潜在変数

(日本語版)

基礎概念（*Foundations*） • *The Valuation Engineer*
Bert Craytor • 2026年 8月 3日
版 1.0.0
DOI: 10.67330/11n3g168
日本語訳。人工知能の支援により翻訳し、編集部が校閲した。正本は英語版である。

形式的定義。 潜在変数（latent variable）とは、観測される結果に影響を与える未観測の属性であり、その値を直接測定することはできず、観測可能な変数への効果から推論しなければならない変数である。

直観的な理解。 先行する五つの項目は、特性ベクトル $\mathbf{z}$ をあたかも完全に測定できるかのように扱ってきた。GLA、敷地、眺望、状態などがすべてデータの列として手に入る、という前提である。しかし実際には、完全な特性ベクトルなど存在しない。ある属性は測定が困難であるためにデータに含まれず（内装仕上げの質、動線の流れ、微視的な立地と方位）、ある属性は標準的なデータ供給に存在しないためにデータに含まれず（MLS の記録は採光の質、遮音性能、最近の無許可の改良を捉えない）、また、ある属性はまだ関連性が認識されていないためにデータに含まれない（分析を経て初めて価格を予測すると判明する近隣水準の特徴）。

このように観測されていないが関連のある属性が潜在変数である。それらは存在し、価格に影響を与えるが、我々の回帰の列には現れない。測定されていないからといって、その効果が消えるわけではない。効果はモデルの残差に蓄積し、含まれている変数の視点からは「雑音」として現れるが、背後にあるデータ生成過程の視点からは、実は構造をもった信号である。

潜在変数は、理想化されたヘドニックの枠組みと、雑然とした鑑定評価データの現実とを架橋するものである。それを明示的に認めることは、モデルの不確実性、残差の構造、そして観測可能な特性が説明しうる範囲の限界について誠実に推論するための前提条件である。

鑑定評価者が出会う場面。 鑑定評価者は、繰り返し現れる三つの文脈で潜在変数に出会う。

品質と状態の調整。鑑定評価の範疇としての「品質」と「状態」は、それ自体が多次元の潜在的構成概念を単一の順序尺度へ圧縮しようとする試みである。UAD 3.6 における C3 の状態等級は、不動産の繰り延べられた維持修繕、仕上げの摩耗、機械設備の経年、そして外観の印象 — 四つから十に及ぶ背後の属性 — を一つの数値に要約することを意図している。単一の数値による要約は潜在的構成概念の代理変数であり、その代理変数が個々の事案において背後の真実をどれほどよく捉えているかについての判断は、鑑定評価者の技量に属する。

比較表において説明されない残差。調整済みの三つの比較事例が価値を指し示しながら、予想以上に食い違うとき、その不一致は情報を含んでいる。それがすべて標本の雑音であることは稀であり、多くの場合、鑑定評価者がまだ特定も調整もしていない属性を反映している。その残差的な食い違いを調べること — 「なぜこの比較事例は他よりも $50,000 高い値を指し示すのか」 — は、鑑定評価者が非公式に行う潜在変数の推論である。

誰にも見えない最近の改良。ある不動産は、キッチンの改装、屋根の葺き替え、あるいは許可を得た ADU の増築を経ていながら、それがまだ MLS や課税評価のデータに反映されていないことがある。鑑定評価者の視点からは、その改良は現地では観測できるがデータには存在しない。回帰の視点からは、その改良は潜在的である。鑑定評価者は現地での観察とデータに基づく指示値とを調整しなければならず、両者はその潜在属性の真の寄与へ収斂すべきである。

防御可能性にとってなぜ重要か。 潜在変数は、項目 005 の加法的線形モデルが直接には扱っていないいくつかの防御可能性上の論点をもたらす。

欠落変数バイアス。潜在変数が観測変数と相関している場合、その効果は観測変数の係数へ漏れ出す。最近売却されたパシフィカの住宅がキッチンを改装している傾向があり（潜在属性）、住宅売買の契約日がデータに含まれているならば、売却からの経過期間（すなわち `sale_age`、「売却後経過日数」）の係数は、キッチン改装の効果の一部を吸収することになる。`sale_age` の推定された潜在価格としての寄与は偏りを帯びる — それは `sale_age` に加えて、測定されていないキッチン改装の状態の一部を反映するからである。防御可能な鑑定評価は、観測変数の係数がモデルから欠けているものに条件づけられていることを認める。

残差の解釈。標準的な回帰理論は、残差を平均ゼロのガウス雑音として扱う。潜在変数の視点は、残差を雑音と信号の混合として扱う。一部は標本変動であり、一部は未観測の属性の累積的な効果である。どちらがどちらであるかを見分けることは、方法論上の実質的な課題である。立地、売却日、担当鑑定評価者、その他いかなる観測可能な変数とも相関する残差は、対応する潜在変数を部分的に復元しうることを示唆する。

モデル選択。観測変数をさらに加えることは欠落変数バイアスに対する自明な処方であるが、最も重要な潜在変数は、しばしばモデルへ追加することが最も難しいもの（状態の細粒度、最近の改良の品質、近隣の微視的効果）である。防御可能な対応は、モデルが決してそれらを捉えないと認め、変数を無理に押し込むのではなく、説明されない分散を残差へ配分することである場合もある。

鑑定評価実務の例題。 項目 005 の回帰は、比較事例 H について異例の残差を生んだ。他の七つの残差が ±$18,000 の帯に収まったのに対し、モデルは比較事例 H の価格を $46,455 だけ過小に予測した。比較事例 H の観測された特性 — GLA 1,600、敷地 7,300、眺望なし、状態は「良好」— は、それを標本のまさに中央に位置づける。観測可能なプロファイルのうち、異常に高い価格を予告するものは何もない。それにもかかわらず、その不動産はモデルが示す額より $46,455 高く売れた。

自然な仮説は、比較事例 H が価格に寄与しながらデータには現れない潜在属性をもつ、というものである。本号のためのデータセットの構築において、まさにそれが事実である。比較事例 H は最近キッチンの改装を経ており、それが束の価値におよそ $60,000 を加えている。この改装は回帰の変数一覧になかったため、その寄与は係数ではなく残差に蓄積した。推定された残差（$46,455）が真の潜在的寄与（$60,000）よりやや小さいのは、回帰が潜在効果のわずかな部分を他の係数へ分散させ、相関する次元でそれらを上方に偏らせるからである。

ここから三つの実務的な帰結が導かれる。

外れ値としての残差を誤差ではなく証拠として扱うこと。比較事例 H の残差に対する最初の適切な対応は、それを外れ値として除外することではなく、その不動産の何が過小予測を説明するのかを調べることである。この場合、現地調査または建築許可記録の検索が改装を明らかにするであろう。モデルの失敗には診断上の価値がある。

他の係数もわずかに偏っていることを認識すること。眺望、状態、敷地の係数は、比較事例 H の観測された特性との相関を通じて改装効果のごく一部を吸収している。$n = 8$ においてそのバイアスは小さく、比較事例と共変量が増えればさらに縮小するが、ゼロではない。

残差は全体として復元可能な構造を担っている。これが、著者が展開した残差制約アプローチ（Residual Constraint Approach, RCA）方法論の実質的な主張である（Craytor, 2025）。住宅の鑑定評価データにおいて、残差は体系的に iid のガウス雑音ではない。残差は未観測の属性 — 状態の細粒度、品質の細粒度、機能的効用、その他の潜在的構成概念 — による構造化された寄与を反映しており、残差そのものを捨て去るのではなく、モデル化すべき潜在変数の信号として扱うことによって部分的に復元しうる。

RCA が状態と品質を扱う方法は、この二層構造を実行可能なものにする。鑑定評価者の順序等級 — 上で論じた粗粒度の代理変数であり、状態-1（状態・粗粒度）および品質-1（品質・粗粒度）と表しうるもの — は、それ自体を潜在変数として扱っても差し支えないほど粗いが、それでも回帰変数として投入し、おおよその潜在価格としての寄与を得ることができる。次いでプロ

トコルは、そうして推定された寄与分を、取引事例比較表の回帰分析部分における測定された属性の外へ移して残差へ戻し、そこからそれを細粒度の潜在変数である状態-2（状態・細粒度）および品質-2（品質・細粒度）へ再配分し、MLS の公開備考欄のような追加的証拠に照らして精緻化する。粗粒度の等級は初期の価値配分を与え、細粒度の潜在変数が復元可能な構造を担う。最終的には、細粒度の価値寄与が粗粒度の価値寄与を置き換える。

ここに示した単一の比較事例による例示は、RCA の直観の最も単純な事例である。データがより多く、モデル化がより豊かになれば、残差の中の構造は識別可能な潜在因子へと分割することができ、各因子は標準的なヘドニックの定式化が到達しえない仕方で束の市場価格に寄与する。潜在変数は制御して取り除くべき厄介物ではない。それは防御可能な評価モデリングの次の最前線である。

参考文献.

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

相互参照： 特性空間（*characteristics space*）; ヘドニック価格関数（*hedonic price function*）; 潜在価格（*implicit price*）; ヘドニック回帰（*hedonic regression*）.

付録：Prolog による形式化

以下の機械可読な付属ファイル（`006-latent-variable.pl`）は、本誌の Prolog 知識ベースにおける本項目の形式化である。このファイルは本文とともに閲覧可能な Prolog ゲラとしても公開される。

```
%% 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).
```