# Латентная переменная

*Foundations • The Valuation Engineer*
Берт Крейтор• 8 августа 2026 г.
Версия 1.0.0
DOI: 10.67330/11n3g168

*Русский перевод, подготовленный с использованием искусственного интеллекта и прошедший редакционную проверку; английская версия является основной.*

---

**Формальное определение.** 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.

**Интуитивное представление.** В первых пяти статьях вектор характеристик**z** рассматривался так, как будто его можно полностью измерить: общая площадь, участок, вид из окна, состояние и т. д. — все эти характеристики представлены в виде столбцов данных. На практике ни один вектор характеристик никогда не бывает полным. Некоторые характеристики отсутствуют в данных, поскольку их сложно измерить (качество внутренней отделки, планировка , микролокация и микроориентация), либо отсутствуют в стандартных источниках данных (записи MLS не отражают качество освещения, звукоизоляцию или недавние несанкционированные улучшения), либо потому что они ещё не были признаны значимыми (особенности района, которые, как оказывается, позволяют прогнозировать цену только после того, как их выявляет анализ).

Эти ненаблюдаемые, но значимые характеристики являются *латентными переменными*. Они существуют, влияют на цену, но не представлены в столбцах нашей регрессии. Их влияние не исчезает из-за того, что они не измеряются; оно накапливается в остатках модели, где с точки зрения включённых переменных оно выглядит как «шум», но на самом деле представляет собой структурированный сигнал с точки зрения лежащего в основе процесса генерации данных.

Латентные переменные служат связующим звеном между идеализированной гедонической моделью и неоднозначной реальностью данных оценки. Их явное учет является необходимым условием для объективного анализа неопределенности модели, структуры остаточных величин и пределов того, что можно объяснить с помощью наблюдаемых характеристик.

**Где оценщик встречается с этим понятием.** Оценщики сталкиваются с латентными переменными в трёх типичных ситуациях.

**Корректировки по качеству и состоянию**. «Качество» и «состояние» как категории оценки сами по себе представляют собой попытку свести многомерный латентный конструкт к единой порядковой шкале. Оценка состояния C3 в UAD 3.6 призвана обобщить отложенный ремонт недвижимости, износ отделки, возраст инженерных систем и внешний вид — от четырёх до десяти базовых характеристик, сведённых к одному числу. Это одноцифровое обобщение является прокси-показателем скрытого конструкта; мастерство оценщика заключается в том, насколько точно этот прокси-показатель отражает суть дела в каждом конкретном случае.

**Необъяснимое отклонение в таблице**. Когда три скорректированных аналога указывают на стоимость, но их значения расходятся больше, чем ожидалось, это расхождение несет информационную ценность. Редко когда это полностью объясняется выборочным шумом; чаще всего оно отражает какую-то характеристику, которую оценщик еще не выявил или не учел при корректировке. Исследование причины расхождения остаточных значений — «почему этот

аналог указывает на цену $ 50,000 выше, чем остальные?» — является неформальной версией вывода о латентных переменных, применяемой оценщиком.

**Недавнее улучшение, которое никто не может увидеть**. В объекте недвижимости могли быть проведены ремонт кухни, замена крыши или построенное с разрешением дополнительное жилое помещение (ADU), что пока не отражено в данных MLS или оценщика. С точки зрения оценщика, это улучшение можно увидеть при личном осмотре, но оно отсутствует в данных. С точки зрения регрессионного анализа, это улучшение является скрытым. Оценщик должен сопоставить результаты личного наблюдения с показателями, основанными на данных, и оба источника должны сходиться в оценке истинного вклада этой скрытой характеристики.

**Почему это важно для обоснованности.** Латентные переменные вводят ряд соображений относительно обоснованности, которые аддитивная линейная модель, приведенная в записи 005, напрямую не учитывает.

**Систематическая ошибка, связанная с пропущенной переменной**. Если латентная переменная коррелирует с наблюдаемой переменной, её влияние проникает в коэффициент наблюдаемой переменной. Если недавно проданные дома в Пасифике, как правило, имеют отремонтированные кухни (скрытая характеристика), а дата заключения договора купли-продажи дома присутствует в данных, то коэффициент срока продажи (также известный как « `sale_age`» или «days since sold») будет включать в себя часть эффекта, связанного с ремонтом кухни . Оцененный неявный вклад цены « `sale_age`» становится смещенным — он отражает « `sale_age`» *плюс* часть неизмеренного статуса ремонта кухни . Обоснованная оценка учитывает, что коэффициенты наблюдаемых переменных зависят от того, что отсутствует в модели.

**Интерпретация остатков**. В стандартной теории регрессии остатки рассматриваются как гауссовый шум с нулевым средним. В подходе, основанном на латентных переменных, остатки рассматриваются как смесь шума и сигнала: часть представляет собой вариацию выборки, а часть — совокупный эффект ненаблюдаемых характеристик. Разграничение того и другого имеет существенное методологическое значение. Остаток, коррелирующий с местоположением, датой продажи, личностью оценщика или любой другой наблюдаемой переменной, указывает на то, что соответствующая латентная переменная можно частично восстановить.

**Выбор модели**. Добавление дополнительных наблюдаемых переменных — это очевидное средство борьбы с систематической ошибкой, связанной с пропущенными переменными, однако латентные переменные, имеющие наибольшее значение, зачастую являются именно теми, которые сложнее всего включить в модель (условия мелкомасштабного уровня, качество недавних улучшений, микроэффекты района). Иногда обоснованным решением является признание того, что модель никогда не сможет их учесть, и отнесение необъяснимой дисперсии к остатку, а не принудительная подгонка дополнительных переменных.

**Разобранный пример оценки.** Регрессия в записи 005 дала необычный остаток для Comp H: модель занизила прогноз цены Comp H на $ 46,455, в то время как остальные семь остатков уложились в диапазон от ±$18,000. Наблюдаемые характеристики объекта Comp H — общая арендуемая площадь (GLA) 1,600, участок 7,300, без вида, состояние «хорошее» — прочно помещают его в середину выборки; ничто в его наблюдаемом профиле не предсказывает необычно высокую цену. И тем не менее он был продан за $ 46,455 больше, чем, согласно модели, он должен был бы стоить.

Естественным предположением является то, что у компании H имеется скрытая характеристика, которая влияет на её цену, но не отражается в данных. При формировании набора данных для данного вопроса ситуация именно такова: в компании Comp H недавно был проведен ремонт кухни, который добавил примерно $60,000 к стоимости набора услуг. Ремонт не входил в список переменных регрессии, поэтому этот вклад накоплен в остатке, а не в коэффи-

циенте. Оцененное значение остатка ($ 46,455) немного меньше, чем истинный скрытый вклад ($ 60,000), поскольку регрессия распределяет небольшие доли скрытого эффекта по другим коэффициентам, смещая их в сторону увеличения в коррелирующих измерениях.

Из этого вытекают три практические последствия:

**Рассматривайте остаток, выпадающий за пределы диапазона, как доказательство, а не как ошибку**. Первым надлежащим ответом на остаток по Comp H является не исключение его как выпадающего значения, а выяснение, что именно в данном объекте недвижимости объясняет недооценку. В данном случае очный осмотр или поиск в архивах разрешений по строительству позволили бы выявить факт реконструкции. Несостоятельность модели имеет ценность для диагностики.

**Следует учитывать, что остальные коэффициенты имеют небольшую систематическую погрешность**. Коэффициенты «вид», «состояние» и «участок» поглотили небольшую часть эффекта реконструкции за счет корреляций с наблюдаемыми характеристиками объекта сравнения H. Эта погрешность невелика на сайте $n = 8$ и уменьшилась бы еще больше при увеличении количества аналогов и ковариат, но она не равна нулю.

**Остатки в целом обладают восстанавливаемой структурой**. В этом и заключается основной тезис методологии «Метода остаточных ограничений» (RCA), разработанной автором (Craytor, 2025): в данных по оценке жилой недвижимости остатки систематически не являются независимым и идентично распределенным гауссовым шумом. Они отражают структурированный вклад ненаблюдаемых характеристик — мелкозернистого состояния, мелкозернистого качества, функциональной полезности и других латентных конструктов — которые можно частично восстановить, рассматривая само остаток как сигнал латентной переменной, подлежащий моделированию, а не отбрасыванию.

Подход RCA к оценке состояния и качества делает эту двухуровневую структуру применимой на практике. Порядковые оценки оценщиков — те упрощённые прокси-показатели, о которых говорилось выше и которые можно обозначить как «состояние-1» (состояние-грубое) и качество-1 (качество-грубое) — настолько приблизительны, что их вполне можно рассматривать как латентные переменные сами по себе, но их всё же можно ввести в качестве регрессоров для получения приблизительных вкладов неявной цены. Затем протокол перемещает эти оцененные вклады из измеренных характеристик в разделе регрессии таблицы продаж обратно в остаток, откуда они перераспределяются в тонкозернистые латентные переменные состояние-2 (состояние-тонкозернистое) и качество-2 (качество-тонкозернистое) и уточняются с учетом дополнительных данных, таких как публичные комментарии в системе MLS. Рейтинг грубой детализации обеспечивает начальное присвоение значений; латентная переменная тонкой детализации несет в себе восстанавливаемую структуру. В конечном итоге вклад значений тонкой детализации заменяет вклад значений грубой детализации.

Приведенная здесь иллюстрация с использованием одной компоненты представляет собой простейший случай интуиции RCA. При наличии большего объема данных и более сложной модели структуру остаточной величины можно разбить на распознаваемые скрытые факторы, каждый из которых вносит вклад в рыночную цену набора ценных бумаг таким образом, который недоступен для стандартной гедонической спецификации. Латентные переменные — это не помеха, которую нужно устранить; они представляют собой новую границу обоснованного моделирования оценки.

*Список литературы.*

Спирман, К. (1904). «Общий интеллект»: объективное определение и измерение. *American Journal of Psychology* , 15(2), 201–292.

Боллен, К. А. (2002). Латентные переменные в психологии и социальных науках. *Ежегодный обзор психологии* , 53, 605–634.

Крейтор, Б. (2025). Метод остаточных ограничений: структура& Протокол. Zenodo.doi:10.5281/zenodo.14787917

---

*Перекрёстные ссылки:* *characteristics space*; *hedonic price function*; *implicit price*; *hedonic regression*.

---

## Приложение: формализация на Prolog

Следующий машиночитаемый вспомогательный файл (`006-latent-variable.pl`) формализует эту статью в базе знаний журнала на Prolog. Он также публикуется рядом с этой статьёй как отдельный просматриваемый файл (galley).

```
%% 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).
```