si pudiera medirse por completo: GLA, lote, vista, estado de conservación, etcétera, todo disponible como columnas de datos. En la práctica, ningún vector de características está jamás completo. Algunos atributos no están en los datos porque son difíciles de medir (la calidad de los acabados interiores, la fluidez de la distribución, la micro-orientación micro-locacional), porque están ausentes de los flujos de datos estándar (los registros de la MLS no capturan la calidad de la iluminación, la insonorización ni las mejoras recientes sin permiso), o porque todavía no se han reconocido como relevantes (rasgos a nivel de vecindario que resultan predecir el precio solo después de que el análisis los revela).

Estos atributos no-observados-pero-relevantes son *variables latentes*. Existen, influyen en el precio, pero no están en las columnas de nuestra regresión. Sus efectos no se desvanecen por no estar medidos; se acumulan en los residuos del modelo, donde aparecen como "ruido" desde la perspectiva de las variables incluidas pero son en realidad señal estructurada desde la perspectiva del proceso generador de datos subyacente.

Las variables latentes son el puente entre el marco hedónico idealizado y la realidad desordenada de los datos de tasación. Reconocerlas explícitamente es una precondición para razonar con honestidad sobre la incertidumbre del modelo, la estructura de los residuos y los límites de lo que las características observables pueden explicar.

**Dónde lo encuentran los tasadores.** Los tasadores encuentran variables latentes en tres contextos recurrentes.

**Los ajustes por calidad y estado de conservación.** La "calidad" y el "estado" como categorías de tasación son en sí mismos intentos de colapsar un constructo latente multidimensional en una sola escala ordinal. Una calificación de estado C3 en el UAD 3.6 pretende resumir el mantenimiento diferido de un inmueble, el desgaste de los acabados, la antigüedad de los sistemas mecánicos y la presentación — de cuatro a diez atributos subyacentes condensados en un solo número. El resumen de un solo número es un proxy de un constructo latente; el oficio del tasador incluye el juicio sobre cuán bien captura el proxy la verdad subyacente en cada caso.

**El residuo sin explicar en el cuadro.** Cuando tres comparables ajustados indican valor pero discrepan más de lo esperado, la discrepancia es informativa. Rara vez es toda ruido muestral; más a menudo refleja un atributo que el tasador aún no ha identificado ni ajustado. Investigar la discrepancia residual — "¿por qué este comparable indica $50,000 por encima de los demás?" — es la versión informal del tasador de la inferencia de variables latentes.

**La mejora reciente que nadie puede ver.** Un inmueble puede haber pasado por una remodelación de cocina, un reemplazo de techo o la construcción de una ADU con permiso que aún

no se refleja en la MLS ni en los datos del catastro. Desde la perspectiva del tasador, la mejora es observable en persona pero no en los datos. Desde la perspectiva de la regresión, la mejora es latente. El tasador debe conciliar la observación en persona con la indicación basada en datos, y ambas deberían converger en la verdadera contribución del atributo latente.

**Por qué importa para la defendibilidad.** Las variables latentes introducen varias consideraciones de defendibilidad que el modelo lineal aditivo de la entrada 005 no aborda directamente.

**Sesgo por variables omitidas.** Si una variable latente está correlacionada con una variable observada, su efecto se filtra en el coeficiente de la variable observada. Si las casas de Pacifica vendidas recientemente tienden a tener cocinas remodeladas (un atributo latente) y la fecha del contrato de la venta está en los datos, entonces el coeficiente de la antigüedad de la venta (alias `sale_age` o "días desde la venta") absorberá parte del efecto de la remodelación de cocina. La contribución estimada del precio implícito de `sale_age` queda sesgada — refleja `sale_age` *más* una porción del estado no medido de la remodelación de cocina. Una tasación defendible reconoce que los coeficientes de las variables observadas están condicionados a lo que falta en el modelo.

**Interpretación de los residuos.** La teoría estándar de la regresión trata los residuos como ruido gaussiano de media cero. La visión de variables latentes trata los residuos como una mezcla de ruido y señal: una parte es variación muestral, otra parte es el efecto acumulado de los atributos no observados. Identificar cuál es cuál es metodológicamente sustantivo. Un residuo que se correlaciona con la ubicación, con la fecha de venta, con la identidad del tasador o con cualquier otro observable sugiere que la variable latente correspondiente puede recuperarse parcialmente.

**Selección del modelo.** Añadir más variables observadas es el remedio obvio para el sesgo por variables omitidas, pero las variables latentes que más importan suelen ser las más difíciles de añadir al modelo (el grano fino del estado, la calidad de las mejoras recientes, los microefectos de vecindario). La respuesta defendible a veces consiste en reconocer que el modelo nunca las capturará y asignar la varianza sin explicar al residuo, en lugar de forzar el ajuste de variables adicionales.

**Ejemplo de tasación resuelto.** La regresión de la entrada 005 produjo un residuo inusual para el comparable H: el modelo subpredijo el precio del comparable H en $46,455 mientras que los otros siete residuos cayeron dentro de una banda de ±$18,000. Las características observadas del comparable H — GLA de 1,600, lote de 7,300, sin vista, estado "bueno" — lo sitúan firmemente en el centro de la muestra; nada de su perfil observable predice un precio inusualmente alto. Y sin embargo se vendió por $46,455 más de lo que el modelo dice que debería.

La hipótesis natural es que el comparable H tiene un atributo latente que contribuye a su precio pero no aparece en los datos. En la construcción del conjunto de datos de este número, ese es exactamente el caso: el comparable H pasó por una remodelación reciente de cocina que añade aproximadamente $60,000 al valor de su conjunto. La remodelación no estaba en la lista de variables de la regresión, de modo que la contribución se acumuló en el residuo y no en un coeficiente. El residuo estimado ($46,455) es ligeramente menor que la verdadera contribución latente ($60,000) porque la regresión reparte pequeñas porciones del efecto latente entre los demás coeficientes, sesgándolos al alza en las dimensiones correlacionadas.

Se siguen tres consecuencias operativas:

**Trate el residuo atípico como evidencia, no como error.** La primera respuesta apropiada al residuo del comparable H no es descartarlo como atípico, sino investigar qué tiene el inmueble que explique la subpredicción. En este caso, una inspección en persona o una búsqueda en los registros de permisos revelaría la remodelación. El fallo del modelo tiene valor diagnóstico.

**Reconozca que los demás coeficientes están ligeramente sesgados.** Los coeficientes de la vista, el estado y el lote han absorbido pequeñas fracciones del efecto de la remodelación a través de correlaciones con las características observadas del comparable H. El sesgo es pequeño con $n = 8$

y se reduciría aún más con más comparables y más covariables, pero no es cero.

**Los residuos en su conjunto llevan estructura recuperable.** Esta es la afirmación sustantiva de la metodología del Residual Constraint Approach (RCA) desarrollada por el autor (Craytor, 2025): en los datos de tasación residencial, los residuos sistemáticamente no son ruido gaussiano iid. Reflejan la contribución estructurada de los atributos no observados — el grano fino del estado, el grano fino de la calidad, la utilidad funcional y otros constructos latentes — que puede recuperarse parcialmente tratando el residuo mismo como una señal de variables latentes que debe modelarse en lugar de descartarse.

El tratamiento que el RCA da al estado y a la calidad hace operativa esta estructura de dos niveles. Las calificaciones ordinales del tasador — los proxies de grano grueso discutidos arriba, que pueden etiquetarse condición-1 (grano grueso del estado) y calidad-1 (grano grueso de la calidad) — son lo bastante burdas como para que bien pudieran tratarse ellas mismas como variables latentes, pero aun así pueden ingresarse como regresores para obtener contribuciones aproximadas de precios implícitos. El protocolo traslada entonces esas contribuciones estimadas fuera de los atributos medidos en la sección de regresión del cuadro de ventas y de vuelta al residuo, desde el cual se reasignan a las variables latentes de grano fino condición-2 (grano fino del estado) y calidad-2 (grano fino de la calidad) y se refinan contra evidencia adicional, como los comentarios públicos de la MLS. La calificación de grano grueso proporciona una asignación de valor inicial; la variable latente de grano fino lleva la estructura recuperable. Al final, la contribución de valor de grano fino reemplaza a la contribución de valor de grano grueso.

La ilustración con un solo comparable presentada aquí es el caso más simple de la intuición del RCA. Con más datos y un modelado más rico, la estructura del residuo puede particionarse en factores latentes reconocibles, cada uno de los cuales contribuye al precio de mercado del conjunto de maneras que la especificación hedónica estándar no puede alcanzar. Las variables latentes no son una molestia que deba controlarse y eliminarse; son la próxima frontera del modelado de valuación defendible.

*Referencias.*

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *American Journal of Psychology*, 15(2), 201–292.

Bollen, K. A. (2002). Latent Variables in Psychology and the Social Sciences. *Annual Review of Psychology*, 53, 605–634.

Craytor, B. (2025). Residual Constraint Approach: Framework & Protocol. Zenodo. doi:10.5281/zenodo.14787917

---

*Referencias cruzadas: espacio de características*; *función de precios hedónicos*; *precio implícito*; *regresión hedónica*.

---

## Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (`006-latent-variable.pl`) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 006-latent-variable.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 006
%%   Title:    Latent Variable
%%   Author:   Bert Craytor
%%   Concept:  latent_variable
```

```
%%   Version:  0.1.0-draft

% ========================================================================
% DICTIONARY IMPORT
% ========================================================================

:- use_module('../../../../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----------
% term(latent_variable, "Latent Variable", measurement_theory,
%      "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.",
%      published, 'dictionary-2026.1').
%
% term(residual, ...).
% term(omitted_variable_bias, ...).
% term(proxy_variable, ...).
% term(residual_constraint_approach, ...).  % synonym: rca
%
% introduced_by(latent_variable, spearman_1904).
% formalized_by(latent_variable, bollen_2002).
% introduced_by(residual_constraint_approach, craytor_2025).
% depends_on(latent_variable, characteristics_space).
% ------------------------------------------------------------------------

:- use_terms([ latent_variable, residual, omitted_variable_bias,
               proxy_variable, residual_constraint_approach,
               characteristics_space, hedonic_price_function,
               implicit_price, hedonic_regression ]).
:- use_relations([related/2, depends_on/2,
                  introduced_by/2, formalized_by/2]).

% ========================================================================
% ARTICLE SUBJECT
% ========================================================================

article_concept(latent_variable).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.006').

% ========================================================================
% LOCAL VOCABULARY
% ========================================================================

local_term(unobserved).
local_term(noise).
local_term(signal).
local_term(structured_signal).
local_term(uad_3_6).
local_term(c3_condition_rating).
local_term(mls).
local_term(permit_records).
local_term(neighborhood_micro_effect).

% ========================================================================
% THE CENTRAL CLAIM
% ========================================================================
```

```
claim(latent_variables_accumulate_in_residuals,
    "Unobserved-but-relevant attributes do not vanish when omitted from the model; their effects
   ↪accumulate in the residual, appearing as noise from the model's perspective but as structured
   ↪ signal from the data-generating process's perspective.").

claim(residuals_are_mixture,
    "Standard regression theory treats residuals as iid Gaussian noise; the latent-variable view
   ↪treats them as a mixture of sampling noise AND signal from unobserved attributes.").

% ======================================================================
% THREE CONTEXTS WHERE APPRAISERS ENCOUNTER LATENT VARIABLES
% ======================================================================

appraiser_encounter(quality_and_condition_proxies,
    "UAD 3.6 'C3 condition' is a single ordinal scale collapsing a multi-dimensional construct (
   ↪deferred maintenance, finish wear, mechanical-systems age, presentation). The rating is a
   ↪proxy_variable for a latent construct.").

appraiser_encounter(unexplained_grid_disagreement,
    "When three adjusted comps disagree by more than expected, the disagreement is rarely all
   ↪sampling noise; usually reflects an attribute not yet identified or adjusted for.").

appraiser_encounter(unobserved_recent_improvement,
    "A recent remodel, roof replacement, or permitted ADU buildout may be observable in person but
   ↪not in MLS/assessor data | observable to the appraiser but latent to the regression.").

% ======================================================================
% DEFENSIBILITY ISSUES
% ======================================================================

defensibility_issue(omitted_variable_bias,
    "If a latent variable is correlated with an observed variable, its effect leaks into the
   ↪observed variable's coefficient. Observed-variable coefficients are conditional on what is
   ↪missing from the model.").

defensibility_issue(residual_interpretation,
    "Treat residuals as a mixture, not pure noise. A residual that correlates with location, sale
   ↪date, appraiser identity, or any other observable suggests the latent variable can be
   ↪partially recovered.").

defensibility_issue(model_selection_limits,
    "The latent variables that matter most are often hardest to add (condition fine-grain, quality,
   ↪ micro-locational effects). Sometimes the defensible response is to allocate the unexplained
   ↪variance to the residual rather than force-fit additional variables.").

% ======================================================================
% WORKED EXAMPLE: COMP H'S ANOMALOUS RESIDUAL
% ======================================================================

:- consult('../../kb/pacifica_comps').

% Comp H's regression residual from entry 005.
comp_h_residual_dollars(46455).

% The latent attribute revealed by investigation.
comp_h_latent_attribute(kitchen_remodel).
comp_h_latent_attribute_value_estimate(60000).
```

```
% Why the estimated residual is smaller than the latent attribute's
% true contribution: regression spreads small portions of the latent
% effect across correlated coefficients, biasing them upward.
residual_smaller_than_truth_explanation(
    "Regression spreads small portions of the latent effect across the other coefficients via
   ↪ correlations, biasing them slightly upward in correlated dimensions.").

% ======================================================================
% THREE OPERATIONAL CONSEQUENCES
% ======================================================================

operational_consequence(outlier_residual_is_evidence,
    "The first response to Comp H's residual is to investigate, not to drop. An in-person
   ↪ inspection or permit search reveals the remodel; the model's failure has diagnostic value.").

operational_consequence(other_coefficients_slightly_biased,
    "View, condition, and lot coefficients absorbed small fractions of the remodel effect via
   ↪ correlations with Comp H's observed characteristics. Bias is small at n=8 but not zero.").

operational_consequence(residuals_carry_recoverable_structure,
    "Residuals in residential appraisal data are systematically not iid Gaussian. They reflect
   ↪ structured contributions of unobserved attributes that can be partially recovered: this is
   ↪ the Residual Constraint Approach intuition.").

% ======================================================================
% CONNECTION TO RESIDUAL CONSTRAINT APPROACH (RCA)
% ======================================================================
%
% The single-comp illustration here is the simplest case of the RCA
% intuition. With richer modeling, residual structure is partitioned
% into recognizable latent factors.

framework_reference(residual_constraint_approach, craytor_2025).

rca_intuition(
    "Treat the residual itself as a latent-variable signal to be modeled rather than discarded.
   ↪ Latent variables are not a nuisance to be controlled away; they are the next frontier of
   ↪ defensible valuation modeling.").

rca_factor_examples(
    [condition_fine_grain, quality_fine_grain, functional_utility]).

% Two-level condition/quality structure (RCA protocol; mirrors the prose).
rating_levels(condition, condition_coarse_grain, condition_fine_grain).
rating_levels(quality,   quality_coarse_grain,   quality_fine_grain).

coarse_to_fine_grain_protocol(
    "Regress on the appraiser's coarse-grain C1-C6/Q1-Q6 ratings to obtain approximate value
   ↪ contributions; move those contributions out of the measured attributes back into the residual;
   ↪  reassign them to the fine-grain latent variables and refine against MLS public remarks. In
   ↪ the end, the fine-grain value contribution replaces the coarse-grain value contribution.").

% ======================================================================
% INFERENCE RULE: anomalous-residual detection
% ======================================================================
%
% A property's residual is "anomalous" if its magnitude exceeds the
% typical residual band reported in entry 005 (+/-$18,000 for the other
% seven comps). Useful for flagging properties whose hedonic prediction
```

```
% suggests a latent attribute warrants investigation.

typical_residual_band(18000).

anomalous_residual(P, R) :-
    comp_h_residual_dollars(R0),  % stand-in: this entry references the
                                  % residual computed in entry 005; future
                                  % integration would pull from there.
    P == comp_h,
    R = R0.

% ======================================================================
% CROSS-REFERENCES
% ======================================================================

cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).
cross_reference(hedonic_regression).
```