

潛變數

繁體中文版

基礎概念 (Foundations) • The Valuation Engineer

Bert Craytor • 2026 年 8 月 4 日

版本 1.0.0

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

繁體中文譯本，由人工智慧輔助翻譯並經編輯部審校；英文版為權威版本。

形式定義。潛變數 (latent variable) 是一項未被觀測的屬性，它影響觀測到的結果，其取值必須由它對可觀測變數的作用推斷得出，而無法直接測量。

直觀闡釋。前五個條目在處理特徵向量 $\mathbf{z}$ 時，彷彿它可以被完整測量：建築面積、地塊、景觀、狀況等等，全都以資料欄位的形式可得。而在實務中，沒有任何特徵向量是完整的。有些屬性不在資料之中，是因為它們難以測量（室內裝修品質、平面配置的流暢度、微區位的朝向細節），是因為標準資料來源中根本沒有它們（MLS 紀錄不會記載採光品質、隔音效果，也不會記載近期未取得許可的改建），或者是因為它們尚未被認定為相關（某些社區層面的特徵，只有在分析揭示之後才被發現能夠預測價格）。

這些未被觀測卻相關的屬性，就是潛變數。它們存在，它們影響價格，但它們不在我們迴歸的資料欄位中。它們的作用不會因為未被測量而消失；這些作用會累積到模型的殘差裡，從已納入變數的視角看它們像是「雜訊」，但從底層資料生成過程的視角看，它們其實是有結構的訊號。

潛變數是理想化的特徵價格框架與雜亂的估價資料現實之間的橋梁。明確承認它們的存在，是誠實地思考模型不確定性、殘差結構以及可觀測特徵解釋力邊界的前提。

估價師在何處遇到它。估價師在三類反覆出現的情境中遇到潛變數。

品質與狀況調整。作為估價類別的「品質」與「狀況」，本身就是把一個多維潛在構念壓縮到單一序數量表上的嘗試。UAD 3.6 中的 C3 狀況評級，意在概括一宗房地產的遞延維修、飾面磨損、機電系統年限與呈現狀態——四到十項底層屬性被濃縮為一個數字。這個單一數字的概括是潛在構念的代理變數；估價師的功力之一，就在於判斷該代理在每個具體案例中對底層真相的捕捉程度。

網格中未獲解釋的殘差。當三個經調整的可比案例給出價值指示，彼此的差距卻超出預期時，這一分歧本身是有資訊的。它很少全部是抽樣雜訊；更常見的情況是，它反映了估價師尚未辨識、尚未調整的某項屬性。追查這種殘餘分歧——「為什麼這個可比案例的指示值比其他高出 \$50,000？」——正是估價師對潛變數推斷的非正式版本。

誰也看不見的近期改建。一宗房地產可能剛做過廚房翻新、屋頂更換，或建成了經許可的附屬住宅單元 (ADU)，而這些尚未反映在 MLS 或稅務估價資料中。從估價師的視角看，這項改建實地可見，卻不在資料之中。從迴歸的視角看，這項改建是潛在的。估價師必須把實地觀察與資料驅動的指示相互調和，二者都應收斂於該潛在屬性的真實貢獻。

為何關乎可辯護性。潛變數帶來了若干可辯護性方面的考量，而條目 005 中的可加線性模型並未直接處理這些問題。

遺漏變數偏誤。如果某個潛變數與一個已觀測變數相關，它的效應就會滲入該已觀測變數的係數。如果近期成交的帕西菲卡 (Pacifica) 住宅往往擁有翻新過的廚房（一項潛在屬性），而房屋成交的合約日期又在資料之中，那麼成交時點（即 `sale_age`，「距成交天數」）的係數就會吸收一部分廚房翻新效應。`sale_age` 的隱含價格貢獻估計值因此被扭曲——它反映的是 `sale_age` 加上未被測量的廚房翻新狀態的一部分。可辯護的估價必須承認，已觀測變數的係數是以模型中缺失的內容為條件的。

殘差的解讀。標準迴歸理論把殘差視為均值為零的高斯雜訊。潛變數視角則把殘差視為雜訊與訊號的混合：一部分是抽樣變異，另一部分是未觀測屬性的累積效應。分辨二者在方法論上具有實質意義。若某殘差與位置、與成交日期、與估價師身分，或與任何其他可觀測量相關，就提示相應的潛變數可以被部分還原。

模型選擇。納入更多已觀測變數是遺漏變數偏誤的顯見對策，但最要緊的那些潛變數，往往恰恰是最難加入模型的（狀況的細顆粒、近期改建的品質、社區微效應）。可辯護的因應方式有時是承認模型永遠無法捕捉它們，從而把未獲解釋的變異數歸入殘差，而不是強行塞進額外的變數。

估價實例演算。條目 005 的迴歸對可比案例 H 產生了一個異常殘差：模型對可比案例 H 價格的低估達 \$46,455，而其餘七個殘差都落在 $\pm \$18,000$ 的區間內。可比案例 H 的已觀測特徵——建築面積 1,600、地塊 7,300、無景觀、狀況「良好」——把它牢牢置於樣本的中間位置；其可觀測輪廓中沒有任何東西預示異常高的價格。然而它的成交價卻比模型所說的高出 \$46,455。

自然的假設是：可比案例 H 具有某項潛在屬性，它對價格有貢獻，卻沒有出現在資料中。在本期資料集的建構中，情況恰恰如此：可比案例 H 經歷了一次近期廚房翻新，為其組合價值增加約 \$60,000。這項翻新不在迴歸的變數清單之內，因此其貢獻累積在殘差中，而不是體現在某個係數上。估計出的殘差（\$46,455）略小於真實的潛在貢獻（\$60,000），因為迴歸會把潛在效應的一小部分分攤到其他係數上，在相關維度上使它們向上偏高。

由此可得三點操作性結論：

把離群殘差當作證據，而不是誤差。對可比案例 H 的殘差，第一個恰當的反應不是把它作為離群值剔除，而是去調查這宗房地產的何種特質解釋了這一低估。在本例中，實地查勘或建築許可紀錄檢索就會揭示那次翻新。模型的失敗具有診斷價值。

認識到其餘係數存在輕微偏誤。景觀、狀況與地塊的係數，透過與可比案例 H 已觀測特徵之間的相關性，吸收了翻新效應的一小部分。在 $n = 8$ 時這一偏誤很小，隨著可比案例與共變數的增加還會進一步縮小，但它並不為零。

殘差整體承載著可還原的結構。這正是作者所提出的殘差約束法（Residual Constraint Approach, RCA）的實質主張（Craytor, 2025）：在住宅估價資料中，殘差在系統上並非獨立同分布的高斯雜訊。它們反映了未觀測屬性的結構化貢獻——狀況的細顆粒、品質的細顆粒、功能效用以及其他潛在構念——而這些貢獻可以透過把殘差本身當作有待建模、而非棄置的潛變數訊號來部分還原。

RCA 對狀況與品質的處理，使這一兩層結構得以落實。估價師的序數評級——即上文討論的粗顆粒代理變數，可標記為狀況-1（狀況粗顆粒）與品質-1（品質粗顆粒）——粗糙到幾乎可以把它們本身也視為潛變數，但它們仍可作為迴歸元納入，以得到近似的隱含價格貢獻。隨後，該協定把這些估計出的貢獻從銷售網格迴歸部分的已測屬性中移出、送回殘差，再由殘差重新分配給細顆粒潛變數狀況-2（狀況細顆粒）與品質-2（品質細顆粒），並對照 MLS 公開備註等補充證據加以精細化。粗顆粒評級提供了初始的價值分配；細顆粒潛變數則承載可還原的結構。最終，細顆粒的價值貢獻取代粗顆粒的價值貢獻。

這裡的單一可比案例演示，是 RCA 直覺最簡單的情形。有了更多資料與更豐富的建模，殘差中的結構可以被劃分為若干可辨識的潛在因子，每個因子都以標準特徵價格設定無法企及的方式對組合的市場價格作出貢獻。潛變數不是需要被控制掉的麻煩；它們是可辯護估價建模的下一個前沿。

參考文獻。

- 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

交叉引用：特徵空間 (*characteristics space*)；特徵價格函數 (*hedonic price function*)；隱含價格 (*implicit price*)；特徵價格迴歸 (*hedonic regression*)。

附錄：Prolog 形式化

以下機器可讀的配套檔案 (006-latent-variable.pl) 在本刊的 Prolog 知識庫中對本條目進行了形式化。該檔案同時作為可瀏覽的 Prolog 版本隨本文一併發布。

```
%% 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).
```