Latente Variable

Foundations • *The Valuation Engineer*

Bert Craytor • 2. August 2026
Version 1.0.0
DOI: 10.67330/11n3g168

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Eine latente Variable ist ein unbeobachtetes Merkmal, das beobachtete Ergebnisse beeinflusst und dessen Wert aus seinen Wirkungen auf beobachtbare Variablen erschlossen werden muss, statt direkt gemessen zu werden.

Intuitive Einordnung. Die ersten fünf Einträge behandelten den Merkmalsvektor $\mathbf{z}$, als ließe er sich vollständig messen: GLA, Grundstück, Aussicht, Zustand und so weiter, allesamt als Datenspalten verfügbar. In der Praxis ist kein Merkmalsvektor jemals vollständig. Manche Merkmale fehlen in den Daten, weil sie schwer zu messen sind (Qualität des Innenausbaus, Fließen des Grundrisses, mikrolokale Feinausrichtung), weil sie in den üblichen Datenquellen nicht vorkommen (MLS-Datensätze erfassen weder Lichtqualität noch Schallschutz noch jüngste ungenehmigte Verbesserungen) oder weil sie noch nicht als relevant erkannt wurden (Merkmale auf Nachbarschaftsebene, die sich erst nach der Analyse als preisrelevant erweisen).

Diese unbeobachteten-aber-relevanten Merkmale sind *latente Variablen*. Sie existieren, sie beeinflussen den Preis, aber sie stehen nicht in den Spalten unserer Regression. Ihre Wirkungen verschwinden nicht dadurch, dass sie ungemessen bleiben; sie sammeln sich in den Residuen des Modells, wo sie aus Sicht der aufgenommenen Variablen als „Rauschen“ erscheinen, aus Sicht des zugrunde liegenden datenerzeugenden Prozesses aber tatsächlich strukturiertes Signal sind.

Latente Variablen sind die Brücke zwischen dem idealisierten hedonischen Rahmenwerk und der unordentlichen Realität der Bewertungsdaten. Sie ausdrücklich anzuerkennen ist Voraussetzung für ehrliches Nachdenken über Modellunsicherheit, Residuenstruktur und die Grenzen dessen, was beobachtbare Merkmale erklären können.

Wo Gutachter dem Konzept begegnen. Gutachter begegnen latenten Variablen in drei wiederkehrenden Zusammenhängen.

Die Anpassungen für Qualität und Zustand. „Qualität“ und „Zustand“ sind als Bewertungskategorien selbst schon Versuche, ein mehrdimensionales latentes Konstrukt auf eine einzige Ordinalskala zusammenzuziehen. Eine Zustandsbewertung C3 nach UAD 3.6 soll den aufgestauten Instandhaltungsrückstand einer Immobilie, den Verschleiß der Oberflächen, das Alter der technischen Anlagen und die Präsentation zusammenfassen — vier bis zehn zugrunde liegende Merkmale, verdichtet zu einer einzigen Zahl. Die Ein-Zahl-Zusammenfassung ist ein Proxy für ein latentes Konstrukt; zur Kunst des Gutachters gehört das Urteil darüber, wie gut der Proxy die zugrunde liegende Wahrheit im Einzelfall erfasst.

Das unerklärte Residuum im Raster. Wenn drei angepasste Comps einen Wert anzeigen, dabei aber stärker voneinander abweichen als erwartet, ist die Abweichung informativ. Sie ist selten reines Stichprobenrauschen; häufiger spiegelt sie ein Merkmal wider, das der Gutachter noch nicht identifiziert und angepasst hat. Der Abweichung im Residuum nachzugehen — „Warum zeigt dieser Comp $50,000 über den anderen an?“ — ist die informelle Version der Latente-Variablen-Inferenz beim Gutachter.

Die jüngste Verbesserung, die niemand sehen kann. Eine Immobilie kann eine Küchenrenovierung, eine Dacherneuerung oder den genehmigten Ausbau einer ADU hinter sich haben, die sich noch nicht in MLS- oder Katasterdaten niederschlägt. Aus Sicht des Gutachters ist die Verbesserung vor Ort beobachtbar, in den Daten aber nicht. Aus Sicht der Regression ist die Verbesserung latent. Der Gutachter muss die Beobachtung vor Ort mit der datengestützten Indikation in Einklang bringen, und beide sollten auf den wahren Beitrag des latenten Merkmals zulaufen.

Warum es für die Vertretbarkeit wichtig ist. Latente Variablen führen mehrere Vertretbarkeitsüberlegungen ein, die das additive lineare Modell aus Eintrag 005 nicht unmittelbar behandelt.

Verzerrung durch ausgelassene Variablen. Ist eine latente Variable mit einer beobachteten Variable korreliert, sickert ihr Effekt in den Koeffizienten der beobachteten Variable. Wenn kürzlich verkaufte Pacifica-Häuser tendenziell renovierte Küchen haben (ein latentes Merkmal) und das Vertragsdatum des Hausverkaufs in den Daten steht, dann wird der Koeffizient des Verkaufsalters (auch `sale_age` oder „Tage seit Verkauf“) einen Teil des Küchenrenovierungseffekts aufnehmen. Der geschätzte implizite-Preis-Beitrag von `sale_age` wird verzerrt — er spiegelt `sale_age` *plus* einen Anteil des ungemessenen Küchenrenovierungsstatus wider. Vertretbare Bewertung erkennt an, dass die Koeffizienten beobachteter Variablen bedingt sind auf das, was im Modell fehlt.

Interpretation der Residuen. Die Standardregressionstheorie behandelt Residuen als gaußsches Rauschen mit Erwartungswert null. Die Latente-Variablen-Sicht behandelt Residuen als Mischung aus Rauschen und Signal: ein Teil ist Stichprobenvariation, ein Teil ist die kumulierte Wirkung unbeobachteter Merkmale. Zu bestimmen, was wovon ist, ist methodisch substanziell. Ein Residuum, das mit der Lage, mit dem Verkaufsdatum, mit der Identität des Gutachters oder mit irgendeinem anderen Beobachtbaren korreliert, legt nahe, dass die zugehörige latente Variable teilweise zurückgewonnen werden kann.

Modellwahl. Mehr beobachtete Variablen aufzunehmen ist das naheliegende Heilmittel gegen die Verzerrung durch ausgelassene Variablen, aber die latenten Variablen, auf die es am meisten ankommt, sind oft gerade diejenigen, die sich am schwersten in das Modell aufnehmen lassen (Feinkörnigkeit des Zustands, Qualität jüngster Verbesserungen, Nachbarschafts-Mikroeffekte). Die vertretbare Antwort besteht mitunter darin anzuerkennen, dass das Modell sie nie erfassen wird, und die unerklärte Varianz dem Residuum zuzuweisen, statt zusätzliche Variablen mit Gewalt anzupassen.

Durchgerechnetes Bewertungsbeispiel. Die Regression in Eintrag 005 erzeugte ein ungewöhnliches Residuum für Comp H: Das Modell unterschätzte den Preis von Comp H um $46,455, während die übrigen sieben Residuen in ein Band von ±$18,000 fielen. Die beobachteten Merkmale von Comp H — GLA 1,600, Grundstück 7,300, keine Aussicht, Zustand „gut“ — platzieren ihn klar in der Mitte der Stichprobe; nichts an seinem beobachtbaren Profil sagt einen ungewöhnlich hohen Preis voraus. Und doch wurde er für $46,455 mehr verkauft, als das Modell sagt.

Die naheliegende Hypothese lautet, dass Comp H ein latentes Merkmal besitzt, das zu seinem Preis beiträgt, aber nicht in den Daten auftaucht. Bei der Konstruktion des Datensatzes für diese Ausgabe ist genau das der Fall: Comp H hat eine jüngste Küchenrenovierung erfahren, die dem Bündelwert etwa $60,000 hinzufügt. Die Renovierung stand nicht in der Variablenliste der Regression, sodass sich der Beitrag im Residuum statt in einem Koeffizienten ansammelte. Das geschätzte Residuum ($46,455) ist etwas kleiner als der wahre latente Beitrag ($60,000), weil die Regression kleine Anteile des latenten Effekts auf die übrigen Koeffizienten verteilt und diese in korrelierten Dimensionen nach oben verzerrt.

Daraus folgen drei operative Konsequenzen:

Behandeln Sie das Ausreißer-Residuum als Evidenz, nicht als Fehler. Die erste angemessene Reaktion auf das Residuum von Comp H besteht nicht darin, ihn als Ausreißer zu

verwerfen, sondern zu untersuchen, was an der Immobilie die Unterprognose erklärt. In diesem Fall würde eine Besichtigung vor Ort oder eine Recherche in den Genehmigungsakten die Renovierung zutage fördern. Das Versagen des Modells hat diagnostischen Wert.

Erkennen Sie an, dass die übrigen Koeffizienten leicht verzerrt sind. Die Koeffizienten für Aussicht, Zustand und Grundstück haben über Korrelationen mit den beobachteten Merkmalen von Comp H kleine Bruchteile des Renovierungseffekts aufgenommen. Die Verzerrung ist bei $n = 8$ klein und würde mit mehr Comps und mehr Kovariaten weiter schrumpfen, aber sie ist nicht null.

Die Residuen als Ganzes tragen zurückgewinnbare Struktur. Dies ist die substanzielle These der vom Autor entwickelten Methodik des Residual Constraint Approach (RCA) (Craytor, 2025): In Daten der Wohnimmobilienbewertung sind Residuen systematisch kein iid-gaußsches Rauschen. Sie spiegeln den strukturierten Beitrag unbeobachteter Merkmale wider — Feinkörnigkeit des Zustands, Feinkörnigkeit der Qualität, funktionale Nützlichkeit und andere latente Konstrukte —, der sich teilweise zurückgewinnen lässt, indem man das Residuum selbst als Latente-Variablen-Signal behandelt, das modelliert statt verworfen werden sollte.

Die Behandlung von Zustand und Qualität im RCA macht diese zweistufige Struktur operational. Die ordinalen Bewertungen des Gutachters — die oben besprochenen grobkörnigen Proxys, die man mit Zustand-1 (Zustand grobkörnig) und Qualität-1 (Qualität grobkörnig) bezeichnen kann — sind grob genug, dass man sie ebenso gut selbst als latente Variablen behandeln könnte; sie lassen sich aber dennoch als Regressoren aufnehmen, um näherungsweise implizite-Preis-Beiträge zu erhalten. Das Protokoll verschiebt diese geschätzten Beiträge dann aus den gemessenen Merkmalen im Regressionsteil des Verkaufsrasters zurück in das Residuum, von wo aus sie den feinkörnigen latenten Variablen Zustand-2 (Zustand feinkörnig) und Qualität-2 (Qualität feinkörnig) neu zugewiesen und anhand zusätzlicher Evidenz wie den öffentlichen MLS-Bemerkungen verfeinert werden. Die grobkörnige Bewertung liefert eine erste Wertzuweisung; die feinkörnige latente Variable trägt die zurückgewinnbare Struktur. Am Ende ersetzt der feinkörnige Wertbeitrag den grobkörnigen Wertbeitrag.

Die Ein-Comp-Illustration hier ist der einfachste Fall der RCA-Intuition. Mit mehr Daten und reicherer Modellierung lässt sich die Struktur im Residuum in erkennbare latente Faktoren zerlegen, von denen jeder auf Weisen zum Marktpreis des Bündels beiträgt, die die hedonische Standardspezifikation nicht erreichen kann. Latente Variablen sind kein Störfaktor, den man wegkontrollieren muss; sie sind die nächste Grenze der vertretbaren Bewertungsmodellierung.

Literatur.

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

Querverweise: Merkmalsraum; *hedonische Preisfunktion*; *impliziter Preis*; *hedonische Regression.*

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (`006-latent-variable.pl`) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 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).
```