

잠재변수

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor • 2026년 8월 3일

판본 1.0.0

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

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 검토했습니다. 영문판이 정본입니다.

형식적 정의. 잠재변수 (latent variable) 란 관측된 결과에 영향을 미치는 미관측 속성으로, 그 값을 직접 측정할 수 없고 관측 가능한 변수에 미치는 효과로부터 추론해야 하는 변수를 말한다.

직관적 이해. 앞선 다섯 항목은 특성 벡터 $\mathbf{z}$ 를 마치 완전히 측정할 수 있는 것처럼 다루었다. GLA, 대지, 조망, 상태 등이 모두 자료의 열로 확보되어 있다는 뜻이 말이다. 실제로는 어떤 특성 벡터도 결코 완전하지 않다. 어떤 속성은 측정이 어려워서 (내부 마감 품질, 동선의 흐름, 미시적 위치와 향), 어떤 속성은 표준 자료원에 없어서 (MLS 기록은 채광 품질, 방음, 최근의 무허가 개량을 담지 않는다), 또 어떤 속성은 아직 관련성이 인정되지 않아서 (분석을 거친 뒤에야 가격을 예측한다고 밝혀지는 근린 수준의 특징) 자료에 들어 있지 않다.

이렇게 관측되지 않았으나 관련이 있는 속성이 잠재변수이다. 그것들은 존재하고 가격에 영향을 미치지만 우리 회귀분석의 열에는 없다. 측정되지 않았다고 해서 효과가 사라지는 않는다. 그 효과는 모형의 잔차에 쌓이며, 포함된 변수의 관점에서는 “잡음” 으로 보이지만 바탕의 자료생성과정의 관점에서는 실은 구조를 가진 신호이다.

잠재변수는 이상화된 헤도닉 체계와 지저분한 감정평가 자료의 현실 사이를 잇는 다리이다. 그것을 명시적으로 인정하는 일은 모형 불확실성, 잔차 구조, 그리고 관측 가능한 특성이 설명할 수 있는 것의 한계에 대해 정직하게 추론하기 위한 전제 조건이다.

감정평가사가 마주치는 지점. 감정평가사는 세 가지 되풀이되는 맥락에서 잠재변수와 마주친다.

품질 및 상태 보정. 감정평가 범주로서의 “품질” 과 “상태” 는 그 자체가 다차원의 잠재 구성개념을 하나의 순서척도로 압축하려는 시도이다. UAD 3.6 의 C3 상태 등급은 부동산의 이연 유지보수, 마감 마모, 기계설비 연식, 외관 상태 — 넷에서 열 개에 이르는 바탕 속성 — 을 하나의 숫자로 요약하려는 것이다. 단일 숫자 요약은 잠재 구성개념의 대리변수이며, 그 대리변수가 각 사안에서 바탕의 진실을 얼마나 잘 포착하는지에 대한 판단은 감정평가사의 기예에 속한다.

격자에서 설명되지 않는 잔차. 보정된 세 개의 비교사례가 가치를 가리키면서도 예상보다 크게 엇갈릴 때, 그 불일치는 정보를 담고 있다. 그것이 모두 표본 잡음인 경우는 드물다. 그보다는 감정평가사가 아직 식별하거나 보정하지 못한 속성을 반영하는 경우가 더 많다. 그 불일치를 파고드는 일 — “왜 이 비교사례는 다른 것들보다 \$50,000 높은 값을 가리키는가?” — 은 감정평가사가 비공식적으로 수행하는 잠재변수 추론이다.

아무도 볼 수 없는 최근의 개량. 어떤 부동산은 주방 리모델링, 지붕 교체, 또는 허가받은 ADU 증축을 마쳤으나 아직 MLS 나 과세평가 자료에 반영되지 않았을 수 있다. 감정평가사의 관점에서 그 개량은 현장에서는 관측되지만 자료에는 없다. 회귀분석의 관점에서 그 개량은 잠재적이다. 감정평가사는 현장 관찰과 자료 기반의 도출값을 조화시켜야 하며, 둘 다 그 잠재 속성의 참된 기여분으로 수렴해야 한다.

왜 방어가능성에 중요한가. 잠재변수는 항목 005 의 가법적 선형모형이 직접 다루지 않는 여러 방어가능성 고려사항을 불러온다.

누락변수 편익. 잠재변수가 관측변수와 상관되어 있으면 그 효과가 관측변수의 계수로 스며든다. 최근 거래된 퍼시픽아 주택들이 주방을 리모델링한 경향이 있고 (잠재 속성) 매매계약일이 자료에 들어 있다면, 거래의 경과 기간 (즉 `sale_age` 또는 “매도 이후 경과일수”) 계수가 주방 리모델링 효과의

일부를 흡수하게 된다. `sale_age` 의 추정된 잠재가격 기여분은 편의를 갖게 된다. 그것은 `sale_age` 에 더하여 측정되지 않은 주방 리모델링 상태의 일부를 반영하기 때문이다. 방어 가능한 감정평가는 관측변수의 계수가 모형에서 빠진 것에 조건부라는 사실을 인정한다.

잔차의 해석. 표준적 회귀이론은 잔차를 평균이 0 인 가우스 잡음으로 취급한다. 잠재변수의 관점은 잔차를 잡음과 신호의 혼합으로 취급한다. 일부는 표본변동이고, 일부는 미관측 속성들의 누적 효과이다. 어느 쪽이 어느 쪽인지를 가려내는 일은 방법론적으로 실질적이다. 위치, 매매일자, 감정평가사 신원, 또는 그 밖의 관측 가능한 변수와 상관되는 잔차는 대응하는 잠재변수를 부분적으로 복원할 수 있음을 시사한다.

모형 선택. 관측변수를 더 넣는 것이 누락변수 편의에 대한 뻔한 처방이지만, 정작 가장 중요한 잠재변수는 모형에 추가하기가 가장 어려운 것들 (세밀한 상태, 최근 개량의 품질, 근린 미시효과) 인 경우가 많다. 방어 가능한 대응은 때때로 모형이 결코 그것들을 포착하지 못하리라는 점을 인정하고, 변수를 억지로 끼워 넣는 대신 설명되지 않은 분산을 잔차에 배분하는 것이다.

감정평가 실무 예제. 항목 005 의 회귀분석은 비교사례 H 에 대해 특이한 잔차를 낳았다. 다른 일곱 개 잔차가 $\pm \$18,000$ 대역 안에 든 반면, 모형은 비교사례 H 의 가격을 \$46,455 만큼 과소 예측했다. 비교사례 H 의 관측된 특성 — GLA 1,600, 대지 7,300, 조망 없음, 상태 “양호” — 은 그것을 표본의 한가운데에 확실히 위치시킨다. 관측 가능한 프로파일 가운데 유난히 높은 가격을 예고하는 것은 아무것도 없다. 그럼에도 그 부동산은 모형이 말하는 것보다 \$46,455 비싸게 팔렸다.

자연스러운 가설은 비교사례 H 가 가격에 기여하지만 자료에는 나타나지 않는 잠재 속성을 지녔다는 것이다. 본 호를 위한 데이터셋 구성에서 실제로 그러하다. 비교사례 H 는 최근 주방 리모델링을 거쳤고 이는 묶음 가치에 약 \$60,000 을 더한다. 리모델링은 회귀분석의 변수 목록에 없었으므로 그 기여분은 계수가 아니라 잔차에 쌓였다. 추정된 잔차 (\$46,455) 가 참된 잠재 기여분 (\$60,000) 보다 다소 작은 까닭은, 회귀분석이 잠재 효과의 작은 부분들을 다른 계수들에 흩뿌려 상관된 차원에서 그 계수들을 위쪽으로 편의시키기 때문이다.

여기서 세 가지 실무적 귀결이 따라 나온다:

이상치 잔차를 오차가 아니라 증거로 다루라. 비교사례 H 의 잔차에 대한 첫 번째 적절한 대응은 그것을 이상치로 버리는 것이 아니라, 그 부동산의 무엇이 과소 예측을 설명하는지 조사하는 것이다. 이 경우에는 현장 조사나 건축허가 기록 검색이 리모델링을 드러낼 것이다. 모형의 실패는 진단적 가치를 가진다.

다른 계수들도 약간 편의되어 있음을 인식하라. 조망, 상태, 대지의 계수는 비교사례 H 의 관측된 특성과의 상관을 통해 리모델링 효과의 일부를 흡수했다. $n = 8$ 에서 그 편의는 작고 비교사례와 공변량이 늘어나면 더 줄어든 것이나, 0 은 아니다.

잔차 전체는 복원 가능한 구조를 담고 있다. 이것이 필자가 발전시킨 잔차제약접근법 (Residual Constraint Approach, RCA) 방법론의 실질적 주장이다 (Craytor, 2025). 주거용 감정평가 자료에서 잔차는 체계적으로 iid 가우스 잡음이 아니다. 잔차는 미관측 속성 — 세밀한 상태, 세밀한 품질, 기능적 효용, 그 밖의 잠재 구성개념 — 의 구조화된 기여를 반영하며, 잔차 자체를 버릴 것이 아니라 모형 화해야 할 잠재변수 신호로 다룸으로써 부분적으로 복원할 수 있다.

RCA 가 상태와 품질을 다루는 방식은 이 두 층위 구조를 실행 가능하게 만든다. 감정평가사의 순서 등급 — 앞서 논의한 조립도 (coarse-grain) 대리변수로서, 상태-1(상태 조립도) 과 품질-1(품질 조립도) 로 표기할 수 있다 — 은 그 자체를 잠재변수로 취급해도 무방할 만큼 거칠지만, 그럼에도 회귀 변수로 투입하여 대략적인 잠재가격 기여분을 얻을 수 있다. 이어서 프로토콜은 그렇게 추정된 기여분을 매매 격자의 회귀분석 부분에서 측정된 속성 바깥으로 옮겨 다시 잔차로 되돌리고, 거기서 그것을 세밀도 (fine-grain) 잠재변수인 상태-2(상태 세밀도) 와 품질-2(품질 세밀도) 에 재배분한 뒤 MLS 공개 비교란 같은 추가 증거에 비추어 정교화한다. 조립도 등급은 초기 가치 배분을 공급하고, 세밀도 잠재변수가 복원 가능한 구조를 담는다. 최종적으로 세밀도 가치기여분이 조립도 가치기여분을 대체한다.

여기 제시한 단일 비교사례 예시는 RCA 직관의 가장 단순한 사례이다. 자료가 더 많아지고 모형 화가 풍부해지면 잔차 속 구조는 식별 가능한 잠재 요인들로 분할될 수 있으며, 각 요인은 표준적 해

도덕 설정이 닿지 못하는 방식으로 묶음의 시장가격에 기여한다. 잠재변수는 통제해서 없애야 할 성가신 것이 아니다. 그것은 방어 가능한 평가 모형화의 다음 전선이다.

참고문헌.

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)

상호 참조: 특성 공간 (*characteristics space*); 헤도닉 가격함수 (*hedonic price function*); 잠재가격 (*implicit price*); 헤도닉 회귀분석 (*hedonic regression*).

부록: Prolog 형식화

아래의 기계 판독 가능한 부속 파일 (006-latent-variable.pl) 은 본지의 Prolog 지식베이스에서 이 항목을 형식화한 것이다. 이 파일은 본문과 함께 열람 가능한 Prolog 갤리로도 발행된다.

```
%% 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).

```