

Variable latente (version française)

Foundations • The Valuation Engineer

Bert Craytor • 2 août 2026

Version 1.0.0

DOI: [10.67330/11n3g168](https://doi.org/10.67330/11n3g168)

Traduction française de l'entrée « Latent Variable » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.

Définition formelle. Une variable latente est un attribut non observé qui influence des résultats observés, et dont la valeur doit être inférée de ses effets sur des variables observables plutôt que mesurée directement.

Cadrage intuitif. Les cinq premières entrées ont traité le vecteur de caractéristiques $\mathbf{z}$ comme s'il pouvait être entièrement mesuré : GLA, terrain, vue, état, et ainsi de suite, tous disponibles comme colonnes de données. En pratique, aucun vecteur de caractéristiques n'est jamais complet. Certains attributs ne figurent pas dans les données parce qu'ils sont difficiles à mesurer (qualité des finitions intérieures, fluidité du plan, micro-orientation micro-localisationnelle), parce qu'ils sont absents des flux de données standard (les fiches MLS ne capturent pas la qualité de l'éclairage, l'insonorisation ou les améliorations récentes réalisées sans permis), ou parce qu'ils n'ont pas encore été reconnus comme pertinents (des attributs à l'échelle du quartier qui ne se révèlent prédictifs du prix qu'après que l'analyse les a mis au jour).

Ces attributs non observés mais pertinents sont des *variables latentes*. Ils existent, ils influencent le prix, mais ils ne sont pas dans les colonnes de notre régression. Leurs effets ne disparaissent pas parce qu'ils ne sont pas mesurés ; ils s'accumulent dans les résidus du modèle, où ils apparaissent comme du « bruit » du point de vue des variables incluses mais sont en réalité un signal structuré du point de vue du processus générateur de données sous-jacent.

Les variables latentes sont le pont entre le cadre hédonique idéalisé et la réalité désordonnée des données d'évaluation. Les reconnaître explicitement est une condition préalable à un raisonnement honnête sur l'incertitude du modèle, la structure des résidus et les limites de ce que les caractéristiques observables peuvent expliquer.

Où les évaluateurs rencontrent ce concept. Les évaluateurs rencontrent les variables latentes dans trois contextes récurrents.

Les ajustements de qualité et d'état. La « qualité » et l'« état » comme catégories d'évaluation sont elles-mêmes des tentatives de condenser un construit latent multidimensionnel en une seule échelle ordinale. Une note d'état C3 dans l'UAD 3.6 est censée résumer l'entretien différé d'un bien, l'usure des finitions, l'âge des systèmes mécaniques et la présentation — quatre à dix attributs sous-jacents condensés en un seul nombre. Le résumé en un seul nombre est un proxy d'un construit latent ; le métier de l'évaluateur inclut le jugement sur la fidélité avec laquelle le proxy capture la vérité sous-jacente dans chaque cas.

Le résidu inexpliqué dans la grille. Lorsque trois comparables ajustés indiquent une valeur mais divergent plus qu'attendu, le désaccord est informatif. Il est rarement tout entier du bruit d'échantillonnage ; le plus souvent il reflète un attribut que l'évaluateur n'a pas encore identifié ni ajusté. Enquêter sur le désaccord résiduel — « pourquoi ce comparable indique-t-il \$50,000 au-dessus des autres ? » — est la version informelle, chez l'évaluateur, de l'inférence sur variables latentes.

L'amélioration récente que personne ne peut voir. Un bien peut avoir fait l'objet d'une rénovation de cuisine, d'un remplacement de toiture ou de la construction d'un ADU avec permis qui n'est pas encore reflété dans les données MLS ou de l'assesseur. Du point de vue de l'évaluateur, l'amélioration est observable sur place mais pas dans les données. Du point de vue de la régression, l'amélioration est latente. L'évaluateur doit réconcilier l'observation sur place avec l'indication issue des données, et les deux devraient converger vers la contribution réelle de l'attribut latent.

Pourquoi cela importe pour la défendabilité. Les variables latentes introduisent plusieurs considérations de défendabilité que le modèle linéaire additif de l'entrée 005 ne traite pas directement.

Biais de variable omise. Si une variable latente est corrélée à une variable observée, son effet fuit dans le coefficient de la variable observée. Si les maisons de Pacifica récemment vendues ont tendance à avoir des cuisines rénovées (un attribut latent) et que la date du contrat de vente figure dans les données, alors le coefficient de l'ancienneté de la vente (alias `sale_age` ou « jours depuis la vente ») absorbera une partie de l'effet de rénovation de cuisine. La contribution estimée du prix implicite de `sale_age` devient biaisée — elle reflète `sale_age` plus une portion du statut non mesuré de rénovation de cuisine. Une évaluation défendable reconnaît que les coefficients des variables observées sont conditionnels à ce qui manque au modèle.

Interprétation des résidus. La théorie standard de la régression traite les résidus comme un bruit gaussien de moyenne nulle. La vision par variables latentes traite les résidus comme un mélange de bruit et de signal : une partie est de la variation d'échantillonnage, une partie est l'effet cumulé des attributs non observés. Identifier laquelle est laquelle est méthodologiquement substantiel. Un résidu qui corrèle avec la localisation, avec la date de vente, avec l'identité de l'évaluateur ou avec toute autre variable observable suggère que la variable latente correspondante peut être partiellement récupérée.

Sélection de modèle. Ajouter davantage de variables observées est le remède évident au biais de variable omise, mais les variables latentes qui comptent le plus sont souvent celles qui sont les plus difficiles à ajouter au modèle (le grain fin de l'état, la qualité des améliorations récentes, les micro-effets de quartier). La réponse défendable consiste parfois à reconnaître que le modèle ne les capturera jamais et à allouer la variance inexpliquée au résidu plutôt que de forcer l'ajustement de variables supplémentaires.

Exemple d'évaluation travaillé. La régression de l'entrée 005 a produit un résidu inhabituel pour le comp H : le modèle a sous-prédit le prix du comp H de \$46,455 alors que les sept autres résidus se situaient dans une bande de $\pm \$18,000$. Les caractéristiques observées du comp H — GLA 1,600, terrain 7,300, sans vue, état « bon » — le placent fermement au milieu de l'échantillon ; rien dans son profil observable ne prédit un prix inhabituellement élevé. Et pourtant il s'est vendu \$46,455 de plus que ce que le modèle dit qu'il aurait dû.

L'hypothèse naturelle est que le comp H possède un attribut latent qui contribue à son prix mais n'apparaît pas dans les données. Dans la construction du jeu de données de ce numéro, c'est exactement le cas : le comp H a fait l'objet d'une rénovation récente de cuisine qui ajoute environ \$60,000 à la valeur de son panier. La rénovation ne figurait pas dans la liste de variables de la régression, de sorte que la contribution s'est accumulée dans le résidu plutôt que dans un coefficient. Le résidu estimé (\$46,455) est légèrement inférieur à la vraie contribution latente (\$60,000) parce que la régression répartit de petites portions de l'effet latent sur les autres coefficients, les biaisant à la hausse dans les dimensions corrélées.

Trois conséquences opérationnelles s'ensuivent :

Traiter le résidu aberrant comme une preuve, non comme une erreur. La première réponse appropriée au résidu du comp H n'est pas de l'écarter comme aberrant mais d'enquêter

sur ce qui, dans le bien, explique la sous-prédiction. Dans ce cas, une inspection sur place ou une recherche dans les registres de permis révélerait la rénovation. L'échec du modèle a une valeur diagnostique.

Reconnaître que les autres coefficients sont légèrement biaisés. Les coefficients de vue, d'état et de terrain ont absorbé de petites fractions de l'effet de rénovation via les corrélations avec les caractéristiques observées du comp H. Le biais est faible à $n = 8$ et rétrécirait encore avec plus de comparables et plus de covariables, mais il n'est pas nul.

Les résidus dans leur ensemble portent une structure récupérable. C'est l'affirmation substantielle de la méthodologie Residual Constraint Approach (RCA) développée par l'auteur (Craytor, 2025) : dans les données d'évaluation résidentielle, les résidus ne sont systématiquement pas un bruit gaussien iid. Ils reflètent la contribution structurée d'attributs non observés — le grain fin de l'état, le grain fin de la qualité, l'utilité fonctionnelle et d'autres construits latents — qui peuvent être partiellement récupérés en traitant le résidu lui-même comme un signal de variable latente à modéliser plutôt qu'à écarter.

Le traitement RCA de l'état et de la qualité rend cette structure à deux niveaux opérationnelle. Les notes ordinales de l'évaluateur — les proxys à gros grain discutés ci-dessus, que l'on peut étiqueter état-1 (état-gros-grain) et qualité-1 (qualité-gros-grain) — sont assez grossières pour qu'on puisse aussi bien les traiter elles-mêmes comme des variables latentes, mais elles peuvent tout de même être saisies comme régresseurs pour obtenir des contributions approchées de prix implicites. Le protocole déplace ensuite ces contributions estimées hors des attributs mesurés de la section régression de la grille des ventes et les renvoie dans le résidu, d'où elles sont réattribuées aux variables latentes à grain fin état-2 (état-grain-fin) et qualité-2 (qualité-grain-fin) et affinées au regard de preuves supplémentaires telles que les remarques publiques de la MLS. La note à gros grain fournit une attribution de valeur initiale ; la variable latente à grain fin porte la structure récupérable. À la fin, la contribution de valeur à grain fin remplace la contribution de valeur à gros grain.

L'illustration à un seul comparable présentée ici est le cas le plus simple de l'intuition RCA. Avec plus de données et une modélisation plus riche, la structure du résidu peut être partitionnée en facteurs latents reconnaissables, chacun contribuant au prix de marché du panier de manières que la spécification hédonique standard ne peut atteindre. Les variables latentes ne sont pas une nuisance à neutraliser ; elles sont la prochaine frontière de la modélisation défendable en évaluation.

Références.

- 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](https://doi.org/10.5281/zenodo.14787917)

Renvois : espace des caractéristiques ; fonction de prix hédoniques ; prix implicite ; régression hédonique.

Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (006-latent-variable.pl) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```

%% 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).
```