

Valuation Engineer Journal

价值的工程学——可测量、有节制、可复现。

第 1 卷 · 第 1 期 · 2026 年夏季



本期内容

房地产估价中学理与实务的落差

Residual Constraint Approach (RCA) —— 框架与流程

earthUI、glmnetUI 与 mgcvUI —— 开放的建模界面

基础概念 (F0–F6) 异质商品 · 特征空间 ·

特征价格函数 · 隐含价格 · 特征价格回归 · 潜变量

近期判例 美国 · 欧盟 · 拉丁美洲

Valuation Engineer Journal

第 1 卷 · 第 1 期

2026 年夏季

ISSN 3143-5033

主编 William Bert Craytor, SRA

目录

宗旨	2
社论	3
创刊号弁言	4
致编辑的信	6
邀您来信	7
论文	8
房地产估价中学理与实务的落差	9
Residual Constraint Approach (RCA) —— 框架与流程	38
earthUI——earth (MARS) 程序包的交互式界面	85
glmnetUI——glmnet (Lasso 与弹性网) 程序包的交互式界面	124
mgcvUI——mgcv (广义可加模型) 程序包的交互式界面	157
基础概念	192
总览 (F0)	193
异质商品 (F1)	199
特征空间 (F2)	206
特征价格函数 (F3)	213
隐含价格 (F4)	219
特征价格回归 (F5)	225
潜变量 (F6)	232
近期判例	238
近期判例——美国 (R1)	239
近期判例——欧盟 (R2)	242
近期判例——拉丁美洲 (R3)	246

宗旨

《*Valuation Engineer Journal*》刊载关于价值工程的严谨且面向实务的研究与案例：将技术判断与经济结果相连的方法、模型与测量。本刊尤重可复现性、假设的公开与论述的清晰。

社论

栏目：编者的话

创刊号致辞

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

William Bert Craytor, SRA — 主编

第 1 卷，第 1 期 · 2026 年夏季

发表日期：2026 年 7 月 1 日

DOI: 10.67330/0tymfa32

© 2026 William Bert Craytor. Published by Pacific Vista Net.

本开放获取文章依据 知识共享署名 4.0 国际许可协议 (CC BY 4.0) 的条款发布。

摘要。第 1 卷第 1 期的编者的话：本刊为何存在，五篇文章与三个栏目各自贡献了什么，以及一份共襄其事的邀约。

关键词：编者的话，估价工程，创刊号

欢迎阅读《*The Valuation Engineer Journal*》（估价工程师学刊）创刊号。

本刊之所以存在，源于一道鸿沟。估价理论长期以来把估价表述为对市场行为的严谨测量，然而日常实务在很大程度上仍然依靠人工挑选可比实例、以及仅凭判断估算的调整值。与此同时，那些能够测量市场究竟为房地产特征支付了多少的统计工具——特征价格回归 (hedonic regression)、现代样条模型 (spline model)、经交叉验证的机器学习——已经成熟了数十年，却极少被实际估价工作所采纳。“估价工程师” (valuation engineer) 是我们对弥合这道鸿沟的从业者的称呼：把持证估价师的专业能力，与数据工程、统计学以及可复现的作业规程结合在一起。

本期的五篇文章奠定了工作基础。开篇文章考察理论与实务之间的鸿沟本身，并论证本行业自身的准则早已要求比其日常方法所能提供的更多的测量。第二篇——也是本刊全部意图的核心——完整呈现了残差约束法 (Residual Constraint Approach)：这一规程让回归通道去测量可以测量的部分，把无法测量的部分规训为一份明确的残差分解，并对总额加以约束，使判断只能对价值加以细化，而绝不能凭空创造价值。随后是三篇软件文章，记录 earthUI 及其用于惩罚回归与广义可加模型的配套界面——本刊所依托的开源工具——它们在付印前不久合并为一个应用程序，各篇文章中的出版者按对此均有说明。

三个栏目与这些文章相伴。*Foundations* (基础) 栏目以七则短文 (F0–F6) 开篇，梳理出一条概念链：从异质商品出发，经由特征空间、特征价格函数、隐含价格与特征价格回归，直至潜藏于残差之中的潜变量。每则短文都注明其学术渊源，拥有各自的 DOI，并以一段机器可读的 Prolog 形式化表述收尾——这是一个知识库的开端，我们打算逐期将其培育成本领域精确而共通的术语体系。*Recent Court Decisions* (近期法院判例) 栏目综述美国、欧盟与拉丁美洲中与估价实务相关的裁判，因为法律环境本身就是这一工程问题的组成部分。*Letters to the Editor* (读者来信) 栏目开启通信专栏：为使用

而设计的方法，总会以其作者从未预料的方式受到检验，而这些发现——以及由此引发的分歧——正属于此处。

本期的每一项内容都以 CC BY 4.0 许可开放获取出版，同时提供 PDF 与 HTML 两种格式，配有 Crossref DOI，并在 Zenodo 上独立保存。我们诚邀投稿——文章、Foundations 条目、来往信函以及不同意见——欢迎估价师、经济学家与工程师共同参与。

William Bert Craytor, SRA

主编

致编辑的信

栏目：读者来信

诚邀来信

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

William Bert Craytor, SRA — 主编

第 1 卷，第 1 期 · 2026 年夏季

发表日期：2026 年 7 月 1 日

DOI: 10.67330/r55mw896

© 2026 William Bert Craytor. Published by Pacific Vista Net.

本开放获取文章依据 知识共享署名 4.0 国际许可协议 (CC BY 4.0) 的条款发布。

摘要。主编开启“读者来信”专栏：本刊希望从读者那里收到什么，以及如何寄送。

关键词：读者来信，通信往来，同行评议

一份只顾自己说话的刊物，只是半份刊物。本专栏正是为另一半而设。

The Valuation Engineer Journal (估价工程师学刊) 刊发的是意在被使用的方法——而被使用的方法，总会以其作者从未预料的方式受到检验。如果您应用过残差约束法 (Residual Constraint Approach)，并发现其中某一步在您所在的市场行不通，我们想要这样的来信。如果某则 Foundations (基础) 条目对某个术语的界定比您的实务所需更为狭窄，我们想要这样的来信。如果我们所综述的某项法院判例，在该辖区的从业者读来另有趣味，我们同样想要这样的来信。不同意见、复现、修正与纯粹的纠错，都在本专栏的范围之内；来自实务的简短报告也是——一项奏效的技术、一个已被解决的数据难题、一次值得提醒他人的失败。

规矩不多。来信应当署名、简明（一页为宜）、有的放矢——请注明您所针对的文章、章节或条目，并像在工作底稿中那样给出证据。我们可能会为篇幅与清晰起见作出编辑，但始终会征得作者同意。来信与本刊其他所有内容一样，依据 CC BY 4.0 许可开放获取出版；其作品被讨论的作者，将获得在同一期作出回应的机会。

请通过本刊网站 (journal.valuation-engineer.com) 或直接向主编寄送来信。本刊收到的第一批真正的来信，将标志着它从一份出版物变为一场对话的时刻——我们对此满怀期待。

William Bert Craytor, SRA

主编

论文

房地产估价中的理论与实践差距：功能、边界与张力的结构化阐述（简体中文版）

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

William Bert Craytor

第 1 卷 第 1 期 · 2026 年 5 月 21 日

发表于 2026 年 7 月 1 日

DOI: [10.67330/x7pbf13](https://doi.org/10.67330/x7pbf13)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

本开放获取文章依据[知识共享署名 4.0 国际许可协议 \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) 的条款发布。

摘要。 房地产估价事关重大公共利益，但其官方理论——以过去与当前条件下的市场价值意见为核心——与估价产品在当代金融流程中的实际用途之间的张力正日益加剧。本文分五个阶段构建这一论证，每个阶段均配有明确、可查询的 Prolog 知识库：前两个阶段确立估价所承担的社会与经济功能；第三个阶段指出其主要最终产品——市场价值意见；第四个阶段考察估价相对于量化预测与金融预测的时间取向和方法论取向，以及将两者分隔开来的监管边界。这些内容汇聚为第五个阶段的核心主张：在抵押贷款中，按贷款价值比折算的市场价值实际上充当着未来止赎价值的隐性替代指标——这是一种面向未来的用途，而现值理论并不予以承认。许多估价师并未意识到这一抵押品估值角色，因而未能以应有的审慎态度对待市场价值的未来走势——本文旨在揭示这一问题，并欢迎后续研究。

关键词： 房地产估价，市场价值，估价理论，知识表示，估价准则，止赎价值

目录

1 社会重要性	1
2 经济重要性	6
3 市场价值估计作为估价最终产品的重要性	10
4 估价活动的时间取向	13
5 市场价值 × LTV 作为止赎价值的替代指标	19
6 结论	25

1 社会重要性

房地产及其他资产估价通过帮助买方、卖方、贷款机构、保险公司、投资者、法院、税务机关以及公众，就财产及其他形式财富的转让、融资、税收与管理作出知情且公平的决定，发挥着不可或缺的社会与经济功能。高质量的估价以证据、分析与专业判断为基础，提供独立且论据充分的价值意见，从而促进市场透明度、公众信心、资本的高效配置以及金融机构的稳定。(LaCour-Little and Malpezzi 2003; K'Akumu and Larsen 2024; Hendershott and Kane 1995; Lee, Peng, and Liao 2020)

相反，低质量的估价可能在多个时间层面造成严重后果。短期内，不准确的估价可能使个别买方或卖方因买价过高、卖价过低、税负过重或不当的放贷决定而蒙受直接经济损失。中期内，系统性的估价缺陷可能扭曲市场信号、损害信贷审核、助长投机行为、导致不公平的税收或保

险承保，并加剧诉讼与交易层面的不信任。长期而言，估价质量的持续缺失可能动摇公众对市场与机构的信心，导致金融不稳定、资源错配、社区衰败或过度投资，并造成波及家庭、企业与社区的更广泛的社会不公。(LaCour-Little and Malpezzi 2003; Mayer and Nothhaft 2022)

由于房地产及其他资产往往占个人、企业与公共财富的相当大比重，估价质量不仅仅是私人层面的技术问题，而是关乎经济韧性、公平与长期公共福祉的重大公共利益事项。(Mayer and Nothhaft 2022)

```
%% appraisal_social.pl
%%
%% A Prolog knowledge base encoding the social and public-interest role
%% of real estate and other asset appraisals. Companion to
%% appraisal_economics.pl, which covers the strictly economic axis.
%%
%% All domain-specific predicates are prefixed with `social_` so this
%% module can be loaded alongside appraisal_economics.pl (and any other
%% companion KBs) without name collisions, and so cross-KB rules can
%% mention both vocabularies unambiguously.
%%
%% Conceptual model:
%%   - Appraisals support social DECISION TYPES.
%%   - Each decision involves identifiable PARTIES.
%%   - Appraisal QUALITY (high / low) drives EFFECTS across time HORIZONS.
%%   - Emphasis: equity, public confidence, public interest.
%%
%% Author: generated for Bert Craytor, ValEngr / RCA context.

:- module(appraisal_social,
    [ social_horizon/1
    , social_party/1
    , social_decision_type/1
    , social_involves/2           % DecisionType, Party
    , social_effect/4           % Quality, Horizon, Domain, Effect
    , social_contribution/1
    , social_hallmark/1
    , social_equity_principle/1
    , social_public_interest_claim/1
    , social_positive_effect/3
    , social_negative_effect/3
    , social_horizon_summary/2
    , social_consequences_of_quality/2
    , social_parties_affected_by/2
    , social_public_harm/2
    , social_private_harm/2
    , social_appraisal_is_public_interest/0
    ]).

:- discontiguous social_effect/4.
:- discontiguous social_involves/2.

%% -----
%% 1. Time horizons
%% -----

social_horizon(short_term).
social_horizon(medium_term).
social_horizon(long_term).
```

```

%% -----
%% 2. Parties whose decisions depend on appraisals
%%
%% The opening sentence enumerates these explicitly.
%% -----

social_party(buyers).
social_party(sellers).
social_party(lenders).
social_party(insurers).
social_party(investors).
social_party(courts).
social_party(taxing_authorities).
social_party(general_public).

%% -----
%% 3. Types of decisions appraisals inform
%% -----

social_decision_type(property_transfer).
social_decision_type(financing).
social_decision_type(taxation).
social_decision_type(stewardship).
social_decision_type(insurance_coverage).
social_decision_type(adjudication).          % courts: condemnation, divorce, estate
social_decision_type(public_administration).

social_involves(property_transfer,    buyers).
social_involves(property_transfer,    sellers).
social_involves(financing,            lenders).
social_involves(financing,            buyers).
social_involves(taxation,              taxing_authorities).
social_involves(taxation,              general_public).
social_involves(stewardship,          investors).
social_involves(stewardship,          general_public).
social_involves(insurance_coverage,   insurers).
social_involves(insurance_coverage,   buyers).
social_involves(adjudication,          courts).
social_involves(public_administration, taxing_authorities).
social_involves(public_administration, general_public).

%% -----
%% 4. What high-quality appraisals contribute (second sentence)
%% -----

social_contribution(market_transparency).
social_contribution(public_confidence).
social_contribution(efficient_capital_allocation).
social_contribution(stability_of_financial_institutions).

%% Characteristics of a high-quality appraisal opinion of value.
social_hallmark(independence).
social_hallmark(well_supported).
social_hallmark(grounded_in_evidence).
social_hallmark(grounded_in_analysis).
social_hallmark(grounded_in_professional_judgment).

%% -----

```

```

%% 5. Effects of appraisal quality by horizon
%% -----

%% ---- SHORT TERM: high quality ----
social_effect(high, short_term, property_transfer, fair_pricing_for_parties).
social_effect(high, short_term, financing, appropriate_lending_decisions).
social_effect(high, short_term, taxation, accurate_individual_assessment).

%% ---- SHORT TERM: low quality ----
social_effect(low, short_term, property_transfer, buyer_overpayment).
social_effect(low, short_term, property_transfer, seller_underselling).
social_effect(low, short_term, taxation, excessive_individual_taxation).
social_effect(low, short_term, financing,
↳ inappropriate_lending_decisions).
social_effect(low, short_term, property_transfer,
↳ immediate_financial_harm_to_individuals).

%% ---- MEDIUM TERM: high quality ----
social_effect(high, medium_term, markets, clear_market_signals).
social_effect(high, medium_term, financing, sound_credit_underwriting).
social_effect(high, medium_term, taxation, equitable_tax_assessment).
social_effect(high, medium_term, insurance_coverage,
↳ equitable_insurance_coverage).
social_effect(high, medium_term, adjudication, reduced_litigation_pressure).

%% ---- MEDIUM TERM: low quality ----
social_effect(low, medium_term, markets, distorted_market_signals).
social_effect(low, medium_term, financing,
↳ impaired_credit_underwriting).
social_effect(low, medium_term, markets,
↳ encouragement_of_speculation).
social_effect(low, medium_term, taxation, inequitable_taxation).
social_effect(low, medium_term, insurance_coverage,
↳ inequitable_insurance_coverage).
social_effect(low, medium_term, adjudication, increased_litigation).
social_effect(low, medium_term, markets, transactional_distrust).

%% ---- LONG TERM: high quality ----
social_effect(high, long_term, institutions, public_confidence_in_institutions).
social_effect(high, long_term, markets, public_confidence_in_markets).
social_effect(high, long_term, institutions, financial_stability).
social_effect(high, long_term, communities, equitable_community_outcomes).
social_effect(high, long_term, public_interest, economic_resilience).
social_effect(high, long_term, public_interest, social_fairness).
social_effect(high, long_term, public_interest, long_term_public_good).

%% ---- LONG TERM: low quality ----
social_effect(low, long_term, markets, eroded_market_confidence).
social_effect(low, long_term, institutions, eroded_institutional_confidence).
social_effect(low, long_term, institutions, financial_instability).
social_effect(low, long_term, markets, resource_misallocation).
social_effect(low, long_term, communities, neighborhood_decline).
social_effect(low, long_term, communities, neighborhood_overinvestment).
social_effect(low, long_term, communities, social_inequities_for_households).
social_effect(low, long_term, communities, social_inequities_for_businesses).
social_effect(low, long_term, communities, social_inequities_for_communities).

%% -----

```

```

%% 6. Equity principles and public-interest claims
%% -----

social_equity_principle(informed_decisions).
social_equity_principle(equitable_decisions).
social_equity_principle(fair_treatment_of_parties).
social_equity_principle(independence_of_valuer).

social_public_interest_claim(
    'Real estate and other assets are a substantial portion of personal, corporate,
    ↪ and public wealth.').
social_public_interest_claim(
    'Appraisal quality is not merely a private technical concern.').
social_public_interest_claim(
    'Appraisal quality is tied to economic resilience.').
social_public_interest_claim(
    'Appraisal quality is tied to fairness.').
social_public_interest_claim(
    'Appraisal quality is tied to the long-term public good.').

%% Top-level conclusion of the passage.
social_appraisal_is_public_interest :-
    social_public_interest_claim(_).    % at least one such claim exists

%% -----
%% 7. Derived / query predicates
%% -----

social_positive_effect(Horizon, Domain, Effect) :-
    social_effect(high, Horizon, Domain, Effect).

social_negative_effect(Horizon, Domain, Effect) :-
    social_effect(low, Horizon, Domain, Effect).

social_horizon_summary(Horizon, Effects) :-
    findall(Quality-Effect,
        social_effect(Quality, Horizon, _Domain, Effect),
        Effects).

social_consequences_of_quality(Quality, List) :-
    findall(consequence(Horizon, Domain, Effect),
        social_effect(Quality, Horizon, Domain, Effect),
        List).

social_parties_affected_by(DecisionType, Parties) :-
    setof(P, social_involves(DecisionType, P), Parties).

%% Harms that fall on the public rather than on transaction principals.
social_public_harm(Horizon, Effect) :-
    social_effect(low, Horizon, Domain, Effect),
    member(Domain, [markets, institutions, communities, public_interest,
    ↪ taxation]).

%% Private (transaction-level) harms.
social_private_harm(Horizon, Effect) :-
    social_effect(low, Horizon, Domain, Effect),
    member(Domain, [property_transfer, financing, insurance_coverage]).

```

2 经济重要性

房地产及其他资产估价通过支持信贷市场、投资决策、税收制度、财务报告、保险承保、遗产管理、公司交易以及公共政策规划的高效运行，发挥着基础性的经济作用。可靠的估价有助于建立可信的抵押品价值与市场价值度量，使资本得以在借款人、贷款机构、投资者、政府与企业之间更高效地流动。(LaCour-Little and Malpezzi 2003; Ebrahim and Hussain 2010)

短期内，准确的估价可以降低交易不确定性、改善放贷决策、促进公平谈判，并有助于防止因高估或低估而产生的直接经济损失。中期内，估价质量通过影响抵押贷款风险、投资配置、税收公平、保险充足性以及市场参与者的定价行为，左右着更广泛市场的稳定与效率。在此期间，薄弱或存在系统性偏差的估价可能助长资产泡沫膨胀、损害贷款组合、扭曲开发激励，并加大违约或诉讼风险。(LaCour-Little and Malpezzi 2003; Hendershott and Kane 1992; Griffin and Maturana 2016)

长期来看，估价质量的累积经济后果更为显著，影响着资本形成模式、基础设施投资、实际财富转移、公共收入、住房可负担性以及整个金融体系的韧性。估价实践的长期缺陷可能导致资本的慢性错配、银行体系不稳定、投资者信心下降以及土地与生产性资产的低效利用；而严格的估价准则则有助于支撑稳定的市场、稳健的经济规划和长期经济增长。(Ghent, Torous, and Valkanov 2019; Hendershott and Kane 1992; Ebrahim and Hussain 2010)

```
%% appraisal_economics.pl
%%
%% A Prolog knowledge base encoding the economic role of real estate
%% and other asset appraisals across short-, medium-, and long-term
%% horizons, including both the consequences of high-quality appraisal
%% practice and the consequences of weak or biased practice.
%%
%% All domain-specific predicates are prefixed with `economic_` so this
%% module can be loaded alongside appraisal_social.pl (and any other
%% companion KBs) without name collisions, and so cross-KB rules can
%% mention both vocabularies unambiguously.
%%
%% Conceptual model:
%%   - Appraisals support a set of economic FUNCTIONS.
%%   - Each function operates over one or more time HORIZONS.
%%   - Appraisal QUALITY (high | low) drives EFFECTS on those functions.
%%   - Effects are POSITIVE (benefits) when quality is high and
%%     NEGATIVE (harms) when quality is low/systematically biased.
%%
%% Author: generated for Bert Craytor, ValEngr / RCA context.

:- module(appraisal_economics,
    [ economic_horizon/1
    , economic_function/1
    , economic_supports/2
    , economic_effect/4
    ↪ Effect
    , economic_stakeholder/1
    , economic_benefits/2
    , economic_positive_effect/3
    , economic_negative_effect/3
    , economic_horizon_summary/2
    , economic_consequences_of_quality/2
    , economic_functions_at_horizon/2
    , economic_stakeholders_affected_by/2
    , economic_reliable_appraisals_improve/1
    , economic_appraisal_quality_matters_at/1
    ]).
```

```

:- discontinuous economic_effect/4.
:- discontinuous economic_supports/2.

%% -----
%% 1. Time horizons
%% -----

economic_horizon(short_term).
economic_horizon(medium_term).
economic_horizon(long_term).

%% -----
%% 2. Economic functions supported by appraisals
%% -----

economic_function(credit_markets).
economic_function(investment_decisions).
economic_function(taxation_systems).
economic_function(financial_reporting).
economic_function(insurance_underwriting).
economic_function(estate_administration).
economic_function(corporate_transactions).
economic_function(public_policy_planning).

%% -----
%% 3. Stakeholders whose decisions depend on appraisals
%% -----

economic_stakeholder(borrowers).
economic_stakeholder(lenders).
economic_stakeholder(investors).
economic_stakeholder(governments).
economic_stakeholder(businesses).
economic_stakeholder(insurers).
economic_stakeholder(taxpayers).
economic_stakeholder(homeowners).

%% Capital flows more efficiently when appraisals are reliable.
economic_benefits(borrowers, credit_markets).
economic_benefits(lenders, credit_markets).
economic_benefits(investors, investment_decisions).
economic_benefits(governments, taxation_systems).
economic_benefits(governments, public_policy_planning).
economic_benefits(businesses, corporate_transactions).
economic_benefits(businesses, financial_reporting).
economic_benefits(insurers, insurance_underwriting).
economic_benefits(taxpayers, taxation_systems).
economic_benefits(homeowners, credit_markets).

%% -----
%% 4. Which functions are most salient at each horizon
%% -----

economic_supports(credit_markets, short_term).
economic_supports(investment_decisions, short_term).
economic_supports(corporate_transactions, short_term).

```

```

economic_supports(credit_markets,      medium_term).
economic_supports(investment_decisions, medium_term).
economic_supports(taxation_systems,    medium_term).
economic_supports(insurance_underwriting, medium_term).

economic_supports(credit_markets,      long_term).
economic_supports(taxation_systems,    long_term).
economic_supports(public_policy_planning, long_term).
economic_supports(financial_reporting,  long_term).
economic_supports(estate_administration, long_term).

%% -----
%% 5. Effects of appraisal quality, by horizon and function
%% -----

%% ---- SHORT TERM: high quality ----
economic_effect(high, short_term, credit_markets,
↳ reduced_transactional_uncertainty).
economic_effect(high, short_term, credit_markets,
↳ improved_lending_decisions).
economic_effect(high, short_term, corporate_transactions, fair_negotiations).
economic_effect(high, short_term, investment_decisions,
↳ prevention_of_immediate_losses).

%% ---- SHORT TERM: low quality ----
economic_effect(low, short_term, credit_markets,
↳ transactional_uncertainty).
economic_effect(low, short_term, credit_markets,      poor_lending_decisions).
economic_effect(low, short_term, investment_decisions,
↳ losses_from_overvaluation).
economic_effect(low, short_term, investment_decisions,
↳ losses_from_undervaluation).

%% ---- MEDIUM TERM: high quality ----
economic_effect(high, medium_term, credit_markets,      stable_mortgage_risk).
economic_effect(high, medium_term, investment_decisions,
↳ efficient_investment_allocation).
economic_effect(high, medium_term, taxation_systems,      tax_equity).
economic_effect(high, medium_term, insurance_underwriting,
↳ adequate_insurance_coverage).
economic_effect(high, medium_term, investment_decisions,
↳ rational_market_pricing).

%% ---- MEDIUM TERM: low quality ----
economic_effect(low, medium_term, investment_decisions,      inflated_asset_bubbles).
economic_effect(low, medium_term, credit_markets,
↳ impaired_lending_portfolios).
economic_effect(low, medium_term, public_policy_planning,
↳ distorted_development_incentives).
economic_effect(low, medium_term, credit_markets,      increased_default_risk).
economic_effect(low, medium_term, corporate_transactions,
↳ increased_litigation_risk).

%% ---- LONG TERM: high quality ----
economic_effect(high, long_term, investment_decisions,      sound_capital_formation).
economic_effect(high, long_term, public_policy_planning,
↳ sound_infrastructure_investment).

```

```

economic_effect(high, long_term, estate_administration,
↳ orderly_intergenerational_wealth_transfer).
economic_effect(high, long_term, taxation_systems,      stable_public_revenues).
economic_effect(high, long_term, public_policy_planning,
↳ supported_housing_affordability).
economic_effect(high, long_term, credit_markets,
↳ financial_system_resilience).
economic_effect(high, long_term, investment_decisions,
↳ long_term_economic_growth).

%% ---- LONG TERM: low quality ----
economic_effect(low, long_term, investment_decisions,
↳ chronic_capital_misallocation).
economic_effect(low, long_term, credit_markets,      banking_instability).
economic_effect(low, long_term, investment_decisions,
↳ reduced_investor_confidence).
economic_effect(low, long_term, public_policy_planning, inefficient_land_use).
economic_effect(low, long_term, public_policy_planning,
↳ inefficient_productive_asset_use).

%% -----
%% 6. Derived / query predicates
%% -----

economic_positive_effect(Horizon, Function, Effect) :-
    economic_effect(high, Horizon, Function, Effect).

economic_negative_effect(Horizon, Function, Effect) :-
    economic_effect(low, Horizon, Function, Effect).

economic_horizon_summary(Horizon, Effects) :-
    findall(Quality-Effect,
        economic_effect(Quality, Horizon, _Function, Effect),
        Effects).

economic_consequences_of_quality(Quality, List) :-
    findall(consequence(Horizon, Function, Effect),
        economic_effect(Quality, Horizon, Function, Effect),
        List).

economic_functions_at_horizon(Horizon, Functions) :-
    setof(F, economic_supports(F, Horizon), Functions).

economic_stakeholders_affected_by(Function, Stakeholders) :-
    setof(S, economic_benefits(S, Function), Stakeholders).

%% Overall claim: reliable appraisals improve allocative efficiency.
economic_reliable_appraisals_improve(capital_flow_efficiency).
economic_reliable_appraisals_improve(market_stability).
economic_reliable_appraisals_improve(long_term_economic_growth).

%% Top-level qualitative claim: appraisal quality matters at every horizon.
economic_appraisal_quality_matters_at(Horizon) :-
    economic_horizon(Horizon),
    \+ \+ economic_effect(high, Horizon, _, _),
    \+ \+ economic_effect(low, Horizon, _, _).

```

3 市场价值估计作为估价最终产品的重要性

市场价值估价在抵押品风险分析中发挥着重要作用，因为它就财产或资产在正常市场条件下可能实现的价格提供了专业形成的估计，从而为评估信贷敞口与潜在止赎结果的贷款机构、投资者和抵押品分析师提供了基础性基准。(Agarwal, Ben-David, and Yao 2015) 尽管由于不利的销售条件、缩短的出售周期、法律费用、实物损耗或负面市场情绪，止赎价值或清算价值可能与市场价值相差甚远，抵押品分析师仍经常以市场价值估价作为在压力情景下估算回收潜力的起点。(Campbell, Giglio, and Pathak 2011; Pennington-Cross 2006; Clauretie and Daneshvary 2009)

通过分析市场价值、贷款余额、贷款价值比 (LTV)、市场趋势、房产状况、流动性以及当地供求状况之间的关系，抵押品分析师评估借款人违约情况下损失发生的可能性及其严重程度。(Qi and Yang 2009)

短期内，准确的市场价值估价有助于贷款机构作出审慎的审贷决定并建立适当的抵押品安全边际。中期内，它们通过帮助机构评估经济周期中抵押品充足性的变化，支持组合风险管理、贷款证券化、准备金配置与监管合规。长期而言，持续可靠的估价实践通过降低过度杠杆、担保不足的放贷以及抵押资产系统性错误定价的风险，为金融体系稳定作出贡献——这些风险会在市场承压时期放大止赎损失和更广泛的经济动荡。(Eriksen et al. 2019)

```
%% appraisal_collateral.pl
%%
%% A Prolog knowledge base encoding the role of market value (MV)
%% appraisals in collateral risk analysis. Companion to
%% appraisal_economics.pl and appraisal_social.pl.
%%
%% Naming convention:
%%   - Predicates that have direct counterparts in the companion KBs
%%     (horizon, effect, summary, consequences, etc.) are prefixed
%%     `collateral_` so this module can be loaded alongside the others
%%     without name collisions, and so bridge rules can mention each
%%     vocabulary unambiguously.
%%   - Predicates that are unique to the collateral-analysis domain
%%     (value_basis, divergence_factor, risk_input, etc.) are NOT
%%     prefixed. They have no analogues in the other KBs and no
%%     collision risk, so the prefix would only add noise.

:- module(appraisal_collateral,
    [ collateral_horizon/1
      , collateral_analyst/1
      , collateral_effect/4                                % Quality, Horizon, Domain,
      ↪ Effect
      , collateral_positive_effect/3
      , collateral_negative_effect/3
      , collateral_horizon_summary/2
      , collateral_consequences_of_quality/2
      , value_basis/1
      , uses/2                                              % Analyst, ValueBasis
      , divergence_factor/1
      , stress_scenario/1
      , risk_input/1
      , risk_output/1
      , inputs_for/2
      , mv_is_benchmark_for/1
    ]).

:- disjoint collateral_effect/4.
:- disjoint uses/2.
```

```

%% -----
%% 1. Time horizons (namespaced -- parallel to economic/social_)
%% -----

collateral_horizon(short_term).
collateral_horizon(medium_term).
collateral_horizon(long_term).

%% -----
%% 2. Who uses MV appraisals in collateral risk analysis
%% -----

collateral_analyst(lenders).
collateral_analyst(investors).
collateral_analyst(collateral_analysts).
collateral_analyst(portfolio_risk_managers).
collateral_analyst(regulators).
collateral_analyst(securitization_arrangers).

%% -----
%% 3. Value bases distinguished in the passage
%% -----

value_basis(market_value).           % "normal market conditions"
value_basis(foreclosure_value).       % distressed
value_basis(liquidation_value).       % distressed, accelerated

%% Which value bases each analyst type works with.
uses(lenders, market_value).
uses(investors, market_value).
uses(collateral_analysts, market_value).
uses(collateral_analysts, foreclosure_value).
uses(collateral_analysts, liquidation_value).
uses(portfolio_risk_managers, market_value).
uses(securitization_arrangers, market_value).
uses(regulators, market_value).

%% Top-level claim of the opening sentence.
mv_is_benchmark_for(credit_exposure).
mv_is_benchmark_for(foreclosure_outcomes).
mv_is_benchmark_for(recovery_potential).

%% -----
%% 4. Why foreclosure / liquidation value diverges from MV
%% -----

divergence_factor(distressed_marketing_conditions).
divergence_factor(accelerated_sale_period).
divergence_factor(legal_costs).
divergence_factor(physical_deterioration).
divergence_factor(adverse_market_sentiment).

%% Stressed scenarios in which divergence factors apply.
stress_scenario(borrower_default).
stress_scenario(foreclosure).
stress_scenario(forced_liquidation).
stress_scenario(market_downturn).

```

```

%% -----
%% 5. Inputs the collateral analyst combines with MV, and the
%%      outputs they produce
%% -----

risk_input(market_value).
risk_input(loan_balance).
risk_input(loan_to_value_ratio).
risk_input(market_trends).
risk_input(property_condition).
risk_input(liquidity).
risk_input(local_supply_demand).

risk_output(likelihood_of_loss).
risk_output(severity_of_loss).
risk_output(recovery_estimate).
risk_output(loss_given_default).

%% inputs_for(+Output, -Inputs)
%% All risk inputs are used to produce each risk output.
inputs_for(Output, Inputs) :-
    risk_output(Output),
    setof(I, risk_input(I), Inputs).

%% -----
%% 6. Effects of appraisal quality by horizon
%% -----

%% ---- SHORT TERM: high quality ----
collateral_effect(high, short_term, underwriting, prudent_underwriting_decisions).
collateral_effect(high, short_term, underwriting, appropriate_collateral_margins).

%% ---- SHORT TERM: low quality ----
collateral_effect(low, short_term, underwriting,
    ↪ imprudent_underwriting_decisions).
collateral_effect(low, short_term, underwriting, inadequate_collateral_margins).
collateral_effect(low, short_term, underwriting, mispriced_credit_exposure).

%% ---- MEDIUM TERM: high quality ----
collateral_effect(high, medium_term, portfolio,
    ↪ effective_portfolio_risk_management).
collateral_effect(high, medium_term, portfolio, sound_loan_securitization).
collateral_effect(high, medium_term, portfolio, appropriate_reserve_allocation).
collateral_effect(high, medium_term, regulatory, regulatory_compliance).
collateral_effect(high, medium_term, portfolio,
    ↪ cycle_aware_collateral_evaluation).

%% ---- MEDIUM TERM: low quality ----
collateral_effect(low, medium_term, portfolio,
    ↪ weakened_portfolio_risk_management).
collateral_effect(low, medium_term, portfolio, mispriced_securitizations).
collateral_effect(low, medium_term, portfolio, inadequate_loss_reserves).
collateral_effect(low, medium_term, regulatory, regulatory_noncompliance_risk).
collateral_effect(low, medium_term, portfolio, misjudged_collateral_adequacy).

%% ---- LONG TERM: high quality ----
collateral_effect(high, long_term, system, financial_system_stability).
collateral_effect(high, long_term, system, restraint_on_excessive_leverage).

```

```

collateral_effect(high, long_term, system, well_secured_lending).
collateral_effect(high, long_term, system, accurate_systemic_collateral_pricing).
collateral_effect(high, long_term, system, dampened_foreclosure_amplification).

%% ---- LONG TERM: low quality ----
collateral_effect(low, long_term, system, financial_system_instability).
collateral_effect(low, long_term, system, excessive_leverage).
collateral_effect(low, long_term, system, poorly_secured_lending).
collateral_effect(low, long_term, system, systemic_collateral_mispricing).
collateral_effect(low, long_term, system, amplified_foreclosure_losses).
collateral_effect(low, long_term, system,
↳ amplified_economic_disruption_under_stress).

%% -----
%% 7. Derived / query predicates (namespaced)
%% -----

collateral_positive_effect(Horizon, Domain, Effect) :-
    collateral_effect(high, Horizon, Domain, Effect).

collateral_negative_effect(Horizon, Domain, Effect) :-
    collateral_effect(low, Horizon, Domain, Effect).

collateral_horizon_summary(Horizon, Effects) :-
    findall(Quality-Effect,
        collateral_effect(Quality, Horizon, _Domain, Effect),
        Effects).

collateral_consequences_of_quality(Quality, List) :-
    findall(consequence(Horizon, Domain, Effect),
        collateral_effect(Quality, Horizon, Domain, Effect),
        List).

```

4 估价活动的时间取向

估价师通常就某一特定生效日期形成时点性的市场价值意见——回顾性的（过去）、当前的，或在明确限制下前瞻性的（未来）——而不是对较长时间段作出概率性预测。这种时间与方法论取向是受监管的估价实践与量化分析师、金融分析师、投资分析师及经济学家的工作之间的核心区别之一。（Quan and Quigley 1991）

在大多数估价业务中，特别是依据《专业估价实务统一准则》（Uniform Standards of Professional Appraisal Practice, USPAP）或《国际估价准则》（International Valuation Standards, IVS）等准则进行的市场价值估价，估价师给出的是一个经调和的点估计（或在允许的情形下给出一个窄幅区间），锚定于某一生效日期的可观察市场证据、当时的市场条件和公认的方法。USPAP 明确允许价值结论以具体金额、数值区间或与某一基准的关系的形式给出，但经调和的意见仍然是与该日期挂钩的一个可信的点位指示。（The Appraisal Foundation 2024; International Valuation Standards Council 2025）

尽管估价师必然会纳入市场参与者对趋势、供求、折旧、资本化率或折现率以及去化速度的合理预期，但这样做是为了支持生效日期上有市场依据的价值，而非预测投机性波动。前瞻性价值意见（例如项目建成后的稳定价值）需要明示的特殊假设或假定条件，并且仍被限定于一个明确的未来生效日期，且须有源自市场的支持。

相比之下，量化分析师、金融分析师与投资分析师经常通过覆盖各时间段内一系列可能结果的估计来处理不确定性，运用概率分布、蒙特卡洛模拟、情景分析、敏感性检验或风险价值（VaR）等度量。他们的关注点往往是预期的未来表现、经风险调整的收益，或特定主体在不同期限内的投资价值，而不是某一日期上的单一市场价值点位。（Hull 2018）

这类工作还预设了估价师执照既不要求也不认证的技术能力。例如，拟合多元自适应回归样条 (MARS) 模型需要控制模型复杂度的判断力，使得强劲的样本内拟合 (高 R^2) 不是以过拟合为代价换来的，而是能够延续到留出数据上 (高交叉验证 R^2)，而这又以对底层算法的切实理解为前提。(Friedman 1991; Steurer, Hill, and Pfeifer 2021) 除这类量化能力之外，这项工作还要求具备处理高度结构化概念、辨析细微语义差别以及吸收陌生方法的概念能力。这些是技能而非资质：持照估价师可能具备它们，无照分析师也可能欠缺它们——这正是它们本身并不能决定某项任务落在监管界线哪一侧的原因。

估价师也可以从事有限的面向未来的工作，例如在折现现金流分析中预测收入流、空置率、费用和归复价值，或在开发或诉讼情境中估计前瞻性价值。只要立足于当前市场证据，并遵循强调可支持性、避免无依据投机的专业准则，这些内容就仍处于估价的边界之内。(Diaz and Wolverton 1998; Gallimore 1996; Crosby, Lizieri, and McAllister 2010) 现代估价实践越来越多地融合统计工具、情景考量与风险分析——尤其是在批量估价、自动估价模型或组合业务中——但核心交付成果通常仍是时点性的市场价值意见。(International Association of Assessing Officers 2017; Office of the Comptroller of the Currency and Board of Governors of the Federal Reserve System and Federal Deposit Insurance Corporation and National Credit Union Administration and Consumer Financial Protection Bureau and Federal Housing Finance Agency 2024; Alexandrov, Goodman, and Neal 2023)

可以这样概括：估价实践主要是证据导向的，面向从过去到现在 (或有限前瞻) 的点估计；而金融与量化分析则更具概率性、分布性，并且明确面向各个时间跨度上的未来。(Geltner 1991) 估价师问的是：“在现有证据和既定假设下，截至指定生效日期，最有依据的市场价值指示是什么？”金融与量化分析师则更常问：“未来一段时期内可能出现何种结果区间、概率几何，以及应如何配置资源？”

这条边界并非绝对，并且随着市场复杂性的提高而持续演变。尽管如此，在目的 (市场价值与投资价值之别)、证据要求和监管框架方面，根本性差异依然存在。

然而，真正起作用的界线不是由技能而是由法律划定的：决定性的问题只是某项任务在法律上是否要求持有估价师执照。量化分析师、金融分析师、经济学家、投资分析师、投资组合经理及相关专业人士可以在不持有房地产估价师执照的情况下，合法地估计、预测、建模或分析未来资产价值——包括分布式或区间式的预测——只要该工作不构成受监管的估价。(Murphy 2012)

关键区别在于：

1. 价值类型 (例如市场价值点估计，还是投资价值或概率性预测)；
2. 时间焦点 (特定生效日期，还是某一时间段或期限)；
3. 输出形式 (点估计或窄幅区间，还是完整的概率分布)；
4. 预期用途以及该工作是否以估价的名义出具；以及
5. 是否涉及要求遵守 USPAP 等准则的受监管放贷或法律情境。

这一框架在很大程度上源于美国 1989 年《金融机构改革、复苏和执行法》(FIRREA) 等银行监管法规，其目的在于确保联邦相关交易能够获得可信、独立的市场价值意见。(United States Congress 1989; K'Akumu and Larsen 2024) 许多高度复杂的估值活动——股票研究目标价、可行性研究、组合风险建模、期权定价或预测分析——都不在执照管辖范围之内，因为它们服务于投资、预测或咨询目的，而非受监管的估价职能。

```
%% appraisal_scope_regulation.pl
%%
%% Knowledge base for the boundary of regulated appraisal practice.
%%
%% TWO TIERS, deliberately kept separate:
%%
%% TIER 1 -- THE DEFINING TEST (regulation cluster). A task is a
%% regulated appraisal IFF it is one for which an appraiser's
%% license is legally required. requires_license/1 is that test;
%% it -- and only it -- decides regulated_appraisal/1 vs.
%% unregulated_valuation/1.
```

```

%%
%% TIER 2 -- DESCRIPTIVE TENDENCIES (scope cluster). How the two
%% populations -- `appraisal' and `quant_financial' -- typically
%% differ in approach, temporal scope, output form, and skill set.
%% These are CORRELATES, not the boundary: the populations overlap
%% (a licensed appraiser may do quant work; a quant may be
%% unlicensed), and nothing here determines licensure. No claim is
%% made about intelligence -- only about training and methodology.
%%
%% `temporal scope' and `approach' are DISTINCT axes: an approach has a
%% typical (defeasible) temporal mode, but the two are not the same
%% viewpoint -- which is why an appraiser using an AVM (appraisal
%% approach, distributional methods) is possible.

:- module(appraisal_scope_regulation,
    [ % --- TIER 1: the defining test (regulation) ---
      regulation_framework/1
    , regulation_framework_origin/2
    , regulation_distinguishing_factor/1
    , regulation_triggers_licensure/1
    , regulation_definition_of_appraisal/1
    , regulation_unlicensed_activity/2
    , regulation_licensed_activity/2
    , regulation_legal_exposure_condition/1
    , requires_license/1
    , exempt_from_appraisal_license/1
    , regulated_appraisal/1
    , unregulated_valuation/1
    , regulatory_asymmetry/2

      % --- TIER 2: descriptive tendencies (scope) ---
    , valuation_approach/1
    , scope_profession/1
    , scope_orientation/2
    , temporal_scope/1
    , temporal_mode/1
    , scope_temporal_focus/2
    , scope_output_form/2
    , scope_skill/1
    , scope_skill_demanded_by/2
    , scope_inputs_considered/2
    , scope_future_oriented_context/2
    , scope_professional_constraint/1
    , scope_contrast/4
    , scope_question/2
    , scope_convergence_area/1
    ]),

:- discontiguous regulation_unlicensed_activity/2.
:- discontiguous regulation_licensed_activity/2.
:- discontiguous scope_contrast/4.

%% =====
%% TIER 1 -- THE DEFINING TEST (the boundary)
%% =====

regulation_framework(firrea_1989).
regulation_framework(uspap).

```

```

regulation_framework(ivs).
regulation_framework(state_licensing).
regulation_framework(state_certification).

regulation_framework_origin(firrea_1989,      savings_and_loan_crisis).
regulation_framework_origin(uspap,           firrea_1989).
regulation_framework_origin(state_licensing, firrea_1989).

regulation_distinguishing_factor(kind_of_value).
regulation_distinguishing_factor(purpose_of_valuation).
regulation_distinguishing_factor(temporal_focus).
regulation_distinguishing_factor(output_form).
regulation_distinguishing_factor(represented_as_appraisal).
regulation_distinguishing_factor(federally_related_transaction_involvement).

regulation_triggers_licensure(federally_related_mortgage_transaction).
regulation_triggers_licensure(work_represented_as_appraisal).
regulation_triggers_licensure(work_meeting_state_appraisal_definition).
regulation_triggers_licensure(regulated_lending_transaction).

regulation_definition_of_appraisal(formal_opinion_of_value).
regulation_definition_of_appraisal(concerning_identified_real_property).
regulation_definition_of_appraisal(developed_per_recognized_methods).
regulation_definition_of_appraisal(for_compensation).

regulation_unlicensed_activity(equity_analyst,      stock_price_targets).
regulation_unlicensed_activity(investment_bank,
    ↪ enterprise_value_estimates).
regulation_unlicensed_activity(hedge_fund,
    ↪ commercial_real_estate_appreciation_forecasts).
regulation_unlicensed_activity(economist,
    ↪ housing_market_trend_projections).
regulation_unlicensed_activity(insurance_analyst,
    ↪ replacement_cost_estimates).
regulation_unlicensed_activity(developer,           feasibility_studies).
regulation_unlicensed_activity(internal_corporate_analyst,
    ↪ asset_impairment_estimates).
regulation_unlicensed_activity(internal_corporate_analyst,
    ↪ portfolio_value_estimates).
regulation_unlicensed_activity(quantitative_analyst,
    ↪ predictive_valuation_models).
regulation_unlicensed_activity(avm_developer,
    ↪ algorithmic_residential_price_estimates).

regulation_licensed_activity(licensed_appraiser,
    ↪ market_value_for_federally_related_transaction).
regulation_licensed_activity(licensed_appraiser,
    ↪ formal_opinion_of_value_for_compensation).
regulation_licensed_activity(licensed_appraiser,
    ↪ appraisal_for_regulated_mortgage_lending).
regulation_licensed_activity(licensed_appraiser,
    ↪ appraisal_represented_as_such_under_state_law).

regulation_legal_exposure_condition(holding_self_out_as_licensed_appraiser).
regulation_legal_exposure_condition(preparing_appraisal_as_defined_by_state_law).
regulation_legal_exposure_condition(involved_in_transaction_requiring_licensed_app|
    ↪ raisal).

```

```

%% --- The test itself: does the TASK require an appraiser's license? ---
requires_license(Task) :-
    regulation_triggers_licensure(Task), !.
requires_license(Task) :-
    regulation_licensed_activity(_, Task), !.

exempt_from_appraisal_license(Task) :-
    regulation_unlicensed_activity(_, Task),
    \+ requires_license(Task).

:- dynamic activity_satisfies_definition/1.

activity_satisfies_definition(Task) :-
    regulation_licensed_activity(_, Task).

regulated_appraisal(Task) :-
    activity_satisfies_definition(Task),
    requires_license(Task).

unregulated_valuation(Task) :-
    regulation_unlicensed_activity(_, Task),
    \+ regulated_appraisal(Task).

%% Meta-tension: the regulatory asymmetry the passage closes on.
regulatory_asymmetry(current_market_value_for_lending,
    future_oriented_valuation_for_investment).
regulatory_asymmetry(heavily_regulated_appraisal_practice,
    lightly_regulated_quantitative_practice).
regulatory_asymmetry(consumer_and_lending_protection_focus,
    investment_advisory_focus).

regulatory_asymmetry_origin(banking_regulation).
regulatory_asymmetry_origin(consumer_protection).
regulatory_asymmetry_origin(legal_evidentiary_concerns).
regulatory_asymmetry_origin_disclaimer(
    'Not based on any exclusive claim that appraisers alone possess valuation
    ↪ expertise.').

%% =====
%% TIER 2 -- DESCRIPTIVE TENDENCIES (NOT the boundary)
%%
%% Overlapping populations; correlates of who typically lands on each
%% side. None of these determine licensure -- Tier 1 does.
%% =====

%% --- Axis: valuation approach (descriptive populations) ---
valuation_approach(appraisal).
valuation_approach(quant_financial).

scope_profession(appraiser).
scope_profession(quantitative_analyst).
scope_profession(financial_analyst).
scope_profession(investment_analyst).
scope_profession(portfolio_manager).
scope_profession(economist).
scope_profession(investment_strategist).
scope_profession(financial_forecaster).

```

```

scope_orientation(appraisal,      evidentiary).
scope_orientation(quant_financial, probabilistic).

%% --- Axis: temporal scope (its OWN viewpoint, distinct from approach) ---
temporal_scope(retrospective).
temporal_scope(current).
temporal_scope(prospective_point).
temporal_scope(future_range_horizon).

temporal_mode(point_in_time).
temporal_mode(distribution_over_horizon).

%% Typical (defeasible) pairing of approach to temporal mode -- a LINK
%% between two distinct axes, not an identity.
scope_temporal_focus(appraisal,      point_in_time).
scope_temporal_focus(quant_financial, distribution_over_horizon).

%% --- Axis: output form ---
scope_output_form(appraisal,      point_estimate).
scope_output_form(appraisal,      narrow_range).
scope_output_form(quant_financial, probability_distribution).
scope_output_form(quant_financial, range_of_outcomes).

%% --- Axis: skills / methodological toolkit (training, NOT intelligence) ---
scope_skill(mars_without_overfitting).
scope_skill(fit_vs_generalization_balance).    % balancing  $R^2$  against  $CV-R^2$ 
scope_skill(algorithmic_understanding).
scope_skill(handling_structured_concepts).
scope_skill(discerning_fine_distinctions_in_meaning).
scope_skill(assimilating_new_methods).

%% Skills typically demanded by advanced quantitative valuation; an
%% appraisal license neither requires nor certifies them.
scope_skill_demanded_by(quant_financial, mars_without_overfitting).
scope_skill_demanded_by(quant_financial, fit_vs_generalization_balance).
scope_skill_demanded_by(quant_financial, algorithmic_understanding).
scope_skill_demanded_by(quant_financial, handling_structured_concepts).
scope_skill_demanded_by(quant_financial, discerning_fine_distinctions_in_meaning).
scope_skill_demanded_by(quant_financial, assimilating_new_methods).

%% --- Inputs the appraiser weighs to support a value at the effective date ---
scope_inputs_considered(appraisal, current_market_evidence).
scope_inputs_considered(appraisal, observable_transactions).
scope_inputs_considered(appraisal, prevailing_market_conditions).
scope_inputs_considered(appraisal, recognized_methodologies).
scope_inputs_considered(appraisal, market_trends).
scope_inputs_considered(appraisal, anticipated_supply_demand).
scope_inputs_considered(appraisal, depreciation).
scope_inputs_considered(appraisal, capitalization_rates).
scope_inputs_considered(appraisal, absorption_patterns).
scope_inputs_considered(appraisal, expectations_already_priced_in).

%% --- Bounded future-oriented contexts that remain within appraisal ---
scope_future_oriented_context(income_property, future_income_streams).
scope_future_oriented_context(income_property, vacancy_rates).
scope_future_oriented_context(income_property, expenses).
scope_future_oriented_context(income_property, reversion_values).
scope_future_oriented_context(income_property, dcf_analysis).

```

```

scope_future_oriented_context(development, prospective_stabilized_value).
scope_future_oriented_context(conservation, future_damages).
scope_future_oriented_context(environmental, future_damages).
scope_future_oriented_context(insurance, future_damages).
scope_future_oriented_context(litigation, future_damages).
scope_future_oriented_context(eminent_domain, future_damages).
scope_future_oriented_context(general, remaining_economic_life).
scope_future_oriented_context(general, anticipated_market_reactions).
scope_future_oriented_context(market_analysis, price_trend_direction).

%% --- Professional constraints on any future-oriented component ---
scope_professional_constraint(supportability).
scope_professional_constraint(market_evidence).
scope_professional_constraint(extraordinary_assumptions_or_hypothetical_conditions,
↪ _disclosed).
scope_professional_constraint(avoidance_of_unsupported_speculation).

%% --- Contrasts across the descriptive axes:
↪ scope_contrast(A,B,Dimension,Distinction) ---
scope_contrast(appraisal, quant_financial, orientation,
evidentiary_vs_probabilistic).
scope_contrast(appraisal, quant_financial, temporal_mode,
point_in_time_vs_distribution_over_horizon).
scope_contrast(appraisal, quant_financial, output_form,
point_or_narrow_range_vs_probability_distribution).
scope_contrast(appraisal, quant_financial, purpose,
market_value_vs_investment_worth).
scope_contrast(appraisal, quant_financial, skill_set,
appraisal_licensing_toolkit_vs_advanced_statistical_toolkit).
scope_contrast(appraisal, quant_financial, typical_methods,
sales_comparison_cost_income_vs_simulation_option_pricing_var).

%% --- The orienting question each population asks (verbatim from the passage) ---
scope_question(appraisal,
'What is the most supportable market value indication as of the specified
↪ effective date, given current evidence and defined assumptions?').
scope_question(quant_financial,
'What range of outcomes is likely over a future period, with what
↪ probabilities, and how should resources be allocated?').

%% --- Where the two converge (appraisal borrows distributional methods) ---
scope_convergence_area(complex_commercial_valuation).
scope_convergence_area(mass_appraisal).
scope_convergence_area(automated_valuation_models).
scope_convergence_area(institutional_portfolio_analysis).

```

5 市场价值 × LTV 作为止赎价值的替代指标

前面各节阐明了估价的用途、估价的产出，以及估价的理论边界与监管边界之所在；本节则触及它们一路铺垫而来的那个张力。

抵押贷款与抵押品分析背后一个更深层的实际现实是：就实际经济功能而言，用于抵押贷款审核的市场价值估价往往充当着对未来某一持有期内预期抵押品可回收性的隐性替代指标，尽管在形式上该估价被表述为截至特定生效日期的当前市场价值意见。(Foster and Van Order 1984; Kau et al. 1995; Deng, Quigley, and Van Order 2000)

80% 的贷款价值比并非随意设定。从历史上看，它在一定程度上是作为保护贷款机构的风

险缓冲而演变形成的，用以抵御：

- 温和的市场下行、
- 止赎费用、
- 清算折价、
- 持有成本、
- 法律费用、
- 应计利息、
- 以及违约后处置抵押品所需时间内的不确定性。(Foote, Gerardi, and Willen 2008; Qi and Yang 2009)

这一缓冲也解释了为何贷款金额本身——市场价值与允许的贷款价值比之乘积——可以被解读为贷款机构对可回收止赎价值的工作估计：将贷款上限设定为市场价值的某一比例，实际上是在押注强制出售所得不会跌破贷款余额。然而，这一折减应当有多大，远比 80% 这个整数所暗示的更难确定。对止赎折价的估计因市场、方法和时期而差异巨大；一项被频繁引用的强制出售研究将其定在比可比正常交易价格低约 27% 的水平 (Campbell, Giglio, and Pathak 2011)，而其他分析——在校正房产状况、区位和样本选择之后——报告的幅度则大不相同。(Pennington-Cross 2006; Clauretie and Daneshvary 2009)

相比之下，常规的 80% 门槛本质上是一个全国统一的单一参数，被套用于千差万别的地方市场和信贷周期的先后阶段。它无法追踪任何特定市场的折价，因此只是一个非常粗略的、一刀切的近似，而非从市场价值到止赎回收额的精确映射。真要说的话，它还是一个偏薄的缓冲：20% 的折减低于上述研究报告的 27% 折价，因此即便尚未扣除前文列举的处置成本，常规贷款的最高额度就可能超过强制出售的毛收入。

在借款人权益过薄、不足以构成这一缓冲的情形下，便以抵押贷款保险取而代之。对此最好的理解不是把它当作一个政策旋钮，而是一种风险分配的权衡：有一部分借款人资产薄弱但收入稳定，向他们发放 LTV 超过 80% 的贷款所带来的增量违约与损失风险由保险人预测并承担——私人渠道是高 LTV 常规贷款上的私人抵押贷款保险，公共渠道则是联邦住房管理局 (Federal Housing Administration)。然而，使这种权衡成为常态的市场本身是联邦住房政策的产物：主导常规贷款市场的政府支持企业依其特许章程被禁止在缺乏此类信用增级的情况下购买 LTV 超过 80% 的贷款 (United States Code 1970)。保险终止的条件同样是制度选择而非风险计算的结果，并且在不同项目之间差异悬殊。¹

因此，尽管估价师在官方口径上估计的是“当前市场价值”，贷款机构在经济上却是透过一个前瞻性风险视角来解读这一价值的。实际上，贷款机构问的问题更接近于：

“如果这位借款人在可预见的未来某个时候违约，抵押品是否仍可能足以覆盖未偿债务和与止赎相关的损失？”

这在本质上是一个面向未来的问题。

这有赖于以下两者之间刻意的概念区分：

- 估价师的“当前市场价值”意见，以及
- 贷款机构在实践中将该意见用作抵御未来抵押品减值的预测缓冲。

从历史与法律上看，监管机构曾力图防止估价师变成投机性预测者，因为：

- 投机性预测曾是以往放贷危机的成因之一，
- 法院希望估价与可观察的证据挂钩，
- 而银行监管机构则寻求标准化、可辩护的估价实践。

于是，估价理论朝着以下方向演进：

- 现值、

¹根据 1998 年《房主保护法》(Homeowners Protection Act)，常规贷款中由借款人支付的私人抵押贷款保险在 LTV 降至 78% 时自动终止，在 80% 时可申请终止。(United States Congress 1998) 联邦住房管理局的保费规则更为严格：对于 2013 年 6 月 3 日及以后发放、原始贷款价值比高于 90%（例如首付 3.5%）的贷款，年度保费贯穿贷款整个存续期，而 LTV 不高于 90% 的贷款则在十一年后取消。(U.S. Department of Housing and Urban Development 2013) 因此，许多 FHA 借款人只能通过再融资转为接近 80% LTV 的常规贷款来摆脱抵押贷款保险。

- 当前市场证据、
- 明确的生效日期、
- 以及受约束的假设。

与此同时，贷款机构、投资者和抵押品分析师却在概率性的未来风险框架中例行使用这些同样的估价。

现实中，抵押贷款审核是一个综合系统，涉及：

1. 当前市场估值、
2. 未来风险估计、
3. 违约概率、
4. 止赎时点、
5. 清算损失严重度、
6. 宏观经济预期、
7. 以及组合压力分析。

估价师正式贡献的主要是第 (1) 项，而贷款机构在内部将这一结果与其余各项相结合。

这一区分在实践中变得模糊。如果贷款机构只有在相信抵押品未来大概率能保有足够价值时才肯放贷，那么估价实际上就间接地充当了未来价值保护模型的一部分——即使估价准则在正式表述上回避这样描述它。

随着以下因素的出现，这一张力愈发突出：

- 自动估价模型 (AVM)、
- 机构组合分析、
- 证券化模型、
- 机器学习估值系统、
- 巴塞尔资本建模、
- 以及大规模抵押贷款压力测试。

现代抵押品分析日益将财产价值视为随时间变化的概率量，而非单点静态估计。(Basel Committee on Banking Supervision 2017; Steurer, Hill, and Pfeifer 2021; Board of Governors of the Federal Reserve System 2021)

可以认为，传统的估价监管反映了一种历史性妥协：

- 估价师提供有纪律的、基于证据的现值之锚，
- 而金融机构承担未来风险解读的责任。

这种分工在现代金融体系中是否仍然完全自治，是一个正当且日益受到讨论的问题。

由此可以得出对估价实践本身的一个现实启示。大多数估价师并未意识到这一抵押品估值角色——贷款机构将市场价值用作前瞻性缓冲的做法在估价业务内部很难看到——因而没有以该角色所要求的、对未来价值走势的审慎态度来对待自己的市场价值结论。认识到现值意见实际上是如何被使用的，是弥合这一差距的第一步，也是本文希望为后续研究开启的领域。

```
%% appraisal_proxy.pl
%%
%% A Prolog knowledge base encoding the thesis that market value
%% appraisals, while formally framed as present-value opinions, function
%% in practice as implicit proxies for future collateral recoverability
%% within mortgage underwriting and lender risk management.

:- module(appraisal_proxy,
    [ % --- central thesis ---
      proxy_thesis/1

      % --- LTV-as-buffer mechanism ---
      , proxy_ltv_buffer_against/1
      , proxy_historical_ltv_threshold/1
```

```

    % --- formal vs practical interpretation ---
    , proxy_formal_framing/1
    , proxy_practical_interpretation/1
    , proxy_underlying_question/1

    % --- the artificial separation ---
    , proxy_separation_party/2          % Party, Role
    , proxy_doctrine_practice_gap/2    % FormalSide, PracticalSide

    % --- historical compromise ---
    , proxy_historical_reason/1
    , proxy_doctrine_emphasis/1
    , proxy_routine_practical_use/1

    % --- modern intensifiers ---
    , proxy_modern_driver/1
    , proxy_modern_treatment/1

    % --- responsibility allocation ---
    , proxy_appraiser_responsibility/1
    , proxy_lender_responsibility/1

    % --- meta-claim about coherence ---
    , proxy_open_question/1

    % --- practical implication for appraisal practice ---
    , proxy_practical_implication/1
    , proxy_appraiser_blindspot/1
    , proxy_recommended_practice/1
    , proxy_further_work/1

    % --- underwriting system view ---
    , underwriting_component/2          % Number, Component
    , underwriting_responsibility/2    % Component, Party
  ]).

:- discontiguous proxy_thesis/1.
:- discontiguous underwriting_component/2.

%% =====
%% PART 1: THE CENTRAL THESIS
%% =====

proxy_thesis(
  'A market value appraisal used in mortgage underwriting often serves as an
  ↪ implicit proxy for expected collateral recoverability over a future holding
  ↪ period.').
proxy_thesis(
  'The appraisal is formally framed as a present-market-value opinion as of a
  ↪ specific effective date.').
proxy_thesis(
  'The lender economically interprets the appraisal through a forward-looking
  ↪ risk lens.').
proxy_thesis(
  'There is a deliberate conceptual separation between the appraiser''s
  ↪ current-market-value opinion and the lender''s use of it as a forecast
  ↪ buffer.').

```

```

proxy_thesis(
    'The appraisal indirectly functions as part of a future-value protection model,
    ↪ even though appraisal standards formally avoid describing it that way.').
proxy_thesis(
    'Whether the separation between appraisal doctrine and lender practice remains
    ↪ fully coherent in modern financial systems is an open and increasingly
    ↪ debated question.').

%% =====
%% PART 2: THE LTV-AS-BUFFER MECHANISM
%% =====

proxy_historical_ltv_threshold(eighty_percent).

proxy_ltv_buffer_against(moderate_market_declines).
proxy_ltv_buffer_against(foreclosure_expenses).
proxy_ltv_buffer_against(liquidation_discounts).
proxy_ltv_buffer_against(carrying_costs).
proxy_ltv_buffer_against(legal_expenses).
proxy_ltv_buffer_against(accrued_interest).
proxy_ltv_buffer_against(uncertainty_during_disposition).

%% =====
%% PART 3: FORMAL FRAMING VS. PRACTICAL INTERPRETATION
%% =====

proxy_formal_framing(present_market_value_opinion).
proxy_formal_framing(opinion_as_of_specific_effective_date).
proxy_formal_framing(grounded_in_current_market_evidence).

proxy_practical_interpretation(forward_looking_risk_lens).
proxy_practical_interpretation(forecast_buffer_against_future_impairment).
proxy_practical_interpretation(input_to_future_collateral_recoverability_estimate).

proxy_underlying_question(
    'If this borrower defaults sometime in the foreseeable future, is the
    ↪ collateral likely to remain sufficient to cover the outstanding debt and
    ↪ foreclosure-related losses?').

%% =====
%% PART 4: THE ARTIFICIAL SEPARATION
%% =====

proxy_separation_party(appraiser, opinion_of_current_market_value).
proxy_separation_party(lender, forecast_buffer_against_future_impairment).
proxy_separation_party(appraiser, evidence_based_present_value_anchor).
proxy_separation_party(lender, future_risk_interpretation).

proxy_doctrine_practice_gap(present_market_value_opinion,
                             implicit_future_value_proxy).
proxy_doctrine_practice_gap(opinion_as_of_effective_date,
                             forecast_over_foreseeable_holding_period).
proxy_doctrine_practice_gap(evidence_based_present_anchor,
                             future_risk_buffer).
proxy_doctrine_practice_gap(static_single_point_estimate,
                             probabilistic_value_over_time).

%% =====

```

```

%% PART 5: HISTORICAL COMPROMISE
%% =====

proxy_historical_reason(speculative_forecasting_contributed_to_past_crises).
proxy_historical_reason(courts_required_observable_evidence).
proxy_historical_reason(regulators_sought_standardized_defensible_practices).

proxy_doctrine_emphasis(present_value).
proxy_doctrine_emphasis(current_market_evidence).
proxy_doctrine_emphasis(defined_effective_dates).
proxy_doctrine_emphasis(constrained_assumptions).

proxy_routine_practical_use(lenders_use_in_probabilistic_future_risk_frameworks).
proxy_routine_practical_use(investors_use_in_probabilistic_future_risk_frameworks).
proxy_routine_practical_use(collateral_analysts_use_in_probabilistic_future_risk_f
↪ rameworks).

%% =====
%% PART 6: MODERN INTENSIFIERS
%% =====

proxy_modern_driver(avms).
proxy_modern_driver(institutional_portfolio_analytics).
proxy_modern_driver(securitization_models).
proxy_modern_driver(machine_learning_valuation_systems).
proxy_modern_driver(basel_capital_modeling).
proxy_modern_driver(large_scale_mortgage_stress_testing).

proxy_modern_treatment(probabilistic_over_time).
proxy_modern_treatment(distributional_rather_than_point_estimate).
proxy_modern_treatment(dynamic_rather_than_static).

%% =====
%% PART 7: ALLOCATION OF RESPONSIBILITY UNDER THE COMPROMISE
%% =====

proxy_appraiser_responsibility(disciplined_present_value_anchor).
proxy_appraiser_responsibility(evidence_based_valuation).
proxy_appraiser_responsibility(current_market_value_opinion).

proxy_lender_responsibility(future_risk_interpretation).
proxy_lender_responsibility(default_probability_estimation).
proxy_lender_responsibility(foreclosure_timing_estimation).
proxy_lender_responsibility(liquidation_severity_estimation).
proxy_lender_responsibility(portfolio_stress_analysis).
proxy_lender_responsibility(macroeconomic_expectation_integration).

%% =====
%% PART 8: OPEN QUESTION (META-CLAIM)
%% =====

proxy_open_question(
    'Whether the separation between disciplined present-value appraisal and
    ↪ lender-side future-risk interpretation remains fully coherent in modern
    ↪ financial systems.').
proxy_open_question(
    'Whether modern probabilistic-over-time collateral analytics make the
    ↪ formal/practical separation untenable in some sectors.').

```

```

%% =====
%% PART 9: THE INTEGRATED MORTGAGE UNDERWRITING SYSTEM
%% =====

underwriting_component(1, current_market_valuation).
underwriting_component(2, future_risk_estimation).
underwriting_component(3, default_probability).
underwriting_component(4, foreclosure_timing).
underwriting_component(5, liquidation_severity).
underwriting_component(6, macroeconomic_expectations).
underwriting_component(7, portfolio_stress_analysis).

underwriting_responsibility(current_market_valuation, appraiser).
underwriting_responsibility(future_risk_estimation, lender).
underwriting_responsibility(default_probability, lender).
underwriting_responsibility(foreclosure_timing, lender).
underwriting_responsibility(liquidation_severity, lender).
underwriting_responsibility(macroeconomic_expectations, lender).
underwriting_responsibility(portfolio_stress_analysis, lender).

%% =====
%% PART 10: PRACTICAL IMPLICATION FOR APPRAISAL PRACTICE
%% =====

proxy_practical_implication(
    'Most appraisers are not aware that market value functions, via the
    ↪ loan-to-value ratio, as a proxy for future foreclosure value.').
proxy_practical_implication(
    'The lender''s forward-looking use of market value is rarely visible from
    ↪ within the appraisal assignment.').
proxy_practical_implication(
    'Appraisers therefore do not treat market value conclusions with the caution
    ↪ toward future value trends that the collateral-valuation role warrants.').

proxy_appraiser_blindspot(collateral_valuation_role).
proxy_appraiser_blindspot(forward_looking_use_of_market_value).
proxy_appraiser_blindspot(market_value_as_ltv_proxy_for_foreclosure_value).

proxy_recommended_practice(recognize_how_present_value_opinion_is_consumed).
proxy_recommended_practice(treat_market_value_with_caution_re_future_trends).

proxy_further_work(closing_the_doctrine_practice_gap).
proxy_further_work(equipping_appraisers_for_the_collateral_valuation_perspective).

```

6 结论

前述知识库所编码的五段论述，从五个互补的视角描述了同一个职业。孤立地看，每段论述都是一份自成一体的阐述：它们所承载的社会与公共利益关切、估价所支持的经济功能、其在抵押品风险分析中的操作角色、将其与相邻量化工作分隔开来的监管边界，以及其正式表述与实际用途之间的张力。合在一起，并借助 Prolog 将结构显性化之后，它们描述的是某种更有意味的东西——一个其官方理论与经济功能被刻意分置的职业，其内在自治性正日益受到如今施加于其产出之上的分析工具的考验。

前三个知识库共享一个共同模式：effect(Quality, Horizon, Domain, Effect)。每一个都将估价质量框定作为一种因果输入，其后果在短期、中期与长期跨度上传播，区别仅在于解读

这些后果所经由的领域。经济视角通过信贷市场、投资配置、税收和资本形成来解读；社会视角通过公平、公众信心和社区结果来解读；抵押品视角通过审贷、组合管理和金融体系稳定来解读。共享模式并非巧合。它反映了贯穿三段论述的一个实质性主张：估值工作的质量是有后果的，这些后果随时间放大，而同样的底层失效模式会因所关注利益主体的不同而以不同方式显现。谓词参数侧的共享原子——尤其是质量与时间跨度词汇——使三套词汇表可以互联，从而像系统性高估这样的单一失效能够同时经由经济上的错误定价、社会上的不公与抵押品上的错误计量得到追踪。

第四个知识库转换了论域。它描述的不是估价质量的后果，而是估价在方法论和法律意义上是什么。其谓词是分类性的而非因果性的，其推理机制回答的是管辖问题而非影响问题。这一转换是必要的，因为前三段论述的后果式框架预设了对被评价活动的定义，而这一定义本身是有争议的。scope 谓词记录估价师做什么——基于证据的现值意见，辅以边界狭窄的前瞻性成分——而 regulation 谓词记录该活动落入持照执业范围的条件。前向链推理规则(requires_license/1、exempt_from_appraisal_license/1、regulated_appraisal/1)提供了一种机制，用以询问某项估值工作落在受监管边界之内还是之外——这个问题是前三个知识库默认已经得到回答的。

第五个知识库提出了一个论题。其他知识库或描述、或分类、或追踪后果，而 proxy 知识库则论证：现值估价与未来价值预测之间审慎的理论分隔——即 regulation 知识库所记录的那种分隔——与估价在抵押贷款中的实际经济功能存在张力。80% 的贷款价值比惯例、它所缓冲的七类具体风险、逐字记录的贷款机构的底层问题（“如果这位借款人在可预见的未来某个时候违约，抵押品是否仍可能足以...”），以及由七个组成部分构成的综合审贷系统，都支持同一个主张：估价在形式上是现值意见，在经济上是未来价值缓冲。促成这一安排的历史性妥协——估价师充当有纪律的现值之锚，贷款机构充当未来风险解读的中枢——在估价执照制度形成的历史条件下是一种可以辩护的应对，但文中列举的现代强化因素（AVM、机器学习估值系统、巴塞尔资本建模、证券化分析、大规模压力测试）将财产价值作随时间变化的概率化处理。这些工具把在理论上被分隔的层次重新拉回同一条计算管道，层与层之间的概念分隔也因此愈发难以维系。

这种调和是否可以实现、以何种条件实现，正是 proxy 段落收束时提出的问题。这些知识库并未回答它。但它们确实为回答该问题提供了一份明确、可查询的相关考量清单：估值质量在各时间跨度与领域中的后果、决定哪些后果落入受监管实践的管辖边界，以及该边界如今承压的结构性原因。这份阐述是进一步构建更具体论证——关于方法论、关于准则改革、关于估价实践与更广泛的量化估值体系之间关系——的基础。

关于知识表示的说明

有三个结构性模式值得明确记录，因为它们产生于编码过程本身，而非源自任何单独一段论述。

第一，这一知识库家族并不是扁平的。五个知识库中有三个共享同一模式：一个提供决定该模式何时适用的管辖性限定；一个就正式表述与经济功能之间的差距提出论题。将五者结合的桥接层并不把它们当作同一矩阵的五个列。它把前三者视为对后果的描述，把第四个视为对这些描述在何处有效的约束，把第五个视为对该约束本身承压之处的识别。这种非对称结构是特性而非缺陷：它映照了职业实践的实际组织方式——在其中，描述性主张、管辖性主张与批判性主张占据着确实不同的认识论角色。

第二，区分各模块的前缀纪律(economic_、social_、collateral_、scope_/regulation_、proxy_/underwriting_)只有在存在冲突风险之处才真正发挥作用。在各模块间有直接对应物的谓词才配得上前缀；某一概念领域独有的谓词则不需要。参数侧的共享取值词汇——质量、时间跨度和领域原子——恰恰是各模块得以互通的关键，给这些取值加前缀反而会破坏这一点。由此浮现的模式具有普遍性：谓词加前缀，参数共享原子，并且在谓词为单一视角独有时采用选择性前缀而非一律加前缀。

第三，proxy 知识库所识别的核心张力与 scope/regulation 知识库所识别的监管不对称密切相关，却又彼此不同。两者通过平行的二元谓词——regulatory_asymmetry/2 与 proxy_doctrine_practice_gap/2——加以记录，并刻意采用不同的名称，因为它们描述的是同一底层安排的不同轴线。监管不对称关乎谁受监管、监管程度多深；理论与实践差距关乎对估价师产品的正式表述与其实际经济用途之间的落差。两种张力相互强化。监管把估价师推

向现值理论；经济功能把他们的产出拉向未来价值用途。合起来的图景是：这一职业的官方自我描述，与其产品所汇入的分析管道之间，正变得越来越难以调和。

参考文献

- Agarwal, Sumit, Itzhak Ben-David, and Vincent Yao (2015). "Collateral Valuation and Borrower Financial Constraints: Evidence from the Residential Real Estate Market". In: *Management Science* 61.9, pp. 2220–2240. DOI: [10.1287/mnsc.2014.2002](https://doi.org/10.1287/mnsc.2014.2002).
- Alexandrov, Alexei, Laurie Goodman, and Michael Neal (Jan. 2023). *Reengineering the Appraisal Process: Better Leveraging Both Automated Valuation Models and Manual Appraisals*. Tech. rep. Urban Institute. URL: <https://www.urban.org/sites/default/files/2023-01/Reengineering%20the%20Appraisal%20Process.pdf>.
- Basel Committee on Banking Supervision (2017). *Basel III: Finalising Post-Crisis Reforms*. Standards d424. Bank for International Settlements. URL: <https://www.bis.org/bcbs/publ/d424.pdf>.
- Board of Governors of the Federal Reserve System (Mar. 2021). *Dodd-Frank Act Stress Test 2021: Supervisory Stress Test Methodology*. Tech. rep. Board of Governors of the Federal Reserve System. URL: <https://www.federalreserve.gov/publications/files/2021-april-supervisory-stress-test-methodology.pdf>.
- Campbell, John Y., Stefano Giglio, and Parag Pathak (2011). "Forced Sales and House Prices". In: *American Economic Review* 101.5, pp. 2108–2131. DOI: [10.1257/aer.101.5.2108](https://doi.org/10.1257/aer.101.5.2108).
- Clauretie, Terrence M. and Nasser Daneshvary (2009). "Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time". In: *Real Estate Economics* 37.1, pp. 43–67. DOI: [10.1111/j.1540-6229.2009.00235.x](https://doi.org/10.1111/j.1540-6229.2009.00235.x).
- Crosby, Neil, Colin M. Lizieri, and Patrick M. McAllister (2010). "Means, Motive and Opportunity? Disentangling Client Influence on Performance Measurement Appraisals". In: *Journal of Property Research* 27.2, pp. 181–201. DOI: [10.1080/09599916.2010.499014](https://doi.org/10.1080/09599916.2010.499014).
- Deng, Yongheng, John M. Quigley, and Robert Van Order (2000). "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options". In: *Econometrica* 68.2, pp. 275–307. DOI: [10.1111/1468-0262.00110](https://doi.org/10.1111/1468-0262.00110).
- Diaz Julian, III and Marvin L. Wolverton (1998). "A Longitudinal Examination of the Appraisal Smoothing Hypothesis". In: *Real Estate Economics* 26.2, pp. 349–358. DOI: [10.1111/1540-6229.00749](https://doi.org/10.1111/1540-6229.00749).
- Ebrahim, M. Shahid and Sikandar Hussain (2010). "Financial Development and Asset Valuation: The Special Case of Real Estate". In: *Journal of Banking & Finance* 34.1, pp. 150–162. DOI: [10.1016/j.jbankfin.2009.07.011](https://doi.org/10.1016/j.jbankfin.2009.07.011).
- Eriksen, Michael D. et al. (2019). "The Influence of Contract Prices and Relationships on Appraisal Bias". In: *Journal of Urban Economics* 111, pp. 132–143. DOI: [10.1016/j.jue.2019.03.001](https://doi.org/10.1016/j.jue.2019.03.001).
- Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen (2008). "Negative Equity and Foreclosure: Theory and Evidence". In: *Journal of Urban Economics* 64.2, pp. 234–245. DOI: [10.1016/j.jue.2008.07.006](https://doi.org/10.1016/j.jue.2008.07.006).
- Foster, Chester and Robert Van Order (1984). "An Option-Based Model of Mortgage Default". In: *Housing Finance Review* 3.4, pp. 351–372.
- Friedman, Jerome H. (1991). "Multivariate Adaptive Regression Splines". In: *The Annals of Statistics* 19.1, pp. 1–67. DOI: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Gallimore, Paul (1996). "Confirmation Bias in the Valuation Process: A Test for Corroborating Evidence". In: *Journal of Property Research* 13.4, pp. 261–273. DOI: [10.1080/095999196368781](https://doi.org/10.1080/095999196368781).
- Geltner, David M. (1991). "Smoothing in Appraisal-Based Returns". In: *The Journal of Real Estate Finance and Economics* 4.3, pp. 327–345. DOI: [10.1007/BF00161933](https://doi.org/10.1007/BF00161933).

- Ghent, Andra C., Walter N. Torous, and Rossen I. Valkanov (2019). “Commercial Real Estate as an Asset Class”. In: *Annual Review of Financial Economics* 11, pp. 153–171. DOI: [10.1146/annurev-financial-110118-123121](https://doi.org/10.1146/annurev-financial-110118-123121).
- Griffin, John M. and Gonzalo Maturana (2016). “Who Facilitated Misreporting in Securitized Loans?” In: *The Review of Financial Studies* 29.2. Evidence of appraisal inflation and its consequences, pp. 384–419. DOI: [10.1093/rfs/hhv130](https://doi.org/10.1093/rfs/hhv130).
- Hendershott, Patric H. and Edward J. Kane (July 1992). *Office Market Values During the Past Decade: How Distorted Have Appraisals Been?* NBER Working Paper 4128. Available at SSRN: <https://ssrn.com/abstract=226786>. National Bureau of Economic Research. DOI: [10.3386/w4128](https://doi.org/10.3386/w4128). URL: <https://www.nber.org/papers/w4128>.
- (1995). “U.S. Office Market Values During the Past Decade: How Distorted Have Appraisals Been?” In: *Real Estate Economics* 23.2, pp. 101–116. DOI: [10.1111/1540-6229.00660](https://doi.org/10.1111/1540-6229.00660).
- Hull, John C. (2018). *Options, Futures, and Other Derivatives*. 10th ed. New York: Pearson. ISBN: 978-0-13-447208-9.
- International Association of Assessing Officers (2017). *Standard on Mass Appraisal of Real Property*. International Association of Assessing Officers, Kansas City, MO. URL: https://www.iaao.org/wp-content/uploads/Standard_on_Mass_Appraisal.pdf.
- International Valuation Standards Council (2025). *International Valuation Standards (IVS)*. International Valuation Standards Council, London. Published 31 January 2024; effective 31 January 2025. URL: <https://www.ivsc.org/>.
- K’Akumu, Owiti A. and James E. Larsen (2024). “A Primer on Regulations and the Practice of Residential Property Appraisal”. In: *Journal of Real Estate Practice and Education*. Focus on regulatory and public interest aspects. DOI: [10.1080/15214842.2024.2377869](https://doi.org/10.1080/15214842.2024.2377869).
- Kau, James B. et al. (1995). “The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment”. In: *The Journal of Real Estate Finance and Economics* 11.1, pp. 5–36. DOI: [10.1007/BF01097934](https://doi.org/10.1007/BF01097934).
- LaCour-Little, Michael and Stephen Malpezzi (2003). “Appraisal Quality and Residential Mortgage Default: Evidence from Alaska”. In: *The Journal of Real Estate Finance and Economics* 27.2, pp. 211–233. DOI: [10.1023/A:1024728420837](https://doi.org/10.1023/A:1024728420837).
- Lee, Ming-Hsiang, Chien-Wen Peng, and Hui-Fen Liao (2020). “An Analysis of Objectivity in the Real Estate Appraisal Process”. In: *International Real Estate Review* 23.4, pp. 483–504.
- Mayer, Yanling G. and Frank E. Nothaft (2022). “Appraisal Overvaluation: Evidence of Price Adjustment Bias in Sales Comparisons”. In: *Real Estate Economics* 50.3, pp. 862–881. DOI: [10.1111/1540-6229.12351](https://doi.org/10.1111/1540-6229.12351).
- Murphy, Edward V. (2012). *Regulation of Real Estate Appraisers*. CRS Report RS22953. Congressional Research Service. URL: <https://sgp.fas.org/crs/misc/RS22953.pdf>.
- Office of the Comptroller of the Currency and Board of Governors of the Federal Reserve System and Federal Deposit Insurance Corporation and National Credit Union Administration and Consumer Financial Protection Bureau and Federal Housing Finance Agency (2024). *Quality Control Standards for Automated Valuation Models*. Final Rule, 89 Fed. Reg. 64538 (Aug. 7, 2024). URL: <https://www.federalregister.gov/documents/2024/08/07/2024-16197/quality-control-standards-for-automated-valuation-models>.
- Pennington-Cross, Anthony (2006). “The Value of Foreclosed Property”. In: *Journal of Real Estate Research* 28.2, pp. 193–214. DOI: [10.1080/10835547.2006.12091177](https://doi.org/10.1080/10835547.2006.12091177).
- Qi, Min and Xiaolong Yang (2009). “Loss Given Default of High Loan-to-Value Residential Mortgages”. In: *Journal of Banking & Finance* 33.5, pp. 788–799. DOI: [10.1016/j.jbankfin.2008.09.010](https://doi.org/10.1016/j.jbankfin.2008.09.010).
- Quan, Daniel C. and John M. Quigley (1991). “Price Formation and the Appraisal Function in Real Estate Markets”. In: *The Journal of Real Estate Finance and Economics* 4.2, pp. 127–146. DOI: [10.1007/BF00173120](https://doi.org/10.1007/BF00173120).

- Steurer, Miriam, Robert J. Hill, and Norbert Pfeifer (2021). “Metrics for Evaluating the Performance of Machine Learning Based Automated Valuation Models”. In: *Journal of Property Research* 38.2, pp. 99–129. DOI: [10.1080/09599916.2020.1858937](https://doi.org/10.1080/09599916.2020.1858937).
- The Appraisal Foundation (2024). *Uniform Standards of Professional Appraisal Practice (USPAP), 2024 Edition*. The Appraisal Foundation, Appraisal Standards Board, Washington, DC. Effective January 1, 2024. URL: <https://www.appraisalfoundation.org/>.
- U.S. Department of Housing and Urban Development (2013). *Revision of Federal Housing Administration Policies Concerning Cancellation of the Annual Mortgage Insurance Premium*. Mortgagee Letter 2013-04, Federal Housing Administration. Effective June 3, 2013; the annual MIP continues for the life of loans with original LTV above 90%, and for 11 years at or below 90%. URL: <https://www.hud.gov/sites/documents/13-04ml.pdf>.
- United States Code (1970). *Federal National Mortgage Association and Federal Home Loan Mortgage Corporation Charter Acts*. 12 U.S.C. § 1717(b)(2) (Fannie Mae) and 12 U.S.C. § 1454(a)(2) (Freddie Mac). Prohibit GSE purchase of conventional mortgages exceeding 80% LTV absent credit enhancement (private mortgage insurance, seller recourse/repurchase, or a retained seller participation).
- United States Congress (1989). *Financial Institutions Reform, Recovery, and Enforcement Act of 1989*. Pub. L. No. 101-73, 103 Stat. 183; Title XI codified at 12 U.S.C. §§ 3331–3356. URL: <https://www.govinfo.gov/content/pkg/STATUTE-103/pdf/STATUTE-103-Pg183.pdf>.
- (1998). *Homeowners Protection Act of 1998*. Pub. L. No. 105-216, 112 Stat. 897; codified at 12 U.S.C. §§ 4901–4910. Automatic termination of borrower-paid private mortgage insurance at 78% LTV; cancellation may be requested at 80%.

残差约束法 (RCA): 框架与规程 (简体中文版)

本文为 2025 年 3 月 5 日最初发表于 Zenodo 的论文之更新版的简体中文译本。译文借助人工智能辅助完成并经编辑部审校；英文版为权威版本。

William Bert Craytor, SRA

第 1 卷 第 1 期 · 2026 年夏季

发表于 2026 年 7 月 1 日

DOI: [10.67330/kgebf46](https://doi.org/10.67330/kgebf46)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

本开放获取文章依据[知识共享署名 4.0 国际许可协议 \(CC BY 4.0\)](#) 的条款发布。

摘要。传统估价方法的一个重大局限在于其依赖过时的手工技术，例如配对比较分析 (matched pair analysis)。这类方法对房地产特征的价值贡献视而不见：传统的销售网格 (sales grid) 没有价值贡献一栏，只有调整一栏。而事实上，价值贡献正是构建数学约束的必要基础——这些约束能够防止传统估价师的高估与低估。

残差约束法 (Residual Constraint Approach, RCA) 通过整合一个多阶段估价流程来推进传统的市场比较法 (Sales Comparison Approach, SCA)：该流程运用多元自适应回归样条 (Multivariate Adaptive Regression Splines, MARS)，并对特征价值贡献施加严格的数学约束，尤其针对未被测量的特征——诸如美学、设计、状况与质量等潜在变量——其价值在传统估价中通常受制于估价师的主观判断与偏差。

RCA 中的数学约束依赖于对 MARS 回归的专业运用，以估计可测量房地产特征对成交价格的价值贡献。然而，这一估计通常只能解释实际成交价格的约 80%，不包括设计、功能效用、状况和质量等主观评估的特征。在旧金山湾区 (San Francisco Bay Area)，代表这些主观成分的残差平均约占房地产总价值的 20%。80/20 这一数字反映的是作者在旧金山湾区典型数据条件下的经验，不应被解读为跨市场不变的常数。在可实现 R^2 较低的市场中，本框架的应对之道是识别并测量未被测量的价值驱动因素——例如沿海社区中与海洋的距离、山坡社区中的海拔高度，以及当 MLS 数据未提供时从外部来源收集的类似变量——而不是接受这一局限并报告更宽的误差区间。虽然在其他情境下 MARS 残差通常被视为估计误差，但在房地产估价中，残差是潜在变量价值的间接度量。尽管这初看似似乎不切实际，但已有证明表明：只要各描述性成分的贡献之和等于残差，残差就可以被有意义地分解为这些成分——且不影响最终的估价结果。

这一方法论源自作者二十年的实证应用经验，通过在 R 与 Python 环境中对 MARS 的迭代实现逐步发展而成。

关键词：房地产估价，住宅估价，残差约束法，MARS，多元自适应回归样条，市场比较法，残差分析，潜在变量，自动估价

目录

1 引言	3
2 背景	5
3 估价工程师	7
4 假设	8

5 RCA 工作流程	10
5.1 MARS 分析的类型	10
5.2 处理工作流程	10
5.3 数据工作流程	11
5.3.1 MARS 的数据要求	11
5.3.2 规程：已成交交易挂牌信息	12
5.4 项目文件夹结构	14
5.4.1 规程：多次 MARS (earth) 运行	15
5.5 核心 Excel 工作簿	16
5.6 备份	16
6 工作簿记号	17
6.1 描述维度	17
6.1.1 项目维度 (“a”)	18
6.1.2 运行版本维度 (“r”)	18
6.1.3 工作流阶段 (“w”)	19
6.1.4 Excel 工作簿工作表 (“s”)	19
6.1.5 房产挂牌信息 (“p”) ——MLSData 行	20
6.1.6 MLS 变量 (“v”) ——MLSData 列	20
6.1.7 最终记号示例	20
6.1.8 理由	21
6.1.9 矩阵 ²	21
6.1.10 第一次简化	21
6.1.11 第二次简化	22
7 RCA 销售网格	23
8 变量	26
8.1 变量来源本体	26
8.2 变量应用本体	27
8.3 二阶与三阶项	28
8.3.1 调整变量	28
8.3.2 汇总变量	29
8.3.3 特殊变量 & 记号	29
8.3.4 分解残差	30
9 RCA 核心证明	31

10 MARS 模型	32
10.1 模型变量与交互作用的“ g ”函数	32
10.2 g 函数的索引	32
10.3 绘图维度	33
10.4 最终模型方程	33
10.5 绘制一个模型所需的图形数量	34
11 数据特征	36
11.1 残差的用途	36
11.2 作为近似回归变量的状况与质量评级	36
11.3 数据错误	37
11.4 什么是准确性?	38
11.5 作为异常值的估价对象	39
11.6 CVR2 在应对意外数据方面的重要性	39
12 残差	41
12.1 排序与评分	41
12.2 CQA-残差曲线	43
12.3 CQA 曲线的特征	43
12.3.1 CQA 曲线各组成部分的分析	43
12.3.2 市场动态与社会含义	44
12.4 估价对象残差	44
12.5 RCA 价值结论	45
12.5.1 一般考虑	45
12.5.2 复核、审计与投诉	45
13 结论	46

1 引言

本文阐述了一种现代方法的逻辑与数学基础,该方法旨在准确估计住宅及其他房地产的市场价值(market value)。这一方法源自 20 年来将多元自适应回归样条(Multivariate Adaptive Regression Splines, MARS)软件应用于北加州住宅——尤其是旧金山湾区各社区——的实践经验,在估价精度上实现了显著进步。MARS 能够娴熟地捕捉房产特征与价值之间的非线性关系,同时高效地识别并优先处理影响最大的变量,从而提升估价质量。这一能力使估价师(appraiser)能够以远快于传统方法的速度发现新市场区域中的价值动态。MARS 构成了本文提出的残差约束法(Residual Constraint Approach, RCA)的骨干。RCA 方法之所以得名,源于其严谨的框架:

- 它对已测量和未测量房产特征的价值贡献(value contribution)均施加严格约束:

1. 所有由 MARS 导出的价值贡献, 包括残差 (residual) 在内, 其总和必须等于每个可比交易实例 (comparable) 的净售价。
 2. 残差可以分解为具有指定价值的子成分, 前提是这些子成分的总和等于该可比交易实例的残差。
- RCA 通过将主观特征——如美学吸引力、状况和质量——的总价值与残差挂钩, 最大限度地减少了估计此类特征时的偏差; 当由胜任的估价工程师 (valuation engineer) 构建 MARS 模型时, 残差能够被准确量化。

尽管 RCA 主要为住宅房地产设计——在此类物业中主观特征对市场价值影响显著——但它在更广泛的资产估价应用中亦具潜力。鉴于美国住宅房地产市场的估计价值约达 \$45 万亿, 准确的估价至关重要。Fannie Mae 指出, 典型估价与实际售价的偏差约为 5%, 但这一数字掩盖了偏差问题, 因为估价师往往面临与售价保持一致的压力。此外, 由于 60%–70% 的估价用于再融资——缺乏可比售价——RCA 所能提供的精度之必要性便显而易见。美联储 2012 年的分析估计, 2008 年危机前有 15%–20% 的估价与售价的偏差超过 $\pm 10\%$, 这一范围如今很可能仍适用于再融资估价, 并因物业类型和市场而异。准确性为何重要? 对于低贷款价值比 (LTV) 的交易——例如信用良好的买家在 \$1 million 的住宅上支付 \$500,000 首付——贷款机构可能免除估价, 认为止赎风险极小。然而, 跳过估价的买家有多付的风险, 往往在购买之后才发现缺陷。这凸显了 RCA 这类稳健估价方法的价值。量化估价的准确性仍然颇具挑战。研究者缺乏针对“市场价值”偏差的直接度量指标, 只能依赖间接代理。传统估价方法根植于主观判断, 且受估价师胜任程度参差不齐的影响, 使分析更加复杂。相关策略包括:

1. Tzioumis (2016) 研究了 11 个州 4,782 名估价师的 259,436 份估价报告, 发现三分之一的估价与合同价的偏差在 1% 以内, 一半高出合同价 1%–3%, 少数平均高出 5%——表明存在向上偏差。该研究识别出两类估价师: “按市场价值”型 (接近合同价) 和“高于市场价值”型 (持续偏高)。
2. Krivorotov and LaCour-Little (2020) 发现 90% 的购房估价达到或超过合同价, 凸显了“锚定偏差”——向某一期望价值靠拢的倾向——在再融资情境中尤为明显。

这些研究揭示了估价实践中的偏差, 而传统工具的局限性又放大了这种偏差。如果估价师普遍胜任且无偏, 他们的估价结果应当趋同; 某些群体持续的高估 (overvaluation) 表明存在系统性问题。

上述偏差研究及其指向的系统性模式, 反映了传统市场比较法 (Sales Comparison Approach, SCA) 实践的一个更深层特征, 而这正是 RCA 旨在解决的问题。传统方法在两个认识论地位截然不同的渠道上运作。在可测量特征渠道上——GLA、房间数、地块面积、房龄、位置——估价师被期望通过配对分析或对可比交易实例的局部线性回归, 客观地推导调整 (adjustment) 值。胜任的估价师会做这项工作, 由此得到的调整至少大体上能在数据面前站得住脚。但即便在这一渠道上, 传统的网格也只记录调整; 它甚至没有一列用于记录价值贡献。传统的调整也并非作为价值贡献之差来计算——它们是通过配对或类似程序直接估计出来的, 底层的价值贡献从未被计算过。相比之下, 在 RCA 中, 每一项调整都是估价对象 (subject property) 与可比交易实例在给定特征上的价值贡献之差——证明 II 正是精确刻画了这一结构的后果。在主观特征渠道上——状况、施工质量、设计、功能效用、美学——可用的支持程序从单薄到近乎空洞不等。主观特征调整可以通过对直接可比集合之外物业的配对分析、诉诸公认的同行惯例, 或引用估价师在该市场中有记录的经验来支持。这三者在形式上都符合专业标准。但没有一个能达到可测量渠道常规实现的统计严谨性。

其结果是, 标准的要求与方法所能提供的支持之间存在程序上的不对称。标准要求对每一项调整——无论可测量还是主观——都提供支持。而方法在一个渠道上提供稳健的支持, 在另一个渠道上则只有一系列较弱的程序, 从样本量为一的轶事式配对, 到引用其他估价师的习惯做法, 不一而足。诚实而胜任的估价师对景观调整进行细致的配对分析, 其提供的支持显著优于毫无依据地援引同行惯例者; 但在标准表格上比较这两份估价报告的审查者却难以区分二者, 因为

表格记录的是调整值，而非其推导的严谨程度。方法之所以容忍这种不对称，是因为对于主观特征，一直没有更强的程序可供执业估价师普遍使用。

RCA 对这两个渠道都有改进，但方式在范畴上截然不同。在可测量渠道上，MARS 回归能够捕捉配对分析和普通线性回归无法表达的非线性关系、变量交互作用和基函数（basis function）结构。MARS 模型是针对市场区域内全部合格销售数据（通常为数百宗）拟合的，而非传统网格中的三到九个可比交易实例，因此由 MARS 导出的价值贡献既客观，又比其所取代的方法准确得多。在主观渠道上，RCA 所做的事情与任何传统支持程序在性质上都不相同。它根本不试图为单个主观特征估计调整值。所有主观特征——潜在变量（latent variable）——的贡献共同表现为残差：即每个可比交易实例的实际净售价与其 MARS 预测价之差，可直接由数据计算，无需估价师输入。将该残差分解为具名成分——设计、功能效用、景观、状况、质量等——受到其总和必须等于残差总额的约束，并且根据证明 II（第 9 节），任何满足该约束的分解都会得出相同的价值结论。工作流程中剩下的判断，是将估价对象放置在可比交易实例残差排名分布中的位置，这一判断是对照可检查的排名做出的，并在报告中明确记录。一份薄弱的 RCA 估价报告仍然可能出现——缺乏经验的 VE 可能构建出过拟合的 MARS 模型、对估价对象做出不合理的定位，或忽视数据中的异常——但其薄弱之处是审查者通过检查模型和排名即可识别的，而不是标准表格审查所无法揭示的那种薄弱。

除再融资之外，估价师对业主负有提供准确估价的义务，尤其是在高价值购置中，错误可能造成数万乃至数十万的损失。然而，许多人不愿在估价上花费 \$600-\$1,000，怀疑其附加价值。RCA 这类先进方法虽然初期资源投入较大，却有望带来更高的精度。随着方法的改进和经验的积累，其效率——以及采用率——应会不断提高。

2 背景

作者的教育背景为数学、计算机科学与统计学。其职业为软件工程师，兼具估价与会计方面的辅助经验。他自 2002 年起断断续续从事估价工作，并自 2004 年起将 MARS 应用于住宅房地产的调整值确定和价格趋势分析。

使用 MARS 进行估价是估价领域的“深度经验”。MARS 生成的模型能比任何其他统计方法都更细致地反映哪些变量对市场价值有贡献，以及这些变量取值的变化如何影响售价。因此，在使用 MARS 估价的过程中，估价师不断考察由 MARS 模型揭示的变量取值与售价之间的因果关系。

出于透明性的考虑，RCA 方法最适用于具备以下特征的住宅区域：

1. 物业彼此差异很大的复杂社区

- 较老的住宅
- 更新程度参差不齐
- 多种不同的建筑风格与设计
- GLA、地块面积、海拔和景观等物业属性差异较大。

2. 复杂的市场状况

- 快速变化的市场状况
- 类型多样的买家

3. 良好的数据来源：RCA 规程（protocol）基于数据挖掘，需要充足的高质量数据。虽然它也可用于数据稀疏的区域，但可能不值得为此投入所需的时间和成本。

4. 需要准确的市场价值估计。RCA 在多数情况下应能提供 $\pm 1-2\%$ 的精度。估价工程师报告应就最终价值结论的精度条件加以说明。如果数据或信息不足，或未能通过数据挖掘发现

市场中的一致模式，精度就会受损，且应将这一情况告知客户。具体而言，较高的 R2 和 CVR2（交叉验证 R2，通过将 MLS 数据集反复随机划分为训练集和测试集计算得到）、有意义的 MARS 模型，以及排名残差中物业 CQA（状况-质量-吸引力）特征所呈现的强模式，是评估价值结论准确性与可靠性的主要手段。

过去 20 余年间这些 RCA 规程的发展，依托于 MLSListings Inc. (Sunnyvale, CA) 及其他目前可通过 MLSListings 访问的 MLS 数据，这些数据占作者数据分析经验的 95% 以上。作者的专业能力源自在北加州 15 个主要县开展的大量估价工作，并在 10 个次要县积累了额外经验。

RCA 方法在满足以下两个关键条件的市场中最为有效：

1. 高质量估价数据的丰富可得性
2. 复杂、异质的住宅区域，其特征为：
 - 多样的建筑风格
 - 程度不一的物业更新
 - 跨越数十年的开发模式

加利福尼亚州各县经验

本文所作论述基于在以下北加州各县开展物业估价的经验：

San Mateo	Santa Clara	Marin	Napa	Monterey	Santa Rosa
Alameda	San Francisco	Contra Costa	Sacramento	El Dorado	San Joaquin
Stanislaus	Placer	Solano	Mendocino	Lake	Colusa
Merced	Nevada	San Benito	Santa Cruz	Sutter	Yuba

上述区域基本涵盖硅谷及其周边各县，并向外延伸至更多乡村地区，包括中央谷地和内华达山麓。应当指出，与美国其他欠发达地区的许多估价师所能获得的数据相比，作者始终拥有充足且相对准确的数据。

所用的 MLS 数据经过 RESO 认证，符合 RCF (RESO Common Format)。尽管有 RCF 认证，一些物业的、通常属次要重要性的数据有时仍会缺失或有误，其原因或是最初税务评估员的录入错误，或是后续销售经纪人的 MLS 输入错误。这些错误往往遵循某种模式，例如特征不存在时干脆不做输入。举例来说，在没有车棚时，“carport spaces”一栏不输入 0，而是什么都不填。在车棚罕见的地区，分析师可以用 0 替代缺失数据，以便得到有用的车棚调整值。虽然 R/earth 及其他 MARS 实现可以处理缺失数据，但如何处理由估价工程师 (Valuation Engineer, VE) 决定。

MLSListings 的数据也可从其他一些州（从佛罗里达到夏威夷）的 MLS 获得。这对于测试规程和代码在不同州的运行效果很有用。

3 估价工程师

定义 1：估价工程师

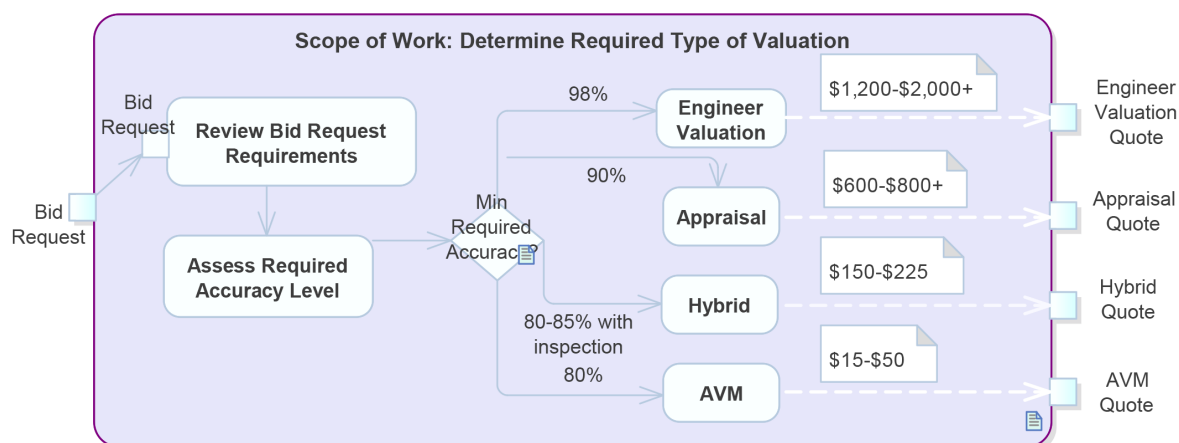
为避免歧义，“估价工程师”（Valuation Engineer，简称“VE”）一词指这样的人员：他既具备持牌估价师的专业能力，又在运用数据挖掘、统计学、编程、详细规程和人工智能来确定复杂物业市场价值方面拥有高级的知识、技能与经验。还应认识到，鉴于既已证实的欺诈可能性，估价是一项需要某种形式执业许可的工作。

贷款机构通常提供以下几种估价选项：

1. 自动估价模型（AVMs）——成本最低的选项
2. 混合估价——由无执照查勘员实地查勘，持牌估价师负责监督
3. 标准估价——由持牌估价师完成

通常不在估价工作考虑范围之内的，是工程师级估价。具备这类技能组合的估价师实在不多，但市场应当存在对他们的需求。借助 RCA 方法（MARS 加约束），他们能够提供显著更高的准确性，并几乎保证相当良好的客观性。

对于较新住宅小区中的合规物业，标准估价通常已经足够。然而，对于翻新历史长达 30 年以上区域中的老宅，则建议采用工程师估价。在这两种方法之间的选择应遵循基于物业特征和市场状况的结构化决策流程。



4 假设

假设 1: 地理区域 & 物业类型

本文的范围主要聚焦于加利福尼亚州大都市及城市地区的独户住宅与共管公寓。虽然市场比较法也适用于其他物业类型和地理区域，但在对农业用地、多户物业和商业房地产进行估价时，收益法和成本法通常发挥更主导的作用。在此类情况下，RCA 方法可能需要修改，或适用性有限。

假设 2: 数据存储

CSV 和 Excel 在 MLS 系统中被大量使用，并被假定为向 RCA 程序馈送数据的主要方式；当然，也可以使用经纪商数据源将数据持续送入 SQL 数据库。备份和中间存储常使用 SQL 数据库，其中扮演重要角色的两个是免费开源的 SQLite 和 PostgreSQL。文本或 RTF 文件也用于报告片段，PNG、JPG、TIFF 文件用于图形。需要注意的是，许多人也使用 Microsoft 的 SQL Server，但其免费版“SQL Server Express”限制很大（例如只能使用 1 个 CPU 核心），而其他版本用于商业用途则需支付高昂的许可费。加利福尼亚州的房地产数据库可能非常庞大，连接（join）操作可能非常耗时，因此对大型连接而言，一台 8 核以上、配备 128G 内存的计算机会有用。当需要查找任意边界区域内的所有物业时，SQL 数据库同样重要。因此，举例来说，如果使用聚类分析来划定社区，就会倾向于使用某种 SQL 数据库。然而，不同数据库处理地理信息的内置函数各不相同，因而互不兼容。在这一系列关于 RCA 的文章中，复杂的 SQL 任务假定使用 PostgreSQL，简单任务使用 SQLite。

假设 3: MARS 处理使用“R/earth”。

2017 年之前，Salford Systems 的 MARS 软件是一个价格亲民的解决方案，每年 \$270。然而，在 Minitab 收购 Salford Systems 之后，MARS 被打包进其 SPM 软件包，年费高达 \$16,000——这一价位令大多数小企业望而却步。幸运的是，由 Stephen Milborrow 维护的 R 语言“earth”包 (Milborrow 2024a) 已发展到与 Minitab 的 SPM MARS 相当乃至更强的水平。earth 包还具有额外优势：它免费可得，与 R 全面的统计生态系统无缝集成，并可通过 Python 借助 Pandas 访问，为不同工作流程提供了出色的灵活性。与 MARS 相关的讨论均假定具体使用 R/earth。

假设 4: 符号约定

来自 CSV 文件和 Excel 电子表格等外部来源的原始数据用标准拉丁字符表示。这包括多重挂牌服务 (MLS) 物业数据、使用 R/earth 包的 MARS (多元自适应回归样条) 配置设置，以及项目特定数据。这一记法同样适用于它们在 R 或 Python 中对应的数据框表示。在讨论物业特征及其各种类型和子集时，记法改用希腊字母，以增强精确性和分析上的清晰度。虽然这种数学记法看起来更为严格，但它有一个关键目的：将抽象的物业特征与其原始数据表示区分开来。按照惯例，希腊字符始终表示物业特征，而拉丁字符表示电子表格数据结构。例如，符号 M 表示 MLS 数据矩阵，其行为各物业，列为相应的特征。

假设 5: 编程语言

在本领域，R 和搭配 Pandas 的 Python 是主导语言，在对速度要求较高时则使用 C 或 C++。R/earth 本身即以 R 和 C 编写。

假设 6：索引

索引方式在不同编程语言与数学之间可能各不相同，这可能令人困惑：

- 数学和 R 从 1 开始索引。
- C、C#、C++ 和 Python 从 0 开始。

在本文中，为保持一致和简单，所有索引均从 0 开始（如 Python 或 C）。具体规则如下：

- 对于列表、矩阵或数据框：
 - 索引 0 保留给“估价对象”（即被研究的主要物业）。
 - 示例：在以行表示物业的表格中，第 0 行是估价对象，第 0 列存放所有物业的唯一 ID。
 - 实用提示：排列物业时将估价对象放在最前（第 0 行），其余物业按销售日期从新到旧排列，并按顺序编号（0、1、2 等）。
- 当没有估价对象时：
 - 如果数据结构不包含物业（例如仅有变量或函数），则跳过索引 0，索引从 1 开始。
 - 给 R 用户的特别说明：由于 R 使用 1 基索引，在用 R 实现本文的数学内容时，需将所有索引加 1，以从 0 基调整为 1 基。

5 RCA 工作流程

理解 RCA 工作流程是进行有效分析的关键。整个过程遵循一系列结构化的分析任务与决策节点

5.1 MARS 分析的类型

MARS 分析的工作流程取决于仅需进行市场分析, 还是需要对特定房地产进行估值或估价。请注意, 对特定房地产的估价通常需要对其所在市场区域进行分析。

- 市场分析: 侧重于县、市或界定区域等较大范围。这种方式不涉及具体的估价对象, 因此无需得出价值结论。工作流程以模型开发、验证以及生成配套分析结果为核心。一旦确定并记录了最优模型, 流程即告结束。
- 单一房地产分析涵盖市场分析的全部内容, 并进一步确定特定房地产的价值。这与传统估价方法有着根本区别。在传统方法中, 价值通常通过对 3-9 个可比交易实例进行调整并计算加权平均值得出。相比之下, RCA 主要通过两个部分来确定价值:

1. MARS 模型的预测值
2. 估计的房地产特定残差 (连同相应的 CQA 评分)

在 RCA 中, 销售网格的主要作用是通过展示各变量的贡献, 来解释估价对象与可比交易实例之间的价值差异。RCA 的一个关键优势在于其稳定性: 尽管客户可能会对可比交易实例的选择提出异议, 但无论选择哪些可比交易实例用于展示, 价值结论都保持一致, 因为所有成交案例 (无论是 15 个还是 600 个) 都是使用同一模型进行调整的。虽然较早的成交案例会被自动调整至当前市场状况, 但一般而言仍以近期成交案例为佳。随着时间跨度的拉长, 未被测量的市场特征变得愈发重要, 模型捕捉价格因素的效果也随之减弱。

5.2 处理工作流程

整体处理工作流程可以用以下步骤来描述:

1. 工作范围。
2. 建立初始项目文件夹与文件。
3. 下载 MLS 数据。关于所需房地产记录数量以及日期范围限制的设定, 请参阅规程 1.1-1.3。
4. 完成各配置工作表: 回归变量、交互项、聚合项、计算项。
5. 运行 MARS (R/earth) 回归以建立价格模型。
6. 由 MARS 模型生成各房地产的价值贡献与残差 (纯残差以及残差/SF)。
7. 按残差/SF 对房地产进行排序。
8. 为每个可比交易实例创建 CQA 评分。
9. 假设已经生成了一份真实反映各房地产整体吸引力的质量排序, 则通过研究排序中的总体模式趋势, 为估价对象在排序中找到最合适的位置, 并为其赋予 CQA 评分。
10. 将相应的残差/SF 赋予估价对象, 即残差/SF · (估价对象) GLA。
11. 将 MARS 价格估计值与所赋残差相加, 得出估价对象的价值结论。
12. 将估价对象的残差分解为描述性组成部分及其价值贡献。

13. 计算可比交易实例由模型产生的价值贡献与调整额, 以及所有小计和调整后的成交价格。
14. 调整后成交价格将等于估价对象的最终价值结论。
15. 选取 6–12 个最相似的房地产纳入销售网格。
16. 将可比交易实例的残差总额分解为残差组成部分的价值贡献, 分解完成后将残差总值设为 0。
17. 重新计算所有小计以及销售网格中可比交易实例的调整后成交价格。核对数值。
18. 将销售网格上传至报告。

请注意, 采用 RCA 时, 最终价值结论并非直接来源于销售网格中的可比价值。在 RCA 方法下, 销售网格仅仅成为一种解释工具, 用于说明为何那些看似与估价对象相似的房地产, 其成交价格会高于或低于最终价值结论或估价对象的评估价值。必须换一种方式来看待这一点: 估价对象的价值结论取决于 MARS 模型的价格估计值, 以及估价工程师所估计的残差。但这两个量在很大程度上依赖于参与 R/earth 回归的全部房地产, 该回归生成了房地产残差的排序, 而估价工程师通过将估价对象置于该排序中的某一位置, 确定了其 CQA 评分, 进而确定了其残差。销售网格中的调整实际上是由 RCA 所施加的数学约束决定的; 尽管残差分解在一定程度上取决于估价工程师的判断, 但分解出的价值贡献合计必须始终等于相应的残差。我们从证明 2 可知, 虽然残差各组成部分的实际数值可能因估价工程师的判断而异, 但对于给定房地产, 其合计值是完全受约束的, 因此无论估价工程师如何改动残差分解, 都不会改变该房地产的调整后成交价格。

5.3 数据工作流程

请注意, 本文仅对 RCA 方法给出高层级的工作流程。其他文章将深入探讨 RCA 分析的各个阶段与子阶段。

5.3.1 MARS 的数据要求

我们需要指定销售交易的起始日期, 并根据估价或估值的基准日确定截止日期。这通常需要考虑每一交互度 (大于 1 时) 所需回归变量与交互项数量的估计值。

5.3.2 规程：已成交交易挂牌信息

规程 1.1：确定所需 MLS 已成交交易挂牌信息的数量

本规程说明如何计算输入 MARS 回归所建议的已成交交易挂牌信息的最低数量。你需要先做一个粗略估计，随后运行一阶、二阶乃至三阶的 MARS 以生成初始模型，然后按照下述规则审查数据行数，判断是否需要更多记录。本规程假设各自变量之间是 100% 相互独立的，而这在房地产领域几乎不可能成立；例如，GLA 与房间数量通常高度相关，GLA 也很可能与地块面积存在相关性。这就涉及到“自由度 (DOF)”的计算。每一个真正独立的变量占一个自由度。但是，当所谓的自变量之间存在依赖关系时，准确计算 DOF 需要进行一定的分析，这已超出本文的讨论范围。只需记住，由于变量之间很可能存在共线性，下述估计值可能偏高。

1. 对于一阶项（无交互项），每个变量约 ~ 15 行的经验规则效果良好
2. 对于每一个允许的双向交互项，应额外增加约 15-30 处房地产
3. 对于更高阶的交互项（3 个及以上变量），每项应有 30 处以上房地产。

不过，这并非硬性规则。实际所需数量可能取决于：

- 数据中的信噪比
- 所要建模的关系的复杂程度
- 响应变量的变异程度
- 数据的质量

示例：

因此，如果你有：

- 10 个主变量
- 5 个允许的双向交互项
- 2 个允许的三向交互项

则可按如下方式计算所需成交案例的最低数量：

$$\begin{aligned} N &= (10 \text{ 个变量} * 15 \text{ 行/变量}) + \\ &\quad (5 \text{ 个双向交互项} * 22 \text{ 行/双向交互项}) + \\ &\quad (2 \text{ 个三向交互项} * 30 \text{ 行/三向交互项}) \\ &= 320 \text{ 行} \end{aligned}$$

或者，也可以采用以下做法：在对一组数据运行 R/earth 之后，可以用模型中基函数的数量加一作为自由度的估计值，再将该数值乘以 20-30，即可得到所需输入记录的数量。

规程 1.2: 销售期间应尽量避免购房者偏好发生变化

在收集历史房地产数据时, 如果市场偏好的变化能够通过适当的变量加以捕捉, 则对时间跨度没有严格限制。在处理市场中建筑风格演变时, 请考虑以下要点: 主流建筑风格的变化 (例如从加州牧场式向地中海式的转变) 可以通过以下方式建模:

- 小区 (subdivision) 标识符
- MLS 区域代码
- 建造日期区间

然而, 如果这些风格变化无法通过特定变量得到充分捕捉, 你将面临一个重要约束: 销售日期通常用于调整市场状况, 若同时用它来控制建筑风格偏好, 会使这两种截然不同的效应相互混淆。在这种情况下, 较为可取的做法是二者择一:

1. 将数据收集范围限制在较新建筑风格出现之后的时期
2. 将可比交易实例的选择限制在具有相同建筑风格的小区之内

这样既能确保分析在房地产特征上保持一致性, 又仍然允许对市场状况随时间的变化进行恰当调整。或者, 至少在这种情形下, 假设你能够从数据中提取出建筑风格, 可以考虑这两个变量之间的交互作用。然而, 这种提取并非总是可行的。

规程 1.3: 若处于成交稀少的区域, 则尽力而为

在使用 MARS 对销售数据建模时, 进行有意义的分析大约需要 15 笔交易。如果成交案例更少 (例如只有 5 处房地产), 可以将每条记录复制两次, 以达到 15 处房地产的门槛。采用这种做法时的重要注意事项:

1. 仅使用一阶回归, 以降低过拟合
2. 由于数据集过于有限, 忽略或跳过交叉验证
3. 在这种情形下, MARS 的表现很可能优于配对比较法或标准线性回归方法, 但要注意, 复制数据会影响不确定性度量与统计显著性的有效性

最好的长期解决方案是通过补充真实成交案例来扩充数据集。不过, 如果必须在数据有限的情况下继续开展工作, 数据复制在承认其统计局限性的前提下, 不失为一种可行的临时方案。

5.4 项目文件夹结构

有必要讨论一下项目文件夹结构。建议为市场分析和单一房地产估价分别设立单独的文件夹。这两个文件夹具有相似的子文件夹结构：

```

RCA_Projects/
├── MarketAnalysis/
│   ├── Burlingame_20240501/
│   ├── Pacifica_20240310/
│   ├── RedwoodCity_20240615/
│   └── ...
├── SingleFamily/
│   ├── Pacific_Rosewood_123_20240510/
│   │   ├── Code/
│   │   │   ├── R/
│   │   │   │   └── ...
│   │   │   └── Python/
│   │   │       └── ...
│   │   └── Data/
│   │       ├── MLS_Original.xlsx
│   │       ├── MLS/
│   │       │   ├── VER202402510153248/
│   │       │   │   ├── MLS.xlsx
│   │       │   │   ├── MLS_Stage2.xlsx
│   │       │   │   ├── MLS_Stage3.xlsx
│   │       │   │   ├── MLS_Stage4.xlsx
│   │       │   │   ├── MLS_Stage5.xlsx
│   │       │   │   └── Documents/
│   │       │   │       ├── LinearModels.pdf
│   │       │   │       ├── Cum_Dist.pdf
│   │       │   │       ├── Model.txt
│   │       │   │       ├── PartContrib.pdf
│   │       │   │       ├── Pairs.pdf
│   │       │   │       ├── Residual1.png
│   │       │   │       ├── Residual2.png
│   │       │   │       ├── ResidualSF.png
│   │       │   │       ├── Rpart.pdf
│   │       │   │       ├── VariablesImp.pdf
│   │       │   │       ├── VariableVsSP.pdf
│   │       │   │       └── ...
│   │       └── ...
│   └── ...
└── ...

```

第 1 阶段处理将创建版本文件夹及其子文件夹。它会复制 `MLS_Original.xlsx`——尽管该电子表格理应永不更改，而是精确反映在某一日期和时间从 MLS 下载的原始内容；但在周期较长的项目中，仍有可能需要重新从 MLS 下载数据。接着，此阶段会将经过扩展的 `MLS.xlsx` 工作簿连同其配置参数复制到版本文件夹中。版本文件夹可以有很多个，每个都对应一次完整的 MARS 运行。

随后的各个阶段将对工作簿进行列的添加、排序及其他更改，并在估价完成时将更新后的版本以 `MLS_Stage(N)` 的形式写入项目数据文件夹。再次强调，根目录下的 `MLS.xlsx` 处于持续变动之中，因此版本副本是对原始 MLS 工作簿的一种持续演化，包含变量的删除与添加以及其他更改。处理用的 R 或 Python 代码很可能以 Data 文件夹中的 `MLS.xlsx` 工作簿作为输入运

行。这样，我们始终保有每一版 RCA 运行背后全部数据的副本。

版本可以根据需要不断创建，直至无法再做出改进为止，此时选出最佳版本作为最终版本。可以在文件夹名称末尾加上“_Final”，以便浏览目录时一目了然。一些包含有趣模型或竞争性模型的旧版本通常应予保留。每个版本文件夹都包含若干文本、图示和图形文件，可嵌入到报告中。最终的 MLS_Stage5.xlsx 工作簿包含报告所需的全部工作表。特别是，它可能包含一张用于转入 Fannie Mae 销售网格的电子表格。它还必然包含聚合项背后的计算或分解过程。这可能是 RCA 方法中最耗时的部分——在一个困难的市场区域中苦苦寻找完美模型的过程。

重建销售网格所需的一切都必须复制到版本文件夹中，包括所使用的 R 或 Python 代码——如果代码经常变动的話。

5.4.1 规程：多次 MARS (earth) 运行

规程 2.1：利用过拟合运行精细剖析市场区域的定价因素

在最初的 earth 运行中，应当有策略地利用过拟合来发现潜藏的价格因素。例如，一个由 800 笔成交案例构成的市场区域中，可能只有 3 处房地产带有工房（workshop），而在这些孤立案例中，工房可能对最终成交价格产生重大影响。仅执行前向传递的运行会产生一个过拟合模型，从中可以看出这些工房的近似价值贡献。然而，由于 800 笔成交中只有 3 笔这样的案例，工房相关项将在后向传递中被剔除，在交叉验证下更是必然如此。可以将仅前向传递赋予工房的价值记录下来，转入潜在变量分析，供销售网格中出现的任何带工房的房地产使用。此类过拟合模型还能在许多其他数据稀薄的特征给出详细估计，而这些特征本会被后向传递剪枝或交叉验证运行所剔除。

将每一次这样的探索性运行都保存为独立的版本文件夹：所记录的价值贡献是估价工程师日后分解残差时会想要引用的证据。当然要明白，估价工程师最终必须依靠剪枝和交叉验证，才能使最终模型对回归从未见过的房地产——例如后来出现的估价对象——保持稳健。过拟合运行是用于研究的工具；它们绝不是最终模型。

规程 2.2：在不同 earth() 参数下进行多次交叉验证运行

市场区域在结构和规模上各不相同，估价工程师通常需要通过一系列运行来进行试验：以了解哪些变量对价格影响最大；判断现有的已测变量集合是否足以估计价值，抑或必须补充新数据（如距水域的距离、海拔、距购物中心的距离等）；并发现新型的交互关系，同时将其他交互关系作为低效或具有误导性的项予以排除。最终，估价工程师希望在训练数据上获得良好的 R^2 ，并在测试数据上取得尽可能好的 CVR^2 。要达到这一目标，可能需要收集新数据、提高现有数据的准确性，或根据市场区域的规模与特点（例如针对可用的自由度——参见规程 1.1）调整 earth() 参数。

这正是上述项目结构支持多个运行版本文件夹的原因：每个候选模型都连同其参数及其 R^2/CVR^2 一并保留，从而可以比较各次运行之间的演变轨迹。若某一版本的 R^2 上升而其 CVR^2 停滞甚至下降，则说明该版本拟合的是噪声而非市场信号，而正是版本历史使这一点得以显现。

规程 2.3: 利用 Prolog 展开备注及描述性特征列

尽管这一功能仍处于一定的试验阶段, earthUI 现已内置 Prolog 引擎 (vProlog), 并使用定子句文法 (DCG) 来解析 MLS 的自由文本列——首先是 Public Remarks (公开备注), 此外还有经纪人备注以及以逗号分隔的特征列表, 例如房间特征或既有建筑物。这种展开会将“owned solar”“3 car garage”“40x50 workshop”之类的短语转化为可供回归使用的普通特征列, 而以此方式挖掘出的数据稀薄特征可直接输入潜在变量分析 (参见后文关于将状况与质量评级作为近似回归变量的小节; 该功能本身在本期的 earthUI 配套文章中有详细说明)。

文本展开会成倍扩大候选预测变量集合, 因而必须进行多次 earth/RCA 运行加以测试: 展开出的列必须在模型中反复取舍试验, 在存在共线性时合并为派生的计数或布尔变量, 并对照 CVR2 检验其承载的是市场信号而非噪声。应将每一次此类试验视为独立的运行版本文件夹, 与规程 2.1 和 2.2 的做法完全一致。此外, 由于解析文法本身仍在开发之中, 估价工程师在将提取出的列用于模型之前, 应先将其与源文本进行核对。

5.5 核心 Excel 工作簿

由于大多数 MLS 服务将房地产数据下载为 Excel CVS 工作表, Excel 可能是存储数据及各类配置参数最便捷的方式。每个项目都有一个名为 MLS.xlsx 的核心 Excel 工作簿, 其中包含一张名为“MLSDData”的工作表, 用于存储从 MLS 下载的已成交房地产销售挂牌信息, 此外还有若干配置与参数工作表:

```
Workbook: MLS.xlsx/
├─ Sheet: Project Data
├─ Sheet: MLSDData
├─ Sheet: Regression Variables
├─ Sheet: Calculations
├─ Sheet: Allowed Interactions
├─ Sheet: Aggregations
└─ Sheet: (MARS run parameters)
```

原始数据存储在一个名称类似 MLS_Original.xlsx 的工作簿中。应当以不同名称——比如“MLS.xlsx”——制作一个副本, 由 R 或 Python 的 RCA 程序加载。这两个工作簿都存放在 Data 文件夹下。此后, Data/MLS.xlsx 工作簿很可能会经历多轮迭代运行, 尝试不同的变量组合、MARS 参数以及其他设置的更改。每次迭代都会创建一个以日期和时间命名的新输入/输出文件夹, 格式为 VER_YYYYMMDDHHMMSS。每个文件夹都会保存输入的 MLS.xlsx 工作簿的一个副本, 并包含四个处理阶段 (第 2-5 阶段) 各自生成的 Excel 工作簿, 以及输出的文本文档、图示和图形。请注意, 最后一个阶段即第 5 阶段, 可能会根据客户类型和工作范围 (SOW) 包含若干供报告使用的其他工作表。

5.6 备份

在将新建或修改后的 Excel 电子表格写入项目版本文件夹时, 还可以在项目文件夹中向 SQLite 或 PostgreSQL 数据库进行备份。数据库文件在后续时期的审阅或参考使用过程中, 较不容易被意外损坏。

6 工作簿记号

6.1 描述维度

工作簿“M”大约有六个描述维度：

1. 估价项目“a”
2. 运行版本“r”
3. 工作流阶段“w”
4. 工作表“s”
5. 房产挂牌信息“p”
6. 变量集“v”

因此，特定工作簿 M 中的某个单元格可以描述为：

$$\begin{matrix} r_* \\ w_* \end{matrix} \begin{matrix} a_* \\ W \\ s_* \end{matrix} \begin{matrix} p_* \\ v_* \end{matrix} \quad (1)$$

(注：“*”表示某个有效索引)

或：

$$\begin{matrix} r_k \\ w_m \end{matrix} \begin{matrix} a_g \\ W \\ s_h \end{matrix} \begin{matrix} p_i \\ v_j \end{matrix} \quad \begin{matrix} i = 1, \dots, n_p \\ j = 1, \dots, n_v \\ k = 1, \dots, n_r \\ m = 1, \dots, n_m \\ g = 1, \dots, n_g \\ h = 1, \dots, n_h \end{matrix} \quad (2)$$

其中：

- n_p = 行数（取决于工作表）
- n_v = 列数（取决于工作表）
- n_r = 运行版本数
- n_m = 工作流阶段数
- n_g = 分析项目数
- n_h = 工作簿工作表数

请理解，“W”指的是整个 Excel 工作簿，它包含六个核心数据工作表，每个工作表都有自己的表格；在运行相应的 R 或 Python 程序时，这些表格会被加载到各自独立的内部数据框中。R 和 Python 都有内部函数，可以根据名称查找它属于数据框的哪一行或哪一列。例如，如果在下标 v_j 中， $v = (\text{“GLA”}, \text{“LotSize”}, \text{“BathRms”}, \text{“BedRms”}, \dots)$ ，且 $j=0$ ，则 $v_j = \text{“GLA”}$ 。R 或 Python 都能够在内部将“GLA”解析为其所存储的列，例如，这可能是相应数据框的第 15 列。因此，在本文中我们对名称数组进行索引，把名称传递给运行中的程序，程序再将该名称重新索引为内部数据框中实际的行或列索引。这有点复杂。但为了精确起见，这样做是必要的，也是最方便的。这是分析师角色与开发者角色之间的“职责分离”；尽管在现实中，估价工程师 (VE) 可能同时承担这两个角色。请注意，R 中的“data frame”在 Python 中称为“dataframe”。

另请注意，MARS 运行时参数工作表并不是 W 的一个维度，因为 W 关系到输入 MARS 的数据版本，而与决定 MARS/earth 如何处理这些数据的参数无关。

为清晰起见，本文给出了一套复杂的多维工作簿记号，它与数据工作簿各工作表所加载到的内部 R 或 Python 数据框更直接地对应。这是一套“沉重”的记号。但当我们进入证明等环节时，通过对所处理对象作出假设，就可以抛掉大部分包袱。尽管如此，这套记号有助于我们从一开始就更好地总览工作流和数据。

我们先从将要处理的工作簿集合的记号开始。

6.1.1 项目维度 (“a”)

一旦工作范围 (SOW) 获得批准, 就需要建立一个项目文件夹来存储输入和输出数据, 包括配置参数、输出图形、图表、电子表格和报告片段。VE 应当用简单的顺序 ID (例如整数) 标识所有项目, 这样一旦某份报告丢失, 就会一目了然。如果项目被取消, 应当有一致的方式在相应项目文件夹内注明该项目已被取消, 但绝不应删除该文件夹。我们假设项目 ID 从 0 开始并逐次加 1, 尽管也存在其他可能。我们进一步假设所有有效的项目 ID 都存储在由处理程序创建的有序集合 “a” 中。

估价项目数据位于工作簿的项目工作表中, 但也可以存储在某个中心 SQL 数据库中,

以下是建议的项目字段:

1. Local ID: 为公司顺序生成的短整数 ID。它非常适合作为存储用的主键——并且可以由数据库在每个后续项目中自动生成。
2. Global GUID: 如果需要, 可以使用以下网址创建一个 GUID (全局唯一 ID)

<https://www.uuidgenerator.net/>

3. 房产位置 (市场分析) 或地址
4. 估价生效日期
5. 委托日期
6. 查勘日期
7. 原始报告日期
8. 第一次更新报告日期
9. 第二次更新报告日期
10. 第三次更新报告日期
11. 客户名称
12. 客户地址
13. 贷款机构名称
14. 贷款机构地址
15. 业主名称
16. 业主地址
17.

项目 ID 列表示例:

```
a = ("1",
      "2",
      "...",
      "141",
      "142")
```

6.1.2 运行版本维度 (“r”)

工作流的第二步是对输入执行 MARS 回归。这很可能需要数次乃至多次执行, 因为我们会调整各种参数以改进模型。每次执行都会为输出创建一个新的版本文件夹, 其名称基于执行的日期和时间。例如,

```
r = ("ver20241210_083124",
      "ver20241209_151130",
      "ver20241208_160944")
```

latest_version : $r_1 = \text{"ver20241210_083124"}$

具有竞争力的模型和输出的有希望的版本应当保留，直到确定最终版本为止；如果它们仍有某些用处，或许还可以保留更久。你可能希望修改版本文件夹名，在名称末尾加上“Final”，以表明它是最终版本。

6.1.3 工作流阶段 (“w”)

RCA 工作流通常分为五个独立的阶段，如下所列。每个阶段必须成功完成后才能开始下一个阶段。VE 应能够在某个阶段结束时停止处理，先审查结果再继续。这是必要的，因为市场区域相当复杂，而且往往颇为独特。可能会遇到需要采用新方法的异常问题。VE 必须能够进入 R 或 Python 代码进行必要的修改，甚至尝试新方法。而且在这种情况下，他可能希望进行试验并尝试不同的方法，这就需要对他正在处理的同一阶段进行多次运行。当然，VE 也应当可以用一条命令运行所有阶段——但有一些前提（例如，三阶聚合必须已通过之前的 Stage 2 运行确定）。这意味着每个阶段完成时，都应把其全部内部数据存入文件或数据库，以便下一个阶段可以通过把上一阶段的数据重新加载到内存中来启动，就好像上一阶段仍驻留在内存中一样。

以下是各个阶段：

```
w = ("stage1", -设置
      "stage2", -MARS 处理
      "stage3", -最终销售网格
      "stage4", -报告片段
      "stage5") -审查/清理
```

stage2 : $w_2 = \text{"stage2"}$

6.1.4 Excel 工作簿工作表 (“s”)

在 RCA 核心 Excel 工作簿中，我们有以下工作表，每个工作表都映射到一个内部数据框：

```
s = ("MLSDData",
      "Regression Variables",
      "Calculations",
      "Interactions",
      "Aggregations 1st & 3rd Degree",
      "Aggregations 2nd Degree")
```

之所以把二阶聚合放在单独的工作表上，是因为它们最容易通过一个对称表来定义：该表展示所有可能的交互对，允许为每一对指定一个聚合变量，将其相应的价值贡献和调整聚合（相加）到该变量上。三阶聚合无法这样处理，因为那需要一个三维对称表，而这在 Excel 中是不可能的。此外，一个 20x20x20 的表将有 8,000 个单元格需要处理。更简单的做法是把任何三阶项的名称直接列在包含一阶聚合的表中，例如：

Variable	Aggregate To
AboveGradeFinArea	GLA
BelowGradeFinArea	BGLA
UnfinArea	UBA
LotSize	LotSize
...	...
GLA___LotSize	GLA
GLA___DateOfSale	GLA
BathRms___DateOfSale	GLA
Lat___Long	Location
Lat___Long___GLA	Location
...	...

6.1.5 房产挂牌信息 (“p”) ——MLSDData 行

房产挂牌信息构成 MLSDData 工作表的行。它们不会出现在其他工作表中，直到 Stage 4——在该阶段，将使用最终选定的可比交易实例生成最终的报告片段。以下是房产 ID 的示例列表：

$$p = (\text{“MLS823211”, “MLS823456”, “MLS823542”, “MLS823721”, “...”})$$

6.1.6 MLS 变量 (“v”) ——MLSDData 列

关于变量的处理，后文还会详述。此处只需指出：MLSDData 工作表通常有 30-40 个类型各异的变量，但通常只有 10-20 个会被选作 MARS 回归的自变量。这些变量也出现在其他工作表中，不过通常是作为可供提交给 MARS 的所有可能变量的列表。变量列表的一个示例是：

$$v = (\text{“GLA”, “SalePrice”, “BedRms”, “BathRms”, “GarageSF”, “...”})$$

6.1.7 最终记号示例

让我们把上述记号汇总成一个最终示例。使用上面的列表，令

$$r_1 \begin{matrix} a_{141} \\ W \\ w_1 \end{matrix} \begin{matrix} p_2 \\ v_0 \end{matrix} \quad (3)$$

请注意，这里我们使用 0 起始索引，即假定列表的第一个元素的索引为 0。Python 也使用 0 起始索引，而 R 使用 1 起始索引。

因此，上式表示我们所讨论的是以下单元格的值：

- 变量 v_0 ，即列 “GLA”
- 房产 p_2 ，即 ID “MLS823542”
- 工作表 s_0 ，即 “MLSDData”
- workflow 阶段 w_1 ，即 “stage2”
- 版本 r_1 ，即
“ver20241209_151130”

- 项目 a_{141} ，即“142”

上述记号也可以写成：

$$\begin{array}{ccc} \text{"ver20241209_151130"} & \overset{\text{"142"}}{W} & \text{"MLS823542"} \\ \text{"stage2"} & \text{"MLSDData"} & \text{"GLA"} \end{array} \quad (4)$$

6.1.8 理由

“为什么不在记号中直接使用索引或数字呢？”我们对记号有两个要求：

1. 子集。例如，分析人员经常选取列表的子集，比如变量的子集。举例来说，MLSDData 工作表中可能有 35-40 个变量，但某次运行只使用其中的 10-20 个。从一次运行到下一次运行，VE 会频繁更改提交给 MARS/earth 的变量。
2. 迭代。例如， $i = 1, 2, \dots, n_1$
我们的算法、规程和证明通常需要能够从头到尾遍历有序列表中的所有项——或挑出某个特定的项。

因此，我们需要一个字母来表示项的子集或列表，并且它必须带有另一个下标，供我们遍历列表中的各项，以给出用于索引数据框行或列的字符串。

6.1.9 矩阵？

那么线性代数中的矩阵呢？矩阵只能处理数字，即整数或实数。R 和 Python 都支持矩阵这种数据结构。然而，我们在 MLS 数据中处理的许多变量是所谓的“字符串”或字符数据。即使有些变量只是数字，它们往往也不是实数，而只是名称。例如，某城市的 MLS 区域代码可能是（“660”、“661”、“662”、“663”、...、“672”）。Earth 回归允许把变量指定为“因子”变量，此时这些数字仅被当作名称处理，不含任何顺序。在这种情况下，生成的模型只会为每个数字提供一个价值贡献，而不是一个在序列上取值的函数。其他例子包括表示状况的“C1”、...、“C6”和表示质量的“Q1”、...、“Q6”。

6.1.10 第一次简化

上述针对核心 Excel 项目工作簿的记号很适合把各部分联系起来，但用于解释算法或数学证明就显得太沉重了。在做繁琐的具体工作时，我们可以假设一切都发生在给定的项目、版本、阶段和工作表之内。如果所处理的对象尚不明确，我们直接说明即可。一旦指明了所处理的对象，我们真正需要的只是：

$$M_{s_h}^{p_i v_j} \quad \begin{array}{l} i = 1, 2, \dots, n_p \\ j = 1, 2, \dots, n_v \\ h = 1, 2, \dots, n_s \end{array} \quad (5)$$

其中：

n_p	=	房产数量
n_v	=	输入变量数量
n_s	=	输入工作表数量

6.1.11 第二次简化

更方便的做法是按工作表所承载的数据框的具体类型来重新命名它们，而不是使用索引。对每个工作表而言，行和列具有不同的功能。因此，W 这一数据框集合被拆分为 M、V、A、I 和 C 数据框，如下所示：

$$\diamond \text{ MLS 数据} \quad M_{v_j}^{p_i} = W_0^{p_i} \quad \begin{matrix} i = 0, 1, \dots, n_p \\ j = 0, 1, \dots, n_v \end{matrix} \quad (6)$$

其中： n_p = 房产数量
 n_v = 输入变量数量

注意：估价对象默认始终位于房产列表首位

在“MLSDData”Excel 电子表格及相应的数据框中，第一行始终保留给估价对象。它绝不会与其下方的其他 MLS 房产一起参与排序。它也从不作为自变量数据输入 MARS 回归，而是作为目标数据，即因变量输入。房产列表同样如此。在所有房产列表中，除非另有说明，索引为 0 的第一个元素始终被假定为估价对象的 ID。

即使任务是为某个没有估价对象的市场区域建立模型，第一行也仍会加载一些虚拟的房产数据。这使得编写程序、给出算法或进行证明都方便得多。因此，如果我们用 c 表示房产列表，则 c_0 是估价对象，而 c_i ($i = 1, \dots, n_p$) 是将输入 MARS/earth 回归的可比挂牌信息。

◇ 回归变量规格说明

$$V_{s_j}^{v_i} = W_1^{v_i} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_s \end{matrix} \quad (7)$$

其中： n_v = 可用变量数量
 n_s = 规格说明数量

注意， $n_s \leq n_v$ ，因为我们通常从可用变量中选取一个子集作为回归的输入。

◇ 计算

$$C_{c_j}^{v_i} = W_2^{v_i} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_c \end{matrix} \quad (8)$$

其中： n_v = 变量数量
 n_c = 计算规格数量

◇ 交互

$$I_{t_j}^{f_i} = W_3^{f_i} \quad \begin{matrix} i = 1, 2, \dots, n_f \\ j = 1, 2, \dots, n_t \end{matrix} \quad (9)$$

其中： n_f = “from” 变量数量
 n_t = “to” 变量数量

请注意，数据框 I 是一个对称数据框，变量按字母顺序排列于“from”行和“to”列。单元格标记为“1”表示允许相应变量对之间的交互；否则，“0”表示不允许交互。对于二阶交互，这相对简单，只可能存在一种交互。对于三阶回归，三个变量之间共有三个变量对交互，必须允许其中每一对交互，才能生成三变量交互项（或变量）。

◇ 聚合

$$A_{v_j}^{v_i} = W_4^{v_i} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_v \end{matrix} \quad (10)$$

其中： n_v = 变量数量

7 RCA 销售网格

请记住，RCA 规程需要一个能够调用 R/earth 并执行所有细节步骤的 R 或 Python 程序。这项工作数据密集，绝不可能以手工方式完成。同时，假设程序没有缺陷，您也需要一个软件程序来确保所需工作精确无误地完成，避免人为差错。

对于从事这项工作的估价工程师来说，能够根据需要修改此类程序也是一种必要能力，尤其是在面对新的、未曾预见的市场区域、社区、房产或其他新需求、新问题时。

在 RCA 处理过程中，我们有三种类型的销售网格：

1. RCA 处理销售网格，在本文中标记为“MLSDData”，它是一个非常庞大的数据框及其关联的电子表格，可能包含许多中间细节，这些细节最终经过聚合处理，得到下文提到的 RCA 销售网格。示例太大，无法在此展示。估价工程师很可能会按照各自的偏好来定制这一电子表格。我打算在将来另撰一篇论文，专门讨论此类网格的开发。
2. RCA 最终销售网格，用于非 GSE 报告，其形式类似于表 2。它确实一定程度上模仿了例如 Fannie Mae 的表格，足以让读者有一定的熟悉感。不过请注意，它为每个可比交易实例提供的列数多于 Fannie Mae 网格：每个可比交易实例除“测量值”和“调整”列之外，还有一个“价值贡献”列，估价对象也同样有一个——这些列是 Fannie Mae 表格所没有的。一个更简单的形式见表 1。
3. GSE 上传销售网格，用于将销售网格上传到诸如 AlamoDe 之类的软件中。但请注意，并非所有估价表格软件包都提供从电子表格上传销售网格数据的功能。出于必要，此类网格不能包含单独的价值贡献列，而 Fannie Mae 等 GSE 对价值贡献毫无概念。我们的策略是：向它们提供其所要求的表格，但以 RCA 销售网格作为支撑，并辅以其他支持性文档，例如各变量聚合的详细信息。

RCA 处理销售网格看起来不像 Fannie Mae 销售网格，而大致相当于 Fannie Mae 销售网格的转置，行与列互换：每一行包含某一房产的数据，每一列包含某一房产变量的数据。房产行可以轻易达到数百或数千行，列数可达 200 列或更多。第一行保留给变量名，第二行保留给估价对象，即使我们只是在做市场分析也是如此。当电子表格被读入数据框后，工作表的第一行成为 names 属性，而估价对象所在的第二行则成为数据框的第一行。

RCA 销售网格中的列包括标识列，然后是主要变量、外部变量和合成变量，它们合并后作为自变量输入 MARS。MARS 从这些变量中输出它认为对成交价格（也称因变量或目标变量）有显著影响的变量。在二阶回归中，我们可能得到成对的变量，它们作为新的变量发挥作用。在三阶回归中，我们还可能得到变量三元组，它们作为所谓的自变量发挥作用，尽管它们当然与底层变量高度相关。

表 1: RCA 电子表格示例：纵向布局

估价对象				可比交易实例 1				可比交易实例 2			
	测量值	价值贡献		测量值	价值贡献	调整		测量值	价值贡献	调整	
成交价格				\$1,150,000				\$1,500,000			
基础价值		\$500,000			\$500,000	\$0.00			\$500,000		\$0
		已测量特征			已测量特征				已测量特征		
居住面积	1,530	\$463,500		1,350	\$382,500	\$81,000		1,700	\$540,000	-\$76,500	
地块面积	6,000	\$200,000		6,500	\$220,000	-\$20,000		7,000	\$240,000	-\$40,000	
房龄	53	\$63,600		57	\$68,400	-\$4,800		51	\$61,200	\$2,400	
浴室数	2	\$15,000		3	\$30,000	-\$15,000		3	\$30,000	-\$15,000	
卧室数	3	\$6,000		4	\$9,000	-\$3,000		4	\$9,000	-\$3,000	
测量合计		\$1,248,100.00			\$1,209,900.00	\$38,200			\$1,380,200.00	-\$132,100	
		残差特征			残差特征				残差特征		
残差	CQA 5.8	\$30,000		CQA 2.5	-\$59,900	\$89,900		CQA 8.2	\$119,800	-\$89,800	
设计	残差分解			残差分解				残差分解			
状况		7,000			-\$15,000	\$22,000			\$35,000	-\$28,000	
质量		5,000			-\$8,000	\$13,000			\$20,000	-\$15,000	
功能效用		8,000			-\$10,000	\$18,000			\$30,000	-\$22,000	
配套设施		5,000			-\$12,000	\$17,000			\$15,000	-\$10,000	
小计		5,000			-\$14,900	\$19,900			\$19,800	-\$14,800	
		30,000			-\$9,900	\$89,900			119,800	-\$89,800	
合计				合计				合计			
价值贡献总计		\$1,278,100.00			\$1,150,000.00				\$1,500,000.00		
调整后成交价格						\$1,278,100.00					\$1,278,100.00

表 2: 死亡日期估价使用的 RCA 销售网格副本 (已删除房产识别信息)

[illegible]

8 变量

在运行 MARS 的输入端，我们只有简单变量，它们是：

1. 从 MLS 电子表格加载的物业特征。
2. 由 VE 创建的派生变量，用于细化 MLS 变量或添加现有变量的变换。例如，VE 可能倾向于将销售日期变换为“DaysOffMarket”，其值等于估价生效日期减去销售日期。
3. VE 可以从第三方来源导入数据，例如按揭利率。或者，如果 MLS 数据中缺少纬度、经度或海拔数据，他可以从其他来源导入。这些变量可称为“外部变量”。

在 MARS 的输出端，我们处理的价格模型是若干项的组合，这些项：

1. 仅包含一个输入变量，并给出该变量价值贡献的方程。这些称为一阶项或（输出）变量。该项本身被视为一个价值贡献值并被赋予名称。
2. 包含两个输入变量，是给定两个变量的交互作用之价值贡献的方程。这些项可称为二阶项或（输出）变量。
3. 包含三个输入变量，是三个变量之间交互作用之价值贡献的方程。在本文中，这些项可称为三阶项或（输出）变量。

于是，MARS 的输出变量即为价值贡献，包括残差价值贡献。它们又在销售网格计算中被作为变量使用，例如计算调整或绘制图形。

我们可以有落入两种基本本体的不同类型变量：（1）变量来源；（2）变量应用，即非残差与残差

8.1 变量来源本体

1. 一阶变量

- 1.1 MLS 回归变量
- 1.2 外部回归变量
- 1.3 派生回归变量

2. 因子变量：

为了处理因子变量，需要一个能够识别它们的函数。由于它们是在 Excel 项目配置工作簿中指定的，因此创建了一个函数：当传入的变量已被归类为因子类型时，该函数返回 true。

3. 交互变量：如果阶数设置为 2 或更高，这些变量由 MARS 在生成模型时创建。然而，由于解释上的困难、过拟合的可能性、对自由度的影响以及为确保模型稳健而随之增加的物业记录数量，最多只建议 3 个变量交互。

3.1 二阶变量：这些很常见，好的例子有 GLA:BathRms、GLA:LotSize、DateOfSale:BathRms、Latitude _Longitude。

3.2 三阶变量：较不常见的交互，提示需要谨慎并可能存在过拟合。

4. 残差分解变量

8.2 变量应用本体

1. MARS 生成的价值贡献

1.1 一阶变量

1.1.1. MLS 回归变量

1.1.2. 外部回归变量

1.1.3. 派生回归变量

1.2 交互变量

1.2.1. 二阶变量

1.2.2. 三阶变量

1.3 残差分解变量

1.4 汇总变量

2. 工程师生成的变量 & 数值

2.1 残差分解变量

2.1.1. 设计/美学

2.1.2. 景观

2.1.3. 功能效用

2.1.4. 视野

2.1.5. 状况

2.1.6. 质量

2.1.7. ...

这些变量组将在下一节中给出其记号并作更详细的描述

规程 3.1: 因子变量

因子变量, 或简称因子, 可出现在所有项中, 包括交互项。它们之所以重要, 是因为在 MARS 回归和绘图中必须对其进行特殊处理。

虽然取字符值的变量必然是因子, 但数值变量必须被标记为因子变量, MARS 才能将其识别为因子。例如, MLS 区域常以数字标识。特别是, 旧金山湾区的邻里区域常以唯一的整数标识。如果这些值作为因子传给 MARS, 那么 MARS 将尝试发现某些区域对其边界内的物业交易是否具有平均贡献价值, 而不考虑其与其他区域的任何关系。相反, 对于普通变量, 它会假设随着变量值的变化, 价值贡献存在某种连续性。这是一个重要的区别。此外, 这些因子变量可能与非因子变量一起出现在模型项中, 这使绘图变得困难。具体而言, 在绘制含因子的单变量项时, 我们将所有此类因子归并在一起, 简单地绘制一张显示其正负贡献的条形图。

如果因子与其他变量一起出现在某个项中, 它们总是以一个可能的取值出现, 因为因子变量对给定物业只能有一个值: 如果某区域的值为 662, 根据定义, 该区域不可能有任何其他值。如果你觉得这令人困惑, 那么更糟的是, 在 MARS 生成的模型中, 因子并没有单一的变量名; MARS 会取输入时给它的因子变量的词根名, 然后通过附加各可能取值来构造新的变量名并创建新列。于是, Excel 电子表格中的单个列 (例如 MlsArea) 会为每个可能的取值创建多个数据框列, 例如 MlsArea660、MlsArea661 等。诸如 MlsArea661 这样的变量, 对位于 661 区域的物业取值为 1, 否则为 0。然后, MARS 将把这些列当作布尔变量进行回归。MARS 提供的模型中将使用这些修改后的名称。

你很可能希望将这些模型名称转换回其原始名称, 使图形对喜欢简单明了 (至少尽可能简单) 的客户更易理解。我们看到的将不是 MlsArea661=1, 而是 MlsArea=661。然后, 我们只需在图形标题中注明诸如“对于 MLS 区域 = 662”之类的文字, 即可绘制各项。这样, 含 M 个因子变量的 N 阶项将作为 (N-M) 阶项绘制, 因子维度则消失于标题之中。注意, 单个项中可能存在两个或更多因子变量, 因此你很可能需要为每一对可能的取值组合绘制一张图。如果具有三个不同取值的变量 x 与具有两个不同取值的变量 y 同非因子变量 z 一起出现在同一项中, 那么你将需要 $2 \cdot 3 = 6$ 张图。一个例子是“MlsArea”+“Architectural Style”, 例如旧金山的维多利亚式住宅, 其价值可能因其所处的旧金山区域不同而或高或低。幸运的是, 在大多数区域, 通常只有 1-3 种建筑风格作为价值贡献者重要到足以用于模型 (尽管可能存在 6 种以上不同的建筑风格)。

8.3 二阶与三阶项

在二阶或三阶回归中, 输出模型中的某些项可能分别包含两个或三个变量。例如:

$$\text{BathRms} ___ \text{GLA} = -52 * \max(0, 1200 - \text{GLA}) * \max(0, \text{BathRms} - 1)$$

注意, 我们通过按字母顺序对变量名排序并将其连接 (或许以下划线相连) 来命名这些项。对于模型中产生的变量名三元组, 我们对三阶项也如此处理。例如“BathRms ___ GLA ___ LotSz”。

在第 3 阶段, 当 RCA 程序为每个物业计算所有变量 (包括交互变量) 的贡献时, 这些变量会被加入到主要输入变量之中。

8.3.1 调整变量

RCA 程序还将在第 3 阶段, 即价值贡献计算完成之后, 通过用估价对象某一特征的贡献减去相应可比物业该特征的价值贡献, 来计算所有变量和所有物业的调整。

8.3.2 汇总变量

汇总变量被定义用来取代变量的某些子集，将它们的所有价值贡献和调整加总到一个公共变量中，以便在报告中使用。例如，我们可能已将已完工和未完工居住面积都拆分为更细的变量：

1. 地上合法已完工面积
2. 地下合法已完工面积
3. 地上非法已完工面积
4. 地下非法已完工面积
5. 地上未完工面积
6. 地下未完工面积

是的，确实存在这样做有意义的地区。

MARS 回归将对所有这些变量进行回归，并生成为它们提供贡献和调整的模型。但对于表格式报告而言，这些不同的贡献可能必须合并或汇总为更少的变量。汇总的细节应在附录中提供。一个更常见的例子是将交互调整汇总到某个主要变量的调整中。例如，我们可以将 GLA_Lotsize 的调整汇总到 GLA 或 LotSize，甚至按 50:50 在两者之间拆分，具体由汇总工作表决定。

8.3.3 特殊变量 & 记号

π_i	物业 $i = 1, 2, \dots, n_p$ 的实际净售价；对于估价对象， π_0 表示价值结论（见证明 I）
α_i	物业 $i = 1, 2, \dots, n_p$ 的调整后净售价
ρ_i	物业 $i = 0, 1, \dots, n_p$ 的 MARS 估计净售价
β	模型的基准值
$\varphi_{i,j}$	物业 $i = 0, 1, \dots, n_p$ 的回归变量 $j = 1, 2, \dots, n_r$ 的价值贡献
$\delta_{i,j}$	物业 $i = 1, 2, \dots, n_p$ 的回归变量 $j = 1, 2, \dots, n_r$ 的调整
ϵ_i	物业 $i = 0, 1, \dots, n_p$ 的总残差
$\epsilon_{i,c}$	物业 $i = 0, 1, \dots, n_p$ 的残差分量价值贡献 c ， $c = 1, 2, \dots, n_c$
$\xi_{i,c}$	物业 $i = 0, 1, \dots, n_p$ 的残差分量 c 的调整， $c = 1, 2, \dots, n_c$

其中

n_p = 可比物业的数量

n_r = 回归变量的数量

n_c = 残差分解变量的数量

8.3.4 分解残差

一旦 MARS 回归创建了模型，它便会根据 MARS 选定的输入变量生成估计售价。它用相应的净售价减去该值，得到每个物业的残差值。此时，RCA 规程规定：一旦 VE 确定了最终的可比物业——通常为 6-12 条物业挂牌信息——他需要逐一查看每条物业挂牌信息，并在忽略随机数据误差可能性的前提下，将计算出的残差分解为他认为最可能影响价值的特征（潜在变量）。例如，对于某一物业，他可能有以下分解变量：

1. 设计
2. 功能效用
3. 高于/低于市场的成交
4. 状况
5. 质量

不同物业很可能会有一些不同的残差分解变量，需要将它们收集到一个有序集合中。

估计稀有特征的贡献。有些承载价值的特征对回归通道而言过于稀有：在 780 笔交易中仅出现 3 次的工作间无法获得站得住脚的系数，前向过程通常会完全跳过这样的变量。这些就是残差特征，其贡献可以在残差空间中估计：先拟合不含该特征的模型，然后用普通最小二乘法将模型残差对被排除的列进行回归。对于二元标志，系数即为一次性总额贡献（等价地，即被标记交易的平均残差偏移）；对于数值列，则是单位贡献。当若干相关设施被联合估计时（例如附属建筑面积与工作间标志一起估计），回归会在它们之间分配贡献，而不是在两行分解中重复计算同一笔金额。由不足约十个观测支持的估计仅具指示性，应通过折旧成本或扩大后的样本（例如从公开备注中提取的标志）予以佐证；残差约束本身会对最终赋予的价值设置上限。earthUI (v0.11+) 以 *Residual Contribution* 选项卡的形式直接实现了这一方法，包括联合估计、逐笔交易的残差证据以及用于工作档案的记录说明。

记号：我们使用上表中的记号，得到：

$$\epsilon_i = \epsilon_{i,1} + \epsilon_{i,2} + \dots + \epsilon_{i,n_c} \quad (11)$$

其中 n_c 是残差分解分量的数量。

VE 必须运用其判断将残差分解为给定的分量。有人可能认为这会让偏见进入最终的价值结论。然而事实并非如此，因为我们有数学证明：只要各分解调整的总和等于残差，它们就不会影响价值结论。分解的目的不是确立价值，而是解释残差相对于其他物业交易的残差为何是这个数值。工程师的偏见唯一可能影响最终价值结论的地方，是对估价对象残差的估计，因为估价对象当然没有实际售价可用来为其计算残差。

每个物业都有其残差价值贡献的集合，但这些集合需要合并为覆盖所有物业的一个集合 C：

$$\begin{aligned} \text{设: } A_i &= \{\{\epsilon_{i,c}\}_{c=1}^{n_c}\} \\ C &= \{x \mid x \in \bigcup_{i=0}^{n_p} A_i\} \end{aligned}$$

其中 $[\epsilon_{i,c}]$ 是值 $\epsilon_{i,c}$ 的变量名， A_i 是此类变量的集合， n_c 是每个 ϵ_i 中值的数量。

因此，当估价对象与可比物业的数据分列于不同列中时，在每个物业的所有回归模型变量价值贡献之下，我们应看到由 n_c 个相应残差价值贡献构成的集合 C。

9 RCA 核心证明

设：

$$\begin{aligned} n_p &= \# \text{ of properties} \\ n_r &= \# \text{ of regression variables} \\ n_c &= \# \text{ of residual component variables} \end{aligned}$$

证明 I： 每一个可比交易实例的调整后售价均等于估价对象的估计售价。

对于任意可比交易实例 i ，调整后售价 α_i 等于净售价 π_i 加上各项调整：即估价对象与可比交易实例在基础价值、回归变量 $j = 1, 2, \dots, n_r$ 以及残差上的价值贡献之差。对于没有实际成交的估价对象，其净售价 π_0 定义为价值结论 $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$ 。于是：

$$\begin{aligned} \alpha_i &= \pi_i + (\beta - \beta) + (\varphi_{0,1} - \varphi_{i,1}) \\ &\quad + (\varphi_{0,2} - \varphi_{i,2}) + \dots + \\ &\quad (\varphi_{0,n_r} - \varphi_{i,n_r}) \\ &\quad + (\epsilon_0 - \epsilon_i) \\ &= \pi_i + (\beta + \varphi_{0,1} + \varphi_{0,2} + \dots \\ &\quad + \varphi_{0,n_r} + \epsilon_0) \\ &\quad - (\beta + \varphi_{i,1} + \varphi_{i,2} + \dots + \\ &\quad \varphi_{i,n_r} + \epsilon_i) \\ &= \pi_i + \pi_0 - \pi_i \\ &= \pi_0 \end{aligned} \tag{12}$$

证明 II： 在所有残差分量价值贡献之和等于可比交易实例残差这一约束下，改变残差的分解方式对 RCA 规程的最终价值结论没有任何影响。

这几乎是显而易见的。然而若不认真思考其中的数学，就很容易忽视这一点，因为传统估价师关注的是调整而非价值贡献。在这里相互抵消的正是价值贡献——这是一个持续了数十年的疏漏。对于某个可比交易实例 k ($k > 0$)，总残差调整等于 n_c 个分解残差分量之和，而后者又等于各价值贡献之差的和，即等于估价对象总残差减去可比交易实例总残差。这与所创建的残差分量的数量或其所代表的内容无关。只要残差分量的价值之和对估价对象和所选可比交易实例都等于总残差， ξ_k 就不会改变。

$$\begin{aligned} \xi_k &= \xi_{k,1} + \xi_{k,2} + \dots + \xi_{k,n_c} \\ &= (\epsilon_{0,1} - \epsilon_{k,1}) + (\epsilon_{0,2} - \epsilon_{k,2}) + \dots + \\ &\quad (\epsilon_{0,n_c} - \epsilon_{k,n_c}) \\ &= (\epsilon_{0,1} + \epsilon_{0,2} + \dots + \epsilon_{0,n_c}) - \\ &\quad (\epsilon_{k,1} + \epsilon_{k,2} + \dots + \epsilon_{k,n_c}) \\ &= \epsilon_0 - \epsilon_k \end{aligned} \tag{13}$$

10 MARS 模型

10.1 模型变量与交互作用的“ g ”函数

RCA 方法的一项核心原则是“解释性推理”至关重要。遗憾的是，为各个价值贡献方程提供有意义的图形——尤其是在使用交互作用和因子变量时——可能颇具挑战。因此，为绘制这些图形制定一套清晰而严格的规程被视为势在必行。本文在此给出这一规程，但请理解，该规程面向的是编写 R 或 Python 代码以生成图形的开发者，对于不参与编码的读者而言，可能仅有粗浅的参考意义。

MARS 生成的价格模型是一个基础常数加上若干附加项之和，每一项涉及 1 到 3 个变量。为了对其进行分析，我们首先将模型拆分为单个的项。然后，将共享同一变量集合、代表单一交互作用的多个项合并为单个交互作用表达式。假设回归的最大阶数为 3。考虑一个三阶回归模型，它由一个基础常数、 n_1 个一阶表达式、 n_2 个二阶交互作用表达式和 n_3 个三阶交互作用表达式组成。这样我们便得到四组不同的表达式，它们或为单变量表达式，或为多变量交互作用，其绘图方式取决于两个因素：（1）它们是否包含因子变量（即分类变量），以及（2）它们所含唯一变量的数量。具体而言：

- 一阶表达式（1 个变量）以二维图形绘制。
- 二阶表达式（2 个变量）以三维图形绘制。
- 三阶表达式（3 个变量）以一组三维图形绘制。
- 若表达式包含因子变量，则每个唯一因子会使图形的维度降低一维（例如，若其中一个变量是因子，则三维图形变为二维图形）。

每个函数 g 的索引提供了确定其绘图方式所需的信息，反映了唯一变量的数量以及是否存在因子变量。

10.2 g 函数的索引

每个函数 g 的索引形式为 $f_{j,k}g_k^j$ ，其中：

- $f_{j,k}$ 是第一个索引（置于 g 之前的上标）：
 - j 的取值范围为 0 到 3，用于标识组别：
 - * $j = 0$ ：基础常数（是通常索引规则的一个例外）。
 - * $j = 1$ ：一阶表达式。
 - * $j = 2$ ：二阶表达式。
 - * $j = 3$ ：三阶表达式。
 - k 是某一具体函数在其所属组内的索引（例如， $k = 1$ 表示组 j 中的第一个函数）。
 - $f_{j,k}$ 的值表示该函数中唯一因子变量的数量：
 - * $f_{j,k} = 0$ ：不含因子变量。
 - * $f_{j,k} > 0$ ：因子变量的数量（例如 1、2 或 3）。
- g 的后置索引：
 - g 之后的上标 j （例如 g^j ）重复标注组号，以求清晰。
 - g 之后的下标 k （例如 g_k ）与 $f_{j,k}$ 中的 k 一致，标识该函数在其组内的位置。

虽然 f 与 g 的索引在数值上一致，但它们的作用不同： $f_{j,k}$ 计数因子变量，这对绘图决策至关重要；而 g 上的 j 和 k 则定义其组别和位置。

10.3 绘图维度

函数 $f_{j,k} g_k^j$ 的图形维度 d 按下式计算:

$$d = j - f_{j,k} \quad (14)$$

其中:

- j 是唯一变量的数量 (即组号, 不考虑常数的情形)。
- $f_{j,k}$ 是因子变量的数量。
- d 是需要绘图的连续 (非因子) 变量的数量, 最大为 3 (因为最大阶数为 3)。含有 d 个连续变量的函数需在 $d+1$ 维中绘制: $d=1$ 对应二维图形, $d=2$ 对应三维图形 (或二维热力图), $d=3$ 则对应一组三维图形。

例如:

- 若 $j=3$ (3 个变量) 且 $f_{j,k}=0$ (无因子变量), 则 $d=3-0=3$, 需要一组三维图形, 每张图固定其中一个变量。
- 若 $j=3$ 且 $f_{j,k}=1$ (一个因子变量), 则 $d=3-1=2$, 需要针对模型中出现的每个因子变量取值绘制一张三维图形 (或二维热力图)。
- 若 $j=0$ (常数), 通常无需绘图, 因为它只是一个单一数值。

当 $d=3$ 时 (例如不含因子变量的三阶表达式), 要将完整函数可视化, 可能需要多张三维图形, 每张图将某一变量固定在不同取值上, 以展示函数行为的一个子集。

10.4 最终模型方程

MARS 模型被拆分为四组表达式 (常数、一阶、二阶和三阶)。索引 $f_{j,k} g_k^j$ 告诉我们:

- j : 组别及变量数量。
- $f_{j,k}$: 因子变量的数量。
- $d = j - f_{j,k}$: 图形维度。

这一结构指导每个 g 函数的可视化方式, 确保绘图方法与表达式的复杂度和变量类型相匹配。

MARS 模型给出的售价估计值 $\hat{P}$ 可以表示为:

$$\begin{aligned} \hat{P} = & f_{0,1} g_1^0 + \sum_{q=1}^{n_1} f_{1,q} g_q^1(x_*) \\ & + \sum_{r=1}^{n_2} f_{2,r} g_r^2(x_*, y_*) \\ & + \sum_{s=1}^{n_3} f_{3,s} g_s^3(x_*, y_*, z_*) \end{aligned} \quad (15)$$

$n_1 = \#$ of first degree terms

$n_2 = \#$ of second degree terms

$n_3 = \#$ of third degree terms

10.5 绘制一个模型所需的图形数量

要将模型中的所有函数全部可视化，需要多少张图？答案取决于变量本身、它们之间的交互作用，以及它们是因子变量还是非因子变量。

估价工程师可以在供 R/earth 使用的 Excel 配置文件中限制变量和交互作用，从而影响哪些交互作用被允许。一种务实的做法是让 R/earth 评估所有可能的交互作用，并选出对售价估计有显著贡献的那些，同时检查过拟合或共线性等问题。某些交互作用可能被禁用，这会影响更高阶的交互作用（例如，若三阶交互作用中的某一变量对被排除，则整个组合都被排除）。

对于因子变量（例如具有 12 个取值的 MLS Area 与 GLA 交互），绘图会变得更为复杂。这可能需要 12 张图——每个区域一张。若另一个因子（如具有 6 个取值的 Architectural Style）再与 MLS Area 和 GLA 交互，则可能需要多达 $12 \cdot 6 = 60$ 张图。不过，R/earth 很少会认定所有组合都显著；通常只有少数关键交互作用（例如特定的区域和风格）值得绘图，而模型会明确指出这些组合。

以下是所需图形数量的简明指南：

1. 单变量函数（非因子）：1 张图。
2. 单变量函数（因子）：1 张将所有因子取值合并展示的条形图。
3. 基础函数：无需绘图。
4. 二阶交互作用（无因子）：1 张图（例如以颜色表示售价的二维热力图）。
5. 二阶交互作用（1 个因子）：模型中每个因子取值 1 张图（例如 2 个取值 = 2 张图）。
6. 二阶交互作用（2 个因子）：1 张以因子组合构成矩阵的图（例如 3 个区域 $\times$ 2 种风格 = 6 个单元格）。
7. 三阶交互作用（无因子）：多张图（例如按 Bath 数量 = 1、2、3、4 绘制 4 张图）。
8. 三阶交互作用（1 个因子）：模型中每个因子取值 1 张图。
9. 三阶交互作用（2 个因子）：模型中每种因子取值组合 1 张图。
10. 三阶交互作用（3 个因子）：以不同取值最少的那个因子的每个取值各绘 1 张图，其余两个因子用颜色矩阵表示。

更复杂的情形可能需要更高级的绘图技术。

此外还存在其他可能性，通常需要更为精细的图形。

由 MARS 生成的完整模型示例见 [图 1](#)。

11 数据特征

以下一节讨论 MARS。MARS 完全依赖于它所接收的数据，而这些数据大多来自某个 MLS（多重挂牌服务）或税务评估机构。不过，数据也可能经由其他数据提供商中转，这些提供商维护并改进数据，然后以一定利润出售。在某些地区，数据相当可靠；在另一些地区则不那么可靠。在关键情形下，房产查勘人员必须亲自进行观察和测量，而其准确性也可能受到质疑。

估价工程师需要理解数据代表什么，以及数据在多大程度上如实地代表了它本应代表的内容。这引出了一系列问题：

- 如何将一个变量归类为测量值？
- 如果一个变量不是测量值，它还能是什么？
- 如果回归所接收的数据不准确，回归又如何能有用？

我们从 MLS 挂牌服务和税务评估机构获得的用于回归的数据，大多是隐含测量值或计数的数值型数据。其中一部分是逻辑型数据，即真或假。

11.1 残差的用途

残差是对未纳入回归模型的变量所增添价值的一种间接量度。用特征价格文献的语言来说，这些未测量的特征就是潜在变量：它们影响成交价格但未被直接测量，多数情况下是因为没有切实可行的方法对其进行客观测量。它们的取值只能通过推断得到，而在实践中，潜在变量的取值往往通过主观判断来估计——在 RCA 中，这种判断以残差排序为锚。通过按残差或每平方英尺残差对房产进行排序，我们得到了一个从吸引力最低到吸引力最高的客观排名。只要由一位对当地市场有深入理解的熟练分析师构建多元自适应回归样条（MARS）模型，这种排序方法就是有效的。以下是一些附加要求：

- 数据与模型的质量：构建良好的回归模型和高质量的数据是获得可靠残差的前提。在旧金山湾区，我的目标是建立一个 R^2 约为 75-80%、交叉验证 R^2 (CV R^2) 为 55-70% 的模型。
- 避免过拟合、限制 R^2 ：在像旧金山湾区这样房价较高地区的多数成熟市场中，估价工程师应以 R^2 最高约 80% 为目标，因为住宅房产竞争性价值的很大一部分在于其主观吸引力，这种吸引力无法量化，因而不能直接纳入回归模型。 R^2 显著高于 80% 可能意味着模型过拟合，需要进一步审查。但需要提醒的是：有些市场区域的 R^2 值高于 90% 而并未过拟合；也有一些难以估价的地区，能达到的最好水平可能只有 50-60% 甚至更低。
- 特定地区的高 R^2 ：不过，在某些区域，例如分块开发的小区或其他同质化的住宅区，由于房产的同质性，较高的 R^2 值出人意料地常见。

11.2 作为近似回归变量的状况与质量评级

房利美的状况与质量评级量表 (C1-C6 和 Q1-Q6) 值得特别说明。这些量表过于粗糙，不能被当作真正的测量特征；而且由于它们由估价师给出，实践中属于主观评级——本身就是潜在变量。明确区分两个层次是有益的：估价师用作回归变量的粗糙评级，即粗粒度状况 (condition-1) 和粗粒度质量 (quality-1)；相对地，从残差中恢复的细粒度潜在变量，即细粒度状况 (condition-2) 和细粒度质量 (quality-2)。VE 可以将估价师的粗粒度评级作为近似值接受并对其进行回归，从而得到近似的价值贡献，这些价值贡献可作为细粒度潜在变量的初始赋值。在采用这种做法时，规程是从模型估计值中减去这两个粗粒度变量经回归估计的价值贡献——把它们从销售网格回归部分下的已测量属性中移出——并将其加回到残差之上。此后，再从残差中移除这部分并将其分配给相应的细粒度潜在变量 `res_condition1` 和 `res_quality1`。随后利用从 MLS 公开

备注 (public_remarks) 及相关字符串列表中提取的附加信息, 对这些潜在变量的取值进行微调。最终结果是, 细粒度价值贡献取代粗粒度价值贡献: 粗糙评级只是近似模型的脚手架, 最终报告的状况与质量取值是从残差推导出的细粒度取值。

11.3 数据错误

根据社交媒体上的评论, 加利福尼亚州的多重挂牌服务 (MLS) 数据总体上比美国许多其他地区质量更高。通常, 在大多数成熟的大都市地区, MLS 数据的质量应当是可以预期的良好, 而在农村地区则较差。

尽管数据中因各种原因偶有错误, 但只要数据量足够, 这些错误一般应相互抵消, 对价格模型所用平均值的影响很小。

然而, 特定地区可能出现系统性偏差。例如, 考虑这样一个社区: 许多住宅最初建有完整的二层, 但部分买家可以选择拆除二楼房间, 为一层的一个或多个房间打造挑高拱顶天花板, 从而显著减少总居住面积 (GLA)。由于种种原因, 这些改动后来没有在评估机构的记录中更新。结果, MLS 可能显示 2,400 平方英尺, 而实测值只有 2,000 平方英尺。在这类情形下, 应在执行多元自适应回归样条 (MARS) 回归之前先更正 MLS 数据。需要注意的是, 为可比交易实例获取这类信息可能需要一些调查工作。发现的差异应予以记录以备将来参考。

综上所述, 可以合理假设, 在 MLS 数据质量较高的地区, 溢入残差的错误与设计、状况、质量等主观且未测量特征的影响相比是微不足道的。在估计 RCA 方法的准确性时, 这是一个重要的考虑因素。例如, 如果某个市场区域的 R^2 值为 80%, 那么价格中约有 20% 归因于 CQA 残差。因此, 在残差排序中定位估价对象时 $\pm 10\%$ 的误差, 传导到最终价值结论上就是 $\pm 10\% \cdot 20\% = \pm 2\%$ 的误差。这就是提高准确性的途径! 在许多情况下, 可以得到误差远低于 $\pm 1\%$ 的准确价值估计。

因此, 在 MARS 回归中把残差当作未建模变量集体价值的间接指标是合理的, 前提是估价工程师具备足够的当地市场经验, 能够识别系统性测量偏差存在于何处以及为何存在。这种方法确保分析准确反映实际房产价值, 同时也承认房地产数据的局限性与细微之处。

实例演算: 在 Sea Ranch 测量未被测量的价值驱动因素

当某个市场区域可实现的 R^2 低于该市场表面同质性所应达到的水平时, 本框架的应对方式是识别并测量缺失的变量, 而不是接受这一局限。索诺马县的 Sea Ranch 社区直接说明了这一点。

这是“测量而非接受”式应对的一个具体实例。索诺马县的 Sea Ranch 社区是一个沿海开发区, 到海洋的距离是主要的价值驱动因素: 在其他特征相同的情况下, 距离悬崖几百英尺的住宅与向内陆四分之一英里的住宅, 价格差异巨大。该社区的 MLS 数据并未把到海距离记录为变量。基于现有 MLS 变量拟合的初始 MARS 模型, 其解释力低于该市场表面同质性所应支持的水平, 且残差结构未能按数据中任何可见特征充分地对房产排序——这一信号表明模型缺失了一个重要的驱动因素, 而不是该市场本身噪声很大。

补救办法是直接测量。作者获取了一张显示房产相对海岸线布局的社区地图, 并为数百套住宅测量了到海距离。将测得的距离作为变量加入 MLS 数据集, 并重新运行 MARS 回归。新变量被模型选为显著变量, 解释力大幅提升, 随后的残差排序按与其余主观特征——景观、状况、质量——相一致的模式对房产排序, 而不再受此前污染残差的潜在到海距离信号的影响。在 MLS 变量缺失或不可靠的山坡社区, 高程问题也以类似方式得到了处理。

这一方法论原则具有普遍性。当可实现的 R^2 低于该市场表面同质性所应支持的水平时, 首要假设是模型缺失了一个重要变量, 恰当的应对是弄清它是什么并直接收集测量数据。数据来源包括社区地图、GIS 数据、税务评估记录、建筑管理部门记录、卫星影像以及实地查勘。这类测量工作的成本是实实在在的, 但有限度; 一旦为某个市场区域收集了该变量, 就可以在该区域的后续估价任务中沿用。RCA 能够顺畅地吸纳新变量, 因为该框架并不在意哪些变量进入

MARS 回归——残差结构会告诉 VE 新变量是否捕捉到了真实信号，各项价值贡献也会相应调整。

另一种应对方式——接受较低的 R^2 并报告更宽的误差区间——在缺失的驱动因素确实无法测量（买方特有的偏好、短暂的市场情绪、交易的特殊情形）时是站得住脚的，但当这些因素只是尚未收集时则不然。面对一个弱模型，VE 的第一步应当是追问缺失的信号是否属于后者。

11.4 什么是准确性？

完美准确的估价要求：

1. 符合市场价值：估价对象的假设性销售交易 100% 符合给定的市场价值定义。
2. 准确而稳健的模型开发：建立一个高质量的 MARS 模型，视地区而定，其 R^2 约为 75-80%， CVR^2 约为 55-70%。
3. 良好的数据错误分析：已测量变量所有可能的数据错误都得到理解和控制，几乎不溢入残差。
4. 系统性偏差核算：数据中任何系统性偏差都应予以核算和调整。
5. 残差排名分析：确保按残差排名的房产从最低到最高排名呈现出合理的特征模式。遗产出售、亏本急售和拍卖出售等异常值应予以调查、解释，并在必要时剔除。
6. 估价对象定位：在透彻理解残差排序的基础上，应准确定位估价对象。找出一处吸引力明显更高的房产和一处吸引力明显更低的房产，将估价对象置于两者之间并打分。然后尽量收窄较高与较低房产之间的排名区间。
7. 解释性推理：绝不追求准确性而牺牲解释性推理。估价师必须能够解释估价对象与可比房产之间的价值差异，至少要让一个具有正常理解能力的人明白。例如，一个 R^2 和 CVR^2 达到最优的良好 MARS 模型可用于解释已测量特征的价值贡献，而估价工程师对残差的分解则可用于解释其他特征的价值贡献。
8. 过拟合：避免模型过拟合，不使用那些单独或组合起来能够识别一处或数处房产的房产特征。尤其要避免使用其取值能够锁定少数房产的因子变量。在规模足够大的数据集中，用于标识社区的因子变量是合适的，但在游泳池极少的地区标识带游泳池房产的因子变量则必须慎用。标识唯一房产的特征，例如精确的 GIS 坐标、宗地号或地址，是绝对不允许的。

如果回归的 R^2 为 80%，我们预计估价对象价值中约有 20% 归因于残差。当然，这只是近似值。从 CQA—残差曲线可知，残差分布的大部分集中在 CQA 的低端和高端两个极值区。基于 CQA—残差曲线以及低、高 CQA 值处的极端斜率，我们可能预期在这些区域对残差的估计不那么准确。然而，这一点被这些区域房产的相对稀少以及相邻房产之间吸引力差异更大这两个因素所抵消。因此，尽管在中间区段我们谈论的影响范围可能是 $\pm 5\%$ ，在两端我们更应关注的则是略窄的 $\pm 2 - \pm 3\%$ 范围。

一般而言，由于 CQA—残差曲线的非线性，计算 RCA 的准确性有一定难度。但只要你有一个合理的残差排序，就可以逐案计算。CQA 分值的范围是 0.0 到 10.0。有了良好的 MARS 模型，你的残差排序应呈现出从吸引力最低到最高的清晰模式，你应该不难把评分控制在总范围的 $\pm 5\%$ 之内，即 CQA 分值的 ± 0.5 之内。如果回归中有 100 处房产，这相当于把估价对象准确地排入一个 10 处房产的组内。那么，参照表 3 中的伯灵格姆销售数据，让我们分别针对 VE 估计的 CQA 为 7.5、5.1 和 1.9 的情形计算近似准确性（或误差），其中他对估价对象的价值结论为 \$3,000,000。注意，由于表 3 只包含所用数百处房产中的一个样本，必要时你需要（假设线性）对残差进行插值。需要明确的是，我们假定 VE 在每种情形下都非常确信，他对估价对象的排名在仅含 100 处房产的残差排序中处于 ± 5 处房产的范围之内。

CQA= 7.5 ±0.5					
CQA	残差	绝对差	预期误差	假设成交价	误差百分比
8.00	\$257,112				
7.50	\$197,113	\$59,999			
7.00	\$154,615	\$42,498	±\$51,249	\$3,000,000	±1.71%

CQA=5.1 ±0.5					
CQA	残差	绝对差	预期误差	假设成交价	误差百分比
5.60	\$50,911				
5.10	\$19,559	\$31,352			
4.60	-\$28,207	\$47,766	±\$39,559	\$3,000,000	±1.32%

CQA= 1.9 ±0.5					
CQA	残差	绝对差	预期误差	假设成交价	误差百分比
2.40	-\$228,411				
1.90	-\$276,381	\$47,970			
1.40	-\$326,938	\$50,557	±\$49,264	\$3,000,000	±1.64%

最后，验证准确性的唯一客观方式，是建立一套严格的规程（本文只是粗线条地部分提及）并核查其遵守情况。之所以必要，是因为受市场不完善的影响，实际成交价格可能偏离市场价值。将估价与其后的成交进行比较，可以在一定程度上反映准确性。然而，房地产市场的复杂性和不稳定性，要求由估价师或高度合格的估价工程师依据既定标准和规程给出价值估计。这些标准和规程应当足够严格，以确保胜任的估价工程师能够各自独立地得出几乎相同的价值结论。在大多数能够获得良好数据的情况下，这应当是一个可以实现的目标。

11.5 作为异常值的估价对象

既然估计一处房产的市场价值需要分析近期的销售交易，那么就有理由追问：对于一处缺乏现代化改造和维修的较旧的估价对象，近期销售交易能否很好地代表其各项特征。在旧金山湾区尤其如此，那里近期成交的住宅通常都经过翻新，以提高其售价。因此，可比销售实例的平均质量和状况通常优于那些很久没有进入市场的老旧住宅。这一价值差异应当由对估价对象残差的估计——换言之，由其 CQA 分值——来捕捉。就估价对象估值中的数据质量与准确性而言，这是一个重要的考虑因素。

11.6 CVR2 在应对意外数据方面的重要性

MARS 回归基于对随机选取的房产子集多次迭代运行 R/earth，该子集通常占 MLS 数据集的 80% 到 90%，称为训练集。其余 10% - 20% 的房产作为测试集。在这一阶段，先由训练数据集建立 MARS 模型，再用它预测测试数据集中房产的成交价格，然后将预测值与实际成交价格进行比较。这一方法已被数据挖掘者公认为构建稳健模型、预测尚未观察到的房产成交价格的最有效策略。

然而，CVR2 指标的主要用途并不在于反映回归模型对实际市场区域成交的事实准确性，而在于优化模型对以指定生效日期为准的假设性房产交易的预测能力。这些房产不在训练数据集中，可能表现出训练集和测试集房产数据中都缺失的新的或意外的特征组合。

在这一背景下，假设一处过去 15 年从未翻新的房产在生效估价日期前一个月挂牌，我们如何估计其成交价格？一个 CVR2 表现强劲的模型，比一个未经过这种反复交叉验证的模型更有可能给出最准确的估计。同样至关重要的是，数据集中要包含足够数量、翻新、维修和现代化改造程度各异的可比房产。

归根结底，如果我们拥有一个稳健的模型，能够准确预测可比房产数据集中各项房产特征的价值贡献，我们就能有效地把新的估价对象放入模型按残差对这些房产的排序之中。这一定位使我们能够为该房产赋予恰当的 CQA 分值和残差。

12 残差

正如“残差约束法”这一名称所示，残差是本方法论的核心。此处的残差是指房产的净成交价格与回归分析所预测的成交价格之间的差额。

正残差表明，基于建筑面积（GLA）、地块面积等典型可测量属性，该房产的成交价高于其估计价格，暗示其吸引力可能高于平均水平。负残差则表明该房产的成交价偏低，可能意味着吸引力较低或存在其他未被捕捉到的负面属性。

考虑到较大的住宅通常具有较大的残差，我们可以使用基于 GLA 的每平方英尺残差进行更规范化的比较。关于这一问题的更深入讨论将留待另一篇论文。

12.1 排序与评分

在为已测量变量建立 MARS 模型之后，为每一处房产计算残差。这些残差表示观测到的净成交价格与模型预测的成交价格之间的差额。还应额外建立一系列“残差/平方英尺”，即残差除以住宅的居住面积——这是基于这样的认识：一般而言，具体房产的残差会随居住面积的大小而增大。其他可能的统计量在某些社区中或许效果更好，但那是另一个话题。

假设 7：“按残差排序”

在本文中，凡提及“按残差排序”，均应理解为按残差排序、按残差/平方英尺排序，或按估价工程师认为最适合该市场区域的其他残差函数排序。作者在旧金山湾区的经验表明，残差/平方英尺通常最为可靠，尤其是在处理 GLA 异常小或异常大的估价对象时。

接下来，将各房产按其残差从最负到最正进行排序。然后为每一处房产生成一个“状况—质量—吸引力”（CQA）评分，取值范围为 0.0 至 10.0。该评分的计算方法如下：

CQA 评分对应于某一房产的残差在所有房产中的百分位排名。具体而言，评分的计算方法是：确定残差低于该房产的房产所占的百分比，再将该百分比除以 10。例如，若某一房产的残差高于 50% 的房产的残差，则其评分为 5.0。

由于四舍五入的原因，可能有一处房产获得 10.0 分，即残差最高的那处房产。然而，10.0 的满分在理论上是不可能的，因为从逻辑上讲，没有任何房产的残差能够高于 100% 的房产（包括其自身）。另一方面，最低分 0.0 是可能出现的，因为至少有一处房产不存在残差更低的其他房产。请注意，当比较的房产数量很多时，由于四舍五入，可能不止一处房产得到 0.0 分。

因此，CQA 评分提供了一个直观的度量，用以衡量某一房产的属性与模型预期的吻合程度——评分越高，表明与模型预测的吻合度越好。

必须承认，只有当销售条件与其市场价值相称时，CQA 评分才可被视为可比房产吸引力的间接指标。估价工程师有责任通过对销售交易进行全面解释和彻底调查，来查明残差价值中是否包含高于或低于市场的因素。虽然显著高于或低于市场的销售并不常见，但在强制出售、遗嘱认证、当事人无行为能力、串通等情形下仍可能发生。

表 3 展示了 MLS 数据的一个半真实节选，连同残差和 CQA 评分（包括总额和每平方英尺两种形式）。本研究分析的房产总数约为 600 处。值得注意的是，若干参与回归的变量并未显示。换言之，所展示的残差是在 MARS 回归已考虑了成交日期等其他变量的影响之后得出的。

表 3: 按残差排序的可比交易实例

成交价格	估计成交价	残差	CQA	残差/平方英尺	CQA SF	GLA	成交日期
\$3,450,000	\$2,318,865	\$1,131,135	10	\$420.50	9.98	2,690	2021-04
\$4,360,757	\$3,399,922	\$960,835	9.9	\$290.28	9.82	3,310	2021-09
\$4,000,000	\$3,134,444	\$865,556	9.8	\$234.57	9.60	3,690	2020-11
\$5,850,000	\$5,150,966	\$699,034	9.7	\$109.91	8.35	6,360	2021-12
\$3,600,000	\$3,023,366	\$576,634	9.5	\$269.46	9.79	2,140	2022-06
\$4,425,000	\$3,987,873	\$437,127	9.2	\$127.44	8.68	3,430	2021-08
\$2,530,000	\$2,164,564	\$365,436	8.9	\$180.02	9.28	2,030	2019-12
\$4,250,000	\$3,884,719	\$365,281	8.8	\$87.18	7.91	4,190	2021-08
\$3,350,000	\$3,034,613	\$315,387	8.5	\$130.87	8.75	2,410	2021-09
\$3,800,000	\$3,522,028	\$277,972	8.2	\$77.43	7.72	3,590	2020-09
\$4,200,000	\$3,963,748	\$236,252	7.8	\$56.12	6.89	4,210	2021-01
\$2,900,000	\$2,702,887	\$197,113	7.5	\$89.19	7.92	2,210	2021-07
\$4,250,000	\$4,103,885	\$146,115	6.9	\$39.38	6.29	3,710	2022-08
\$5,000,000	\$4,864,958	\$135,042	6.8	\$28.25	5.85	4,780	2021-07
\$4,320,000	\$4,199,672	\$120,328	6.5	\$33.90	6.04	3,550	2022-07
\$2,600,000	\$2,479,904	\$120,096	6.5	\$60.96	7.00	1,970	2021-10
\$3,400,000	\$3,301,146	\$98,854	6.2	\$30.32	5.90	3,260	2021-06
\$2,000,000	\$1,901,987	\$98,013	6.2	\$73.14	7.54	1,340	2019-11
\$3,000,000	\$2,919,718	\$80,282	5.9	\$26.50	5.79	3,030	2020-05
\$6,750,000	\$6,699,089	\$50,911	5.6	\$7.75	5.20	6,570	2020-08
\$5,652,000	\$5,601,570	\$50,430	5.5	\$12.83	5.37	3,930	2022-05
\$2,600,000	\$2,579,082	\$20,918	5.2	\$10.51	5.33	1,990	2022-03
\$1,925,000	\$1,905,441	\$19,559	5.1	\$11.51	5.36	1,700	2021-01
\$3,200,000	\$3,209,500	-\$9,500	4.8	-\$3.21	4.88	2,960	2020-11
\$4,150,000	\$4,187,560	-\$37,560	4.5	-\$10.49	4.68	3,580	2021-04
\$2,700,000	\$2,776,510	-\$76,510	3.9	-\$23.61	4.14	3,240	2021-07
\$3,450,000	\$3,567,668	-\$117,668	3.4	-\$31.05	3.90	3,790	2020-09
\$1,975,000	\$2,109,255	-\$134,255	3.3	-\$55.48	3.24	2,420	2020-02
\$1,950,000	\$2,085,431	-\$135,431	3.2	-\$68.40	2.86	1,980	2021-02
\$3,695,000	\$3,859,154	-\$164,154	3.1	-\$47.04	3.56	3,490	2021-05
\$2,288,000	\$2,455,747	-\$167,747	2.9	-\$56.86	3.19	2,950	2020-06
\$2,635,000	\$2,802,781	-\$167,781	2.9	-\$56.87	3.18	2,950	2020-09
\$2,050,000	\$2,217,794	-\$167,794	2.9	-\$82.66	2.46	2,030	2021-07
\$1,550,000	\$1,731,804	-\$181,804	2.7	-\$130.79	1.32	1,390	2021-05
\$2,800,000	\$3,028,411	-\$228,411	2.4	-\$81.58	2.51	2,800	2021-02
\$3,700,000	\$3,976,381	-\$276,381	1.9	-\$76.99	2.62	3,590	2021-08
\$2,517,500	\$2,834,974	-\$317,474	1.5	-\$120.71	1.48	2,630	2019-11
\$2,550,000	\$2,905,330	-\$355,330	1.1	-\$115.74	1.64	3,070	2019-12
\$1,815,000	\$2,263,944	-\$448,944	0.7	-\$273.75	0.15	1,640	2021-11
\$3,711,000	\$4,683,454	-\$972,454	0	-\$226.68	0.37	4,290	2021-12
\$2,225,000	\$3,332,960	-\$1,107,960	0	-\$370.56	0.03	2,990	2021-08
\$3,550,000	\$4,746,489	-\$1,196,489	0	-\$274.42	0.14	4,360	2022-06

12.2 CQA-残差曲线

将某一社区的 CQA 与残差/平方英尺作图，必然得到图 2 所示的 CQA-残差特征曲线。该图所示为 2022 年前后加州 Burlingame 的数据。将 CQA 评分与残差值作图通常也会观察到类似的曲线；不过，我更倾向于使用残差/平方英尺，因为它能够结合估价对象的建筑面积（GLA）来解读 CQA。将残差/平方英尺乘以 GLA，即可把某一房产的 CQA 转换为相应的残差价值。

估价工程师必须仔细评估房产的规模——具体而言是建筑面积（GLA）——是否显著影响残差的大小。这一评估至关重要，因为它可以决定计算房产价值的最准确方法。例如，高端厨房电器和豪华卫浴设施可以大幅提升住宅的残差价值。然而，在厨房和卫生间的尺寸与布局普遍统一的社区中，这些升级可能与 GLA 的相关性并不强。它们对残差价值的影响，可能更多在于它们所代表的质量，而非它们所占据的空间。

相反，昂贵的外部特征和材料能够显著提升房产的门面吸引力，可能与房产的规模有更直接的关系。较大的住宅可以容纳更精致的外部设施，例如大规模的园林绿化、更优质的屋顶材料或宽敞的露台，这些自然会提升房产的整体吸引力，进而提升其残差价值。

因此，估价工程师在分析残差时必须考虑这些动态因素。他们不仅应评估高品质设施和特征所带来的内在价值，还应理解这些改进如何与房产的物理尺寸相互作用。这种全面的方法可确保估价准确反映房产的定性与定量属性。

因此，建议同时绘制 CQA-残差曲线和 CQA-残差/平方英尺曲线并研究其差异，这可以揭示某一市场区域中的模式。估价工程师可以让程序同时计算两种残差，并取某种加权平均来计算分配给估价对象的残差。其影响可能大，也可能不大。

12.3 CQA 曲线的特征

“状况—质量—吸引力”（CQA）曲线提供了一个精细的模型，用以理解房产价值如何随给定区域内的吸引力、状况和市场动态而波动。该模型捕捉了影响房地产市场的经济因素与物理因素之间的相互作用，反映了不同细分市场中房产估价的微妙现实。

12.3.1 CQA 曲线各组成部分的分析

1. 曲线左侧的指数衰减：

当 CQA 评分从约 2 降至 0 时，房产表现出快速恶化的特征。曲线的这一部分可以很好地用指数衰减函数来表示，突显了缺乏适当维护和更新的房产会以多快的速度丧失其吸引力和价值。木材腐烂、颜色褪去、金属锈蚀等物理退化会加速这种衰退，尤其是在未定期维护的住宅中。

2. 中间区段的线性增长：

市场的主体通常属于这一类别，其吸引力在评分约 2 至 8 之间近乎线性地增长。这一区段反映了那些总体上维护达标、但不一定具备突出品质或区位的房产。它代表了住房市场中广泛的中间层级，涵盖了住房市场的大多数——这些住宅基于其标准特征和整体状况进行有竞争力的定价。

3. 曲线右侧的指数增长：

在此，曲线在评分 8 至 10 之间呈指数上升，映照出市场中吸引力较高的一端。这一细分市场中的房产往往属于较富裕的人群，他们不仅投入于更优质的维护，还投入于能显著提升房产吸引力和价值的改进。指数增长反映了此类高品质房产的稀缺性以及富裕买家的旺盛需求。而且要明白，一套价格较低的住宅在吸引力方面可能评分很高，而一套面积大、价格昂贵的住宅评分却可能很低。这些动态可通过对 MARS 的专业运用而揭示出来，而大多数估价师通常不会注意到它们。

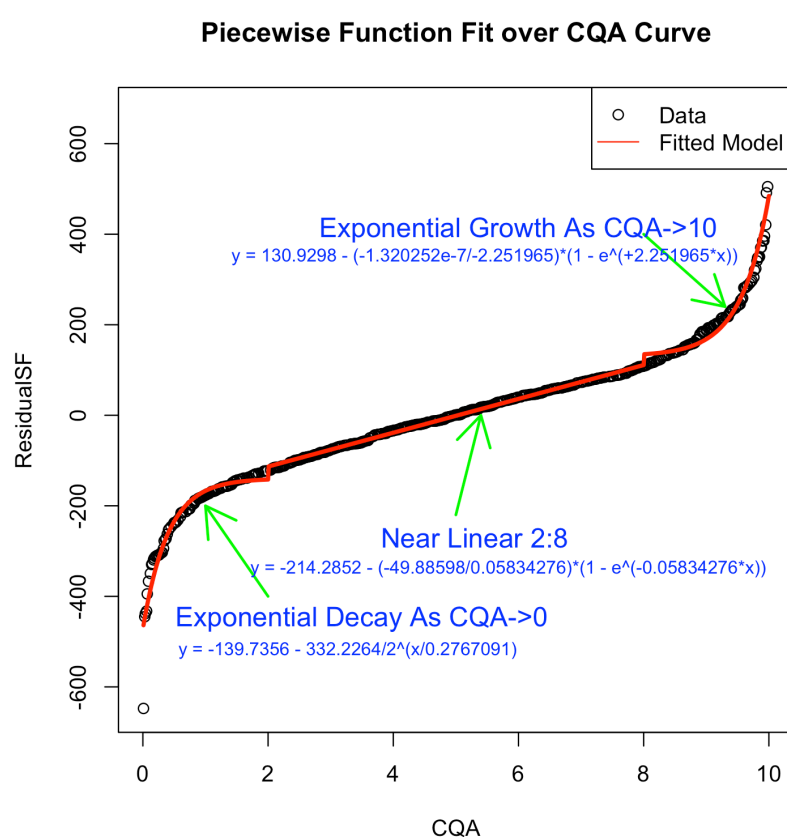
12.3.2 市场动态与社会含义

CQA 曲线隐含地讲述了不同区域内相对财富与经济差距的故事。在圣马特奥县这样的地方，Burlingame 等区域与 Hillsborough、Woodside 或 Atherton 等更富裕社区之间的房产价值差异，展示了显著的社会经济分化。这种差距影响着从房产维护到市场估值的方方面面，并最终影响买家群体构成和房地产市场趋势。

此外，还有这样一个例子：一位相对富有的人士，在自己感到舒适、且出于种种原因更愿意居住的中档社区中投资建造高品质住宅——这凸显了房地产经济学的另一个关键方面——“过度建造”房产的概念。此类房产虽然极具吸引力，但可能在当地找不到能够支付其在更高价位市场区域本可卖出的价格的买家。

CQA-残差曲线是每个市场区域所特有的，在进行分析时不应被忽视。

图 2: CQA-残差特征函数



12.4 估价对象残差

与可比交易实例不同，估价对象并不具有客观确定的残差；它必须由估价工程师（VE）估计。这一估计为残差约束法（RCA）规程引入了一定程度的主观性，标志着个人判断发挥关键作用的一个重要环节。

为了估计“状况—质量—吸引力”（CQA）评分，VE 会细致地分析已排序的房产，在排序中寻找可辨识的模式。这一分析通常涉及对每处房产相关 MLS 照片的详细审阅。对于大型数据集（如 600 处房产）而言，逐一详尽审阅并不现实。相反，VE 运用实务推理来近似判断估价对象在整体排序中的可能位置。需要大量修缮的房产（即“待修房”）通常位于排序底部附近，而更新程度高、吸引力强的房产则往往排在顶部附近。

随后, VE 采用更聚焦的方法, 在排序中“跳跃”, 以识别与估价对象吸引力最接近的可比交易实例。随着范围的缩小, VE 会更加关注各房产之间的相似之处, 尤其是在质量、状况、设计美感(包括油漆、木作、石作和窗户等元素)、功能效用及其他相关特征等方面

必须理解的关键一点是, VE 的判断从根本上受市场所确立的排序的引导。这确保了该过程中的主观因素被锚定于市场现实, 力求对估价对象作出公平而准确的估价。

12.5 RCA 价值结论

12.5.1 一般考虑

确定 CQA 评分对于确立房产的估计残差价值至关重要, 因为这两个要素本质上是相互关联的。虽然不使用 CQA 也可以对房产进行评估, 但这一指标通过展示该房产相对于市场区域内其他房产所处的位置, 提供了有价值的背景信息。例如, CQA 评分为 3.2 表明该房产的吸引力高于周边 32% 的可比房产。如果对所赋评分存在异议, 异议方必须查阅排序, 并基于照片证据或其他实质性数据来论证其提出的任何调整。如果排序体系是稳健的, 且估价工程师(VE)已在此框架内准确定位了该房产, 则不太可能出现重大分歧。

在报告阶段, 通常会选择 6 至 12 个近期且相关的可比交易实例来支持估价结论。对这些可比交易实例与估价对象之间的差异进行分析, 并对每个变量作相应调整。然后将这些调整应用于可比交易实例的净成交价格, 得出调整后成交价格。正如数学上已证明的那样, 这些调整后价格应收敛于估价对象的估计成交价格。

12.5.2 复核、审计与投诉

一旦确定了估价对象的估计残差, 就可以为所有可比交易实例计算残差调整。这些调整对 VE 针对其认为相关的特定 CQA 变量所作的进一步手工分解构成约束。必须强调的是, 正如证明 II 所确认的, 这些细化分解仅具解释性, 并不影响最终估值。

对于不满意估价价值结论的用户所提出的投诉, 传统估价师通常不得不浪费时间解释为什么某处价格更高或更低的房产并不具有用户所期望的价值影响。而使用 RCA, 所有可能的可比交易实例(通常为 100-600 个)通常都被评估为与价格结论完全相同的价值, 因此展示任意某处房产的调整几乎不费力气。除估价对象残差的评分外, 整个 RCA 过程都是自动化且客观的。因此, RCA 的另一个优势在于, 针对 VE 主观判断不完善之处的批评主要局限于这一个环节——并且可以得到有效而高效的管理。此外, 假设 R^2 约为 80%, 估计估价对象残差时 $\pm 10\%$ 的误差对最终价值估计的影响仅为 $\pm 2\%$ 的误差。

在许多情况下, 估价对象在房产残差排序中的契合度如此之高, 以至于可以预期其准确度在 $\pm 1\%$ 以内。

估计残差与估价对象的 MARS 估计成交价格之和, 即得出最终的价值结论, 称为残差约束法(RCA)价值结论。这一过程确保了房产估价方法的全面性与透明度, 增强了人们对评估结果准确性与公正性的信心。

13 结论

残差约束法是一个具有三项实质性贡献的框架：其一，源于在整个市场而非小规模可比交易实例集合上拟合模型的结构抗锚定性（证明 I）；其二，一项不变性结果，使残差分解得以作为解释层而不扰动价值结论（证明 II 及其后的注记）；其三，一项经验规律——CQA-残差特征曲线——它似乎以特征性的“指数—线性—指数”形态在各市场区域中反复出现。该框架的局限性同样具有实质性：残差分解具有不变性但不具识别性；CQA 定位步骤依赖尚未经正式评分者间一致性研究验证的 VE 判断；其准确性主张尚待与替代方法进行基准比较。

本文所呈现的是一个框架和一套规程。本文尚未呈现、但目前正在进行的更大研究计划所提供的，是一条可执行、可审计追溯的流水线，它以具有可辩护默认设置的软件将该规程付诸操作。这条流水线——用于 MARS 基函数阶段的 CRAN 上的 earthUI 包、用于在这些基函数上构建正则化结构模型的 glmnetUI 包，以及用于空间平滑的 mgcvUI 包——是本期姊妹文章的主题，也是从“作为方法的 RCA”通往“作为可交付成果的 RCA”的桥梁。估价监管中的理论与实践鸿沟（Craytor, 2026, 本期）说明了这座桥梁为何重要：估价产出在经济上被当作前瞻性风险输入来使用，而一个使现值意见可审计、可复现的框架，是对这种前瞻性用途进行任何诚实探讨的前提条件。RCA 是这一更大计划中的一步棋。它并非整个计划，本文也并未如此宣称。

参考文献

- Ambrose, Brent W. et al. (2025). “Do Appraiser and Borrower Race Affect Mortgage Collateral Valuation?” In: *Review of Finance*. DOI: [10.1093/rof/rfaf046](https://doi.org/10.1093/rof/rfaf046).
- Craytor, William Bert (2026). “The Doctrine–Practice Gap in Real Estate Appraisal: A Structured Account of Functions, Boundaries, and Tensions”. In: *Valuation Engineer Journal* 1.1.
- Friedman, Jerome H. (1991). “Multivariate Adaptive Regression Splines”. In: *The Annals of Statistics* 19.1, pp. 1–141.
- Hastie, Trevor, Robert Tibshirani, and Jerome H. Friedman (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second. Springer Series in Statistics. New York, NY: Springer. ISBN: 978-0-387-84857-0.
- James, Gareth et al. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics 103. New York: Springer. ISBN: 978-1-4614-7137-0.
- Krivorotov, George and Michael LaCour-Little (2020). “AVM versus Appraisal-Based Underwriting in Refinance Mortgages: The Trade-off Between Noise and Bias”. In: *SSRN Electronic Journal*. DOI: [10.2139/ssrn.3603327](https://doi.org/10.2139/ssrn.3603327).
- Kuhn, Max and Kjell Johnson (2016). *Applied Predictive Modeling*. Corrected at 5th printing. New York: Springer. ISBN: 978-1-4614-6848-6.
- Milborrow, Stephen (2024a). *Earth: Multivariate Adaptive Regression Splines*. R Package. URL: <https://CRAN.R-project.org/package=earth>.
- (2024b). *Notes on the Earth Package*. Tech. rep. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (2024c). *Variance Models in Earth*. Tech. rep. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.
- Perry, Andre M., Jonathan Rothwell, and David Harshbarger (2018). *The Devaluation of Assets in Black Neighborhoods*. Tech. rep. Brookings Institution.
- Perry, Andre M., Jonathan Rothwell, J. Williamson, et al. (2024). *How Racial Bias in Appraisals Affects the Devaluation of Homes in Majority-Black Neighborhoods*. Tech. rep. Brookings Institution.
- Property Appraisal and Valuation Equity Task Force (2022). *Action Plan to Advance Property Appraisal and Valuation Equity*. Tech. rep. U.S. Department of Housing and Urban Development.

- Tzioumis, Konstantinos (2016). “Appraisers and Valuation Bias: An Empirical Analysis”. In: *Real Estate Economics* 45.3, pp. 679–712. DOI: [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- Williamson, Jake and Mark Palim (2022). *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Tech. rep. Fannie Mae.
- Zhang, Wengang (2020). *MARS Applications in Geotechnical Engineering Systems: Multi-Dimension with Big Data*. Singapore: Springer. ISBN: 978-981-13-7421-0.

earthUI: earth (MARS) 软件包的交互式界面

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

William Bert Craytor

第 1 卷，第 1 期 · 2026 年夏季

发表日期: 2026 年 7 月 1 日

DOI: [10.67330/7fj41b65](https://doi.org/10.67330/7fj41b65)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

本开放获取文章依据 [知识共享署名 4.0 国际许可协议 \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) 的条款发布。

摘要。 earthUI 是 R 语言 earth 软件包的图形用户界面，该软件包实现了多元自适应回归样条 (MARS, multivariate adaptive regression splines)。它提供三种用途模式——通用预测建模、房地产估价和市场区域分析——并引导用户完成数据导入、模型配置、模型拟合，以及对 earth 诊断信息和图形输出的解读。本文记录了 earthUI 的数据格式要求、建模工作流程、输出显示界面以及完整的功能参考。

关键词： MARS, 多元自适应回归样条, earth, R, 图形用户界面, 房地产估价

出版者按 (2026 年 7 月)： 本文描述的是 earthUI 0.10.0。付印前不久，earthUI 0.11.0 吸收了配套应用 glmnetUI 和 mgcvUI (见随后两篇文章)，将 MARS、弹性网络和 GAM 工作流程整合到单一应用程序中。本文所述的界面与工作流程不受此次整合的影响。

目录

1	引言	3
1.1	什么是 earthUI?	3
1.2	一个包，三个例程	4
1.3	三种用途模式	4
1.4	房地产专用功能	4
1.5	安装 earthUI 与 vProlog	5
1.6	快速入门	5
2	MLS 输入数据要求	7
2.1	文件格式与结构	7
2.2	估价模式所需的列	7
2.3	特殊列命名约定	8
2.4	数据质量与完整性	8
2.5	估价对象行的位置	8
3	通用模式	9
3.1	概述	9
3.2	侧边栏 workflow	9
3.3	主面板选项卡	9
3.4	设置持久化	10
3.5	深色模式	10
3.6	语言环境与区域设置	10
3.6.1	支持的国家	11
3.6.2	纸张覆盖设置	11
3.6.3	保存默认设置	11

4	估价模式	12
4.1	估价对象行的处理	12
4.2	生效日期与销售日龄	12
4.3	特殊列指定	12
4.4	残差贡献（稀有特征估计）	13
4.5	RCA 调整概述	13
5	市场区域分析模式	14
5.1	与估价模式的区别	14
5.2	何时使用市场模式	14
6	变量选择	15
6.1	目标变量	15
6.2	预测变量表	15
6.3	数据类型检测与覆盖	15
6.4	Factor 与 Linear 标志	15
6.5	特殊列类型参考	16
6.6	多个目标变量	16
7	参数选择	17
7.1	基本参数	17
7.2	前向过程	17
7.3	允许的交互	18
7.4	剪枝	18
7.5	交叉验证	18
7.6	子集过滤	18
7.7	推荐参数值	19
7.7.1	前向过程参数	19
7.7.2	剪枝参数	20
7.7.3	交叉验证参数	20
7.7.4	方差模型	21
7.7.5	Earth 输出文件	21
7.8	设置默认值	21
8	拟合 Earth 模型	22
8.1	Fit 按钮	22
8.2	拟合模态框与跟踪输出	22
8.3	拟合日志	22
9	下载数据	23
9.1	输出列	23
9.2	列排序与 Excel 格式	23
9.3	CQA 评分	24
9.4	排序与估价对象行	24
9.5	mgcvUI 自动导出	24
9.6	glmnetUI 导入	24
9.6.1	主要优势	24
9.6.2	重要提示：权重列必须一致	25
9.6.3	工作流程	25

10 RCA 计算与下载	26
10.1 打开 RCA 对话框	26
10.2 CQA 评分插值	26
10.3 输出列	26
10.4 RCA Adjustments 选项卡	27
11 销售比较网格	28
11.1 概述与用途	28
11.2 可比案例选择模态框	28
11.3 网格布局	28
11.4 售价、让价优惠与 Net SP	29
11.5 分组行（位置、地块、房龄）	29
11.6 模型变量行	29
11.7 CQA Residual 与残差特征行	29
11.8 调整后售价公式	29
11.9 工作表保护与可编辑单元格	30
11.10 在 Excel 中使用网格	30
12 下载报告	31
12.1 生成 Quarto 报告	31
12.2 转换 Quarto 报告	31
12.3 报告内容	31
13 后处理模块：弹性网络与 GAM	33
13.1 为什么需要辅助拟合?	33
13.2 启动与共享状态	33
13.3 弹性网络例程	33
13.4 GAM 例程	33
13.5 一个实用的工作流	33
14 演示数据集：Appraisal_1.csv	35
14.1 描述	35
14.2 数据列	35
14.3 建议的快速上手步骤	36
15 实验性功能：Prolog 文本提取与推导规则	37
15.1 它的作用	37
15.2 环境要求：vProlog (Trealla)，而非 SWI-Prolog	37
15.3 启用与运行	37
15.4 推导规则	38
15.5 带版本号的活动数据文件	38
15.6 编程接口与随附文件	38
15.7 相关实验：气候区先验	39
15.7.1 超越加利福尼亚：辖区包与市场区域模块	39

1 引言

1.1 什么是 earthUI?

earthUI 是 R 语言 `earth` 包的图形用户界面，该包实现了多元自适应回归样条 (MARS, multivariate adaptive regression splines) (Friedman 1991)。建模引擎是 Stephen Milborrow 的

earth 包 (Milborrow 2024), Milborrow (n.d.[a]) 与 Milborrow (n.d.[b]) 对其有详细的文档说明。它以本地 Shiny 应用程序的形式运行——无需登录、无需服务器、也没有账户。您从 R 中启动它, 导入数据集 (CSV 或 Excel), 配置模型, 然后交互式地进行拟合。

该应用程序提供了完整的工作流程: 数据导入、变量配置、模型拟合、诊断图、变量重要性、模型方程, 以及可下载的 HTML、Word 或 PDF 格式报告。

1.2 一个包, 三个例程

从当前版本起, earthUI 包还包含了它的两个后处理模块, 这两个模块在早期版本中是独立的包 glmnetUI 和 mgcvUI。安装 earthUI 即安装了全部三个例程; 每个例程都作为独立的本地 Shiny 应用程序运行:

- `earthUI::launch()` ——本文所描述的 earth (MARS) 应用程序, 运行在端口 7878 上。`earth()` 是核心估值引擎。
- `earthUI::launch_glmnet()` ——弹性网后处理模块 (端口 7879), 它导入 earth 模型的基函数, 并在 lasso / 弹性网正则化下重新估计它们 (参见下文“glmnetUI Import”)。
- `earthUI::launch_mgcv()` ——GAM 后处理模块 (端口 7880), 它以 earth 模型的节点位置作为 mgcv 平滑项的起点 (参见下文“mgcvUI Auto-Export”)。

这三个例程共享同一个项目树、同一个设置数据库和同一套支撑基础设施, 因此在一个应用程序中打开的项目可以立即供其他应用程序使用。glmnet 和 mgcv 例程是对 earth 模型价值结论的佐证, 而不是相互竞争的工作流程: 预期的使用模式是先拟合并完善 earth 模型, 然后在需要佐证时针对同一项目运行这些模块。

1.3 三种用途模式

earthUI 以项目的形式组织工作——项目是有名称的实体, 将您的输入文件、输出和设置集中保存在一个固定位置。每个项目都带有一个 **Purpose** (用途), 在创建项目时于 New Project wizard (新建项目向导) 中选择。共有三种模式:

- **General** (通用) ——适用于任何类型的总体或数据集的 Earth 回归。这是默认模式。它提供完整的 MARS 建模工作流程, 不含任何特定领域的附加功能。
- **For Appraisal** (估价用途) ——为房地产估价量身定制的 Earth 回归。增加了针对单一房产估值的功能, 包括估价对象 (subject) 处理、特殊列指定、残差约束法 (RCA), 以及销售比较网格 (Sales Comparison Grid) 的生成。
- **Market Area Analysis** (市场区域分析) ——为市场区域研究量身定制的 Earth 回归。增加了在既定市场中分析成组房产的功能, 而不是对单一估价对象进行估值。

在这三种模式下, 核心建模引擎完全相同——您始终是在拟合一个 Earth (MARS) 模型。用途设置控制的是哪些附加工具和界面元素可用。

每个项目的用途是固定的。项目的用途——连同其国家、州、县和城市——在项目创建时即被锁定。切换到另一个项目会将 earthUI 重置为干净的默认状态 (已导入的数据、模型结果、标签页、变量配置和 earth 参数全部清空), 然后恢复该项目针对其用途保存的配置。

1.4 房地产专用功能

当选择 **For Appraisal** 或 **Market Area Analysis** 时, earthUI 会激活若干为房地产分析设计的功能:

- **特殊列指定**——每个预测变量都可以被标记为某种特殊角色, 例如 `contract_date`、`dom`、`concessions`、`latitude`、`longitude`、`living_area`、`lot_size`、`actual_age`、`effective_age`、`area`、`site_dimensions` 或 `display_only`。这些指定控制该列在拟合、输出和销售网格生成过程中的处理方式。完整参考请参见第 6 章。
- **经纬度取整**——被指定为 `latitude` 或 `longitude` 的列会自动取整到小数点后 3 位, 以防止过拟合。

- **Sale Age** 列——当某列被指定为 `contract_date` 且提供了 Effective Date（生效日期）时，earthUI 会计算一个 `sale_age` 列（销售日期与生效日期之间的天数），并将其作为预测变量代入。
- **RCA** 计算（仅限估价模式）——在估价模式下，模型拟合完成后，earthUI 可以计算残差约束法（RCA）输出，生成逐个可比销售案例的调整、净 / 总调整汇总，以及估价对象的调整后销售价格。
- **销售比较网格**（仅限估价模式）——生成一个多工作表的 Excel 工作簿，其中包含可比案例调整、残差特征分解和调整后的销售价格的公式。参见第 11 章。

提示：General 用途模式适用于任何数据集——金融、科学、工程或房地产。估价模式和市场模式只是为房地产专业人士增加了便利功能。

1.5 安装 earthUI 与 vProlog

earthUI 已发布在 CRAN 上，因此标准安装只需在 R 中执行一行命令：

```
install.packages("earthUI")
```

本文描述的是当前版本。如果您阅读本文时 CRAN 上仍是较旧的版本，可以从 GitHub 安装开发版（注意 `subdir` 参数——该包位于仓库的 `pkg/` 文件夹中）：

```
remotes::install_github("wcraytor/earthUI", subdir = "pkg")
```

vProlog（可选）。实验性的备注处理功能（参见本文末尾附近的“Prolog Text Extraction”）使用 vProlog 包——一个自包含的嵌入式 Prolog 引擎。它被有意设计为非强制依赖：earthUI 在没有它的情况下也能安装并运行，Prolog 相关功能只是在其安装之前保持禁用状态。vProlog 通过作者的 `r-universe` 仓库分发，该仓库为 Windows 和 macOS 提供预编译的二进制文件（无需编译器工具链）：

```
install.packages("vProlog",
  repos = c("https://wcraytor.r-universe.dev",
    getOption("repos")))
```

不需要系统级的 Prolog 安装（无论是 SWI-Prolog 还是其他）——Trealla 引擎已编译进包本身。本文中的其他所有内容在没有 vProlog 的情况下均可正常使用。

1.6 快速入门

使用 earthUI 的步骤：

1. **安装包**——按上文所述从 CRAN 安装。
2. **启动应用程序**：在 R 中运行 `earthUI::launch()`，或从命令行运行 `Rscript -e 'earthUI::launch()'`。应用程序将在您的网页浏览器中打开，端口为 7878。您也可以直接在浏览器中访问 `http://localhost:7878` 来打开应用程序。（后处理模块以同样的方式启动：`launch_glmnet()` 使用端口 7879，`launch_mgcv()` 使用端口 7880。）
3. **创建或打开项目**——在侧边栏的第 1 节中进行。New Project wizard 会记录项目名称、Purpose（General、For Appraisal 或 Market Area Analysis）、位置（国家、州、县、城市），以及可选的初始数据文件，该文件会被复制到项目的 `in/` 文件夹中。
4. **导入数据**——在第 2 节中进行，该节列出活动项目 `in/` 文件夹中的 CSV 和 Excel 文件。您通过 Finder 或资源管理器手动添加的文件，在点击 Refresh（刷新）后会显示出来。
5. **配置变量**——选择目标变量和预测变量，设置数据类型，并指定任何特殊列角色。
6. **设置 Earth 参数**——`degree`、`nprune`、子集过滤器以及其他 `earth()` 参数。
7. **拟合模型**——点击“Fit Earth Model”（拟合 Earth 模型），并在主面板中查看结果。

8. 导出——将预测结果下载为 Excel，生成 Quarto 报告包，或（在估价模式下）计算 RCA 调整和销售比较网格。

配置通过第 3 节和第 4 节中的“Save current as default”（将当前设置保存为默认）按钮保存在服务器端的项目数据库中，以项目和用途为键（参见第 3 章“Settings Persistence”）。

2 MLS 输入数据要求

对于房地产估价和市场分析工作流程，您的输入数据通常来自多重挂牌服务（MLS，Multiple Listing Service）的导出文件。本章描述预期的文件结构以及 earthUI 可以使用的列。

2.1 文件格式与结构

earthUI 接受 CSV 和 Excel (.xlsx、.xls) 文件。导入时，列名会被自动转换为 snake_case——例如，“Living SqFt”变为 living_sqft，“Contract Date”变为 contract_date，“Sale Price”变为 sale_price。这种规范化确保了整个工作流程中列引用的一致性。导入时使用的 CSV 分隔符和小数点符号由区域设置决定（参见第 3 章“Locale & Regional Settings”）。

您的数据文件应为平面表格，每个房产占一行，每个属性占一列。文件的第一行必须包含列标题。

2.2 估价模式所需的列

虽然 earthUI 可以处理任何列集合，但完整的估价工作流程（RCA 调整 + 销售比较网格）在您的 MLS 导出文件中包含以下列时效果最佳：

列	特殊类型	用途
Sale Price	(目标变量)	模型的响应变量
Contract Date	contract_date	用于计算 sale_age（距生效日期的天数）
Listing Date	listing_date	在没有 DOM 列时，与合同日期一起用于计算 DOM
Days on Market	dom	在市天数；显示在销售网格中
Concessions	concessions	销售让步；销售网格中 Net SP = Sale Price – Concessions
Living Area (SF)	living_area	启用每平方英尺残差 (residual_sf、cqa_sf)
Lot Size	lot_size	在销售网格中归入”Site Size Dimensions”行
Site Dimensions	site_dimensions	在销售网格中与地块面积归为一组
Latitude	latitude	取整到小数点后 3 位；用于邻近度计算和 Location 组
Longitude	longitude	取整到小数点后 3 位；用于邻近度计算和 Location 组
Area ID	area	在销售网格中归入”Loc: Long Lat Area”行
Actual Age	actual_age	在销售网格中归入”Actual Age Effective Age”行
Effective Age	effective_age	在销售网格中与实际房龄归为一组
Parcel Number	(无)	显示在销售网格的 APN 行中
Listing ID	(无)	显示在销售网格的 APN 行中
Address	display_only	显示在销售网格的 Address 行中；不参与模型

电子表格的列名可以使用外语——“特殊”名称使用英语，以便 R 程序能对它们进行特殊处理。除此之外，给定的列名会出现在回归模型、图表，以及（在进行估价时）中间销售网格 (Intermediate Sales Grid) 中。

并非所有列都是必需的。earthUI 会自行适应——如果某列缺失，相应的功能会被直接省略。例如，如果没有指定 concessions 列，销售网格中的 Net SP 行将显示未扣除让步的销售价格。不过，对于房地产定价模型，强烈建议提供某些列以获得可接受的拟合效果：

1. **Sale Age**——合同销售日期与估价或分析的生效日期之间的天数。如果使用多年的销售历史数据，尤其是超过 5 年的时段，sale_age 往往在销售价格估计中扮演核心角色。事实上，它常常重要到没有它 earthUI 根本无法给出任何模型。
2. **Living Area**——也称作“Living Sqft”、“GLA” (gross living area, 总居住面积) 等。这是销售价格的另一个主要决定因素。

3. **Total Bath Count**——全卫、四分之一卫、半卫和四分之三卫浴的总数。例如，两个全卫加一个半卫的值为 2.5。
4. **Garage Bays** 或 **Garage Area**——车库车位数量或车库面积（平方英尺）。
5. **Lot Size**——房产的土地面积，通常以平方英尺或英亩为单位。
6. **Longitude、Latitude**，如果有的话再加上 **Area ID**。这些变量的调整将在销售网格中合并为单一的 Location 调整项。

2.3 特殊列命名约定

earthUI 通过列的特殊类型指定而非列名来识别列。您可以在 MLS 导出文件中为列取任何名称——关键在于您在变量配置表（第 6 章）中指定了正确的特殊类型。

例如，您的 MLS 可能将居住面积导出为“GLA”、“Living SqFt”、“liv_area”或“gross_living_area”。导入之后（此时它已变为 snake_case），您只需在 Special dropdown（特殊类型下拉框）中将其指定为 living_area。此后，无论其原始名称如何，earthUI 都会将其用于每平方英尺残差计算和销售网格分组。

2.4 数据质量与完整性

- **缺失值 (NA)**：任何预测变量或目标变量列中含有 NA 值的行会在拟合前自动删除。na.action 在内部始终设置为 na.fail，NA 的删除在调用 earth() 之前完成。
- **日期列**：日期应采用 R 可解析的格式（例如 2025-06-15、06/15/2025、June 15, 2025）。当至少 80% 的值能够成功解析时，earthUI 会自动检测日期列。两位数年份（例如 06/15/25）会被读取为最接近的匹配年份，因此近期的 MLS 导出文件可以正确解析；四位数年份始终按字面值处理。
- **数值列**：销售价格、居住面积、地块面积、让步及类似字段必须是数值型。如果您的 MLS 导出的价格带有货币符号或千位分隔符（例如“\$350,000”或“350.000,00”），您可能需要在导入前清理这些数据。
- **因子列**：区域 ID、风格或状况等分类变量应包含数量适中的唯一值。earthUI 会自动检测因子，但您可以覆盖该检测结果。

提示：导入后请查看预测变量表中的 T/F/NA 列——它显示每列的 true/false/NA 计数。缺失值超过 30% 的列会以红色标出。请考虑排除高缺失率的列，或在导入前清理数据。

2.5 估价对象行的位置

在 **Appraisal**（估价）模式下，第 1 行必须是估价对象 (subject)。其余所有行均为可比销售案例。估价对象的销售价格可以留空 (NA) 或设为任意值——无论如何，earthUI 在拟合时都将其视为 NA。

在 **Market Area Analysis** 模式下，将估价对象放在第 1 行是可选的，且不会自动排除任何行。若要将估价对象所在行（或任何其他行）排除在拟合之外，可在指定的权重列中将其值设为零，或使用子集过滤器。

在 **General** 模式下，没有特殊的行处理——所有行一视同仁。

3 通用模式

3.1 概述

通用模式 (General Purpose mode) 是启动 earthUI 时的默认模式。它为任何数据集——而不仅仅是房地产数据——提供完整的多元自适应回归样条 (MARS, multivariate adaptive regression splines) 建模工作流。您可以将 earthUI 用于科学数据、金融分析、工程研究, 或任何希望探索变量间非线性关系与交互作用的回归问题。

在通用模式下, 界面会省略房地产专用功能 (特殊列、销售日龄、坐标舍入、RCA)。侧边栏经过精简, 聚焦于变量选择、参数配置、模型拟合与导出。

3.2 侧边栏工作流

侧边栏按编号组织为可折叠的分区, 引导您从项目选择一路走到导出:

1. Project (项目) —— 选择一个现有项目, 或通过 New Project (新建项目) 向导创建新项目; 该向导会记录用途、国家和行政区位置 (创建后全部锁定)。当前活动项目决定了输入文件的读取位置以及所有输出的写入位置——不存在单独的输出文件夹字段。

2. Import Data (导入数据) —— 从活动项目的 in/ 文件夹中选择 CSV 或 Excel 文件; Refresh (刷新) 链接会重新扫描该文件夹, 以发现手动添加的文件。对于包含多个工作表的 Excel 文件, 会出现工作表选择器。列名会被自动转换为 snake_case。

3. Variable Configuration (变量配置) —— 目标变量选择器 (支持多个目标变量)、Effective Date (生效日期), 以及带有 Include、Factor、Linear 和 Latent 复选框的预测变量表。详见第 6 章。

4. Earth Call Parameters (Earth 调用参数) —— earth() 函数的全部参数: degree、penalty、nk、剪枝方法、交叉验证等。完整的参数参考见第 7 章。

Execute Prolog Processing (执行 Prolog 处理) —— 位于第 4 与第 5 分区之间的一个可选、无编号分区, 除非在 Settings (设置) 中启用了实验性的 Prolog 功能, 否则处于禁用状态。见第 15 章。

5. Fit Earth Model (拟合 Earth 模型) —— 一个随机种子字段和异步运行模型的按钮。拟合的详细说明见第 8 章。

6. Download Estimated Target Variable(s) & Residuals (下载估计的目标变量与残差) —— (在估价模式和市场模式下标题为“Download Estimated Sale Prices & Residuals”, 即下载估计售价与残差) 将预测值、残差、CQA 评分以及各 g 函数的贡献导出为一个 Excel 文件。见第 9 章。

7–8. RCA Adjustments and Sales Grid (RCA 调整与销售网格) —— 仅限估价模式。见第 10 章和第 11 章。

9. Generate Quarto Report (生成 Quarto 报告) —— 将一个自包含的 Quarto 报告包 (.qmd 源文件及资源文件) 写入项目的输出文件夹。在非估价模式下编号为 7。见第 12 章。

10. Convert Quarto Report (转换 Quarto 报告) —— 将任意 .qmd 文件渲染为 HTML、Word 或 PDF。见第 12 章。

3.3 主面板选项卡

拟合完成后, 主面板提供十个选项卡:

选项卡	内容
Data (数据)	导入数据的预览。在估价模式下，拆分为“Subject Property”（估价对象）和“Comparable Sales”（可比销售案例）两个表格。
Equation (方程)	以 LaTeX/MathJax 显示的拟合模型方程，展示每个基函数及其系数。
Summary (摘要)	关键指标 (R^2 、GRSq、GCV、RSS、 $CV R^2$) 和系数表。
Variable Importance (变量重要性)	预测变量重要性得分的条形图和排名表。
Contribution (贡献)	g 函数贡献图。从下拉菜单中选择一个 g 函数；单变量预测变量显示折线图，双变量交互项 ($\text{degree} \geq 2$) 显示 3D 透视图和等高线图。
Correlation (相关性)	预测变量相关性的热力图。
Diagnostics (诊断)	三幅图：残差 vs 拟合值、正态 Q-Q 图，以及实际值 vs 预测值。
RCA Adjustments (RCA 调整)	（仅估价模式）各可比销售案例的残差调整百分比 (Residual Adj %)、净调整百分比 (Net Adj %) 和总调整百分比 (Gross Adj %) 的直方图，在第 7 步 (Calculate RCA Adjustments) 完成后填充。
Residual Contribution (残差贡献)	通过将模型残差对模型从未使用过的列进行联合回归，估计稀有特征或被排除特征的市场贡献（见 subsection 4.4 ）。
ANOVA	ANOVA 分解表，显示每个基函数对 RSS 的贡献。
Earth Output (Earth 输出)	来自 ‘earth::summary()’ 的原始文本输出，包括剪枝过程的详细信息。

3.4 设置持久化

earthUI 将您的配置保存在服务器端的项目数据库中（位于 regProj 根目录下的 `projects.sqlite`），以项目和用途为键。保存是按钮驱动的：第 3 分区中的“Save current as default”（将当前设置保存为默认）按钮保存目标变量、生效日期和预测变量配置，而第 4 分区中的按钮保存 earth 调用参数和 Allowed Interactions（允许的交互项）矩阵。每个按钮只更新自己负责的那一半，因此它们绝不会相互覆盖，并且拟合不会自动保存——您可以自由试验而不必担心覆盖已保存的默认设置。一个项目的多个输入文件（一个小型测试抽样、完整数据集）共享同一份配置；而且由于设置存放在项目目录树中，当您同步、备份或共享 regProj 文件夹时，设置会随之迁移。

3.5 深色模式

点击右上角的月亮/太阳图标可在浅色与深色主题之间切换。主题偏好保存在本地存储中，并在会话之间持续生效。

3.6 语言环境与区域设置

earthUI 通过基于国家的语言环境系统支持国际化的数字、日期和 CSV 格式约定。顶部菜单栏 **Settings**（设置，齿轮图标）菜单中的 **Country**（国家）下拉菜单可为 31 个受支持的国家选择预设。每个预设配置以下内容：

- **CSV 分隔符**——美国/英国/日本使用逗号（,），欧洲大部分国家使用分号（;），因为那里逗号被用作小数点。
- **小数点符号**——句点（.）或逗号（,）。
- **千位分隔符**——逗号（美国/英国/日本）、句点（德国/意大利/西班牙）、空格（芬兰/法国/波兰/波罗的海国家/乌克兰/俄罗斯）或撇号（瑞士）。

- 日期格式——MM/DD/YYYY（美国）、DD/MM/YYYY（欧洲大部分国家）或 YYYY-MM-DD（瑞典/立陶宛/日本/加拿大）。
- 纸张尺寸——Letter（美国/加拿大/墨西哥）或 A4（其他所有国家）。

3.6.1 支持的国家

国家	CSV 分隔符	小数点	千位分隔符	日期	纸张	
美国、加拿大 (csv)	,	.	,	MDY/YMD	Letter	
英国、爱尔兰、澳大利亚、新西兰	,	.	,	DMY	A4	
德国、奥地利、比利时、荷兰、意大利、西班牙、葡萄牙、丹麦、巴西	;	,	.	DMY	A4	
法国、波兰、捷克、挪威、拉脱维亚、爱沙尼亚	;	,	(空格)	DMY	A4	
芬兰、瑞典	;	,	(空格)	DMY/YMD	A4	
瑞士	;	,	'	DMY	A4	
乌克兰、俄罗斯	;	,	(空格)	DMY	A4	
土耳其	;	,	.	DMY	A4	
日本、韩国	,	.	,	YMD	A4	
墨西哥	,	.	,	DMY	Letter	
立陶宛	;	,	(空格)	YMD	A4	

3.6.2 纸张覆盖设置

在国家选择器下方，**Paper**（纸张）下拉菜单（Letter 或 A4）允许您在不切换国家的情况下覆盖报告页面尺寸。CSV 分隔符、小数点符号和日期格式顺序仍遵循所选国家的预设；当您更改国家时，纸张覆盖设置会重置为该国家的默认值。

3.6.3 保存默认设置

点击 Settings 菜单中的 **Save**（保存）可全局存储您的语言环境偏好。这些默认设置适用于今后所有会话，无论您打开哪个项目。项目级设置（目标变量、预测变量、参数）单独保存在项目数据库中；语言环境默认设置则通过用户级 SQLite 数据库在所有项目间持续生效。这种两级设计面向审计事务所等需要处理来自多个国家数据的机构——将您最常用的国家设为默认，需要时再进行调整。

提示：图表坐标轴和斜率标签上的数字格式会自动适配您的语言环境。德国语言环境使用句点作为千位分隔符（200.000），芬兰使用空格（200 000），瑞士使用撇号（200'000）。不显示任何货币符号——*earthUI* 与货币无关。

4 估价模式

当您选择 **For Appraisal**（用于估价）作为用途（Purpose）时，earthUI 会将自身配置为面向单一物业估值。第 3 章描述的所有功能仍然可用；本章仅涵盖估价模式特有的附加功能。

4.1 估价对象行的处理

在估价模式下，数据集的第 1 行是估价对象（**subject**）物业，其余所有行均为可比销售案例。您的输入文件必须按此方式组织（见第 2 章）。估价对象的售价可以留空或设为任意值——earthUI 在拟合时会自动将其视为 NA。

导入后，Data 选项卡会拆分为两个部分：**Subject Property**（估价对象物业，第 1 行）和 **Comparable Sales**（可比销售案例，第 2 行及之后）。第 1 行始终被排除在模型拟合之外——通知消息“Skipping row 1 (subject). Fitting on N rows.”（跳过第 1 行（估价对象）。在 N 行上拟合。）会对此加以确认。拟合后，模型仍会为估价对象行生成预测值，在输出中显示为 `est_<target>`。

4.2 生效日期与销售日龄

在每种用途模式下，Variable Configuration 分区中都会出现一个 **Effective Date**（生效日期）字段（默认为当天日期）——通用分析和市场分析同样有生效日期，而不仅仅是估价——但它只在估价模式和市场模式下才会被使用。在这两种模式下，如果您在 Special（特殊列）下拉菜单中将某列指定为 `contract_date`，earthUI 会计算一个 `sale_age` 列——即每笔销售的合同日期与生效日期之间的整数天数。该列会取代原始日期列作为预测变量。

在指定合同日期后第一次点击 Fit 时，earthUI 会创建 `sale_age` 列，并通知您再次点击 Fit 以将其纳入模型。

4.3 特殊列指定

在估价模式和市场模式下，Variable Configuration 表中的每个预测变量旁都会出现一个 **Special**（特殊）下拉菜单。特殊类型及其作用的完整列表如下：

特殊类型	作用
no	默认——无特殊处理
contract_date	触发基于生效日期的 sale_age 自动计算
listing_date	在不存在 dom 列时，与合同日期一起用于计算 DOM
dom	标识“在售天数”(Days on Market) 列；显示在销售网格 (Sales Grid) 中
concessions	标识销售让步 (concessions)；用于销售网格中的净售价公式 (售价 - 让步)
latitude	数值舍入到小数点后 3 位；用于 Haversine 邻近度计算及销售网格中的 Location 分组
longitude	数值舍入到小数点后 3 位；用于 Haversine 邻近度计算及销售网格中的 Location 分组
area	在销售网格中归入“Loc: Long Lat Area”行
living_area	在下载输出中启用按平方英尺计算的残差 (residual_sf、cqa_sf)
lot_size	在销售网格中归入“Site Size Dimensions”行
site_dimensions	在销售网格中与地块面积归为一组
actual_age	在销售网格中归入“Actual Age Effective Age”行
effective_age	在销售网格中与实际房龄归为一组
display_only	该列包含在导出中，但排除在模型拟合之外（允许多列）
remarks	（实验性）自由文本备注列，由可选的 Prolog 步骤解析为特征列（允许多列）；见第 15 章
list	（实验性）逗号分隔的列，由可选的 Prolog 步骤拆分为独热 (one-hot) TRUE/FALSE 列（允许多列）；见第 15 章

每种特殊类型只允许指定一列 (display_only、remarks 和 list 除外)。将某个特殊类型指定给第二列时，会自动从第一列上清除该指定。变量名旁边会出现一个蓝色小徽章，显示其被指定的特殊类型。

4.4 残差贡献（稀有特征估计）

对于因过于稀有而无法进入回归通道的特征——例如仅在少数几笔销售中出现的车间 (workshop)——它们属于 RCA 残差层。**Residual Contribution** (残差贡献) 选项卡通过普通最小二乘法，将拟合模型的残差对模型从未使用过的列进行回归，以估计这些特征的市场贡献。选择多列时会对它们进行联合估计，因此相关的配套设施 (例如 outbuilding_sqft 加上 pr_has_workshop) 会分摊贡献，而不会重复计算。二元标志列产生一次性总额估计，并列出每笔带该标志的销售的残差作为证据；数值列产生每单位贡献；报告的标准误为一个 SE (通常以 $t \geq 2$ 作为显著性判断标准)。支持观测数少于十个的项会在生成的工作底稿注记中自动获得 CAUTION (警示) 标记；该注记以一个段落记录方法、估计值和注意事项，供报告使用。只有模型实际使用过的预测变量才会被排除在候选列表之外——一个曾提供给拟合但从未被前向过程选中的稀有标志列仍然符合条件，因为模型没有吸收它的任何价值。

4.5 RCA 调整概述

Calculate RCA Adjustments & Download (计算 RCA 调整并下载) 按钮 (侧边栏第 7 分区，仅在估价模式下拟合完成后可见) 计算每个可比销售案例相对于估价对象的市场推导调整值。完整的 RCA 工作流在第 10 章中描述。

计算完 RCA 调整后，**Generate Sales Grid & Download** (生成销售网格并下载) 按钮 (侧边栏第 8 分区) 变为可用。销售比较网格 (Sales Comparison Grid) 工作流在第 11 章中描述。

提示：您为估价对象指定的 CQA 评分决定了残差分布中有多大比例归属于估价对象。评分 5.00 会将估价对象置于可比销售案例的中位数位置。

5 市场区域分析模式

当您选择 **Market Area Analysis**（市场区域分析）作为用途时，earthUI 提供与估价模式相同的房地产专用功能（特殊列、销售日龄、坐标舍入），但其定位是分析一组物业，而不是对单一估价对象进行估值。

5.1 与估价模式的区别

- **不自动排除估价对象**——与估价模式不同，市场模式不会自动将第 1 行排除在拟合之外。所有行都会被纳入；若要将估价对象行排除在拟合之外，可在指定的权重列中为其赋零值（如演示数据集所示），或使用子集过滤器。
- **没有 RCA 或销售网格分区**——“Calculate RCA Adjustments & Download”和“Generate Sales Grid & Download”步骤不可用。市场模式专注于模型拟合与输出，而非逐个可比案例的调整。
- **侧边栏编号**——由于没有 RCA 和销售网格步骤，Generate Quarto Report 分区的编号为 7 而非 9（Convert Quarto Report 分区仍为 10）。

5.2 何时使用市场模式

以下情形适合使用市场区域分析模式：

- 为某个社区或市场区域构建回归模型，以理解价值驱动因素
- 分析建筑面积、房龄、地块面积和区位等变量如何影响一组物业的售价
- 为市场状况分析或社区界定准备支持材料
- 使用的数据集第 1 行包含一个估价对象物业，但您希望可以自由选择将其纳入或排除在拟合之外

提示：市场模式也适用于一般的房地产回归场景——您需要特殊列功能（销售日龄、坐标舍入），但不需要 RCA 调整 workflow。

6 变量选择

侧边栏第 3 节——**Variable Configuration**（变量配置）——用于选择哪些列参与模型，以及它们如何被处理。

6.1 目标变量

第 3 节顶部的 **Target (response) variable(s)**（目标（响应）变量）下拉框列出数据集中的每一列。选择一列可拟合标准的单响应模型；选择多列则拟合多响应模型。选择多个目标变量时，模型使用 `earth(cbind(y1, y2, ...) ~ .)` 同时拟合所有响应。

被选为目标变量的列会自动从预测变量列表中排除。

6.2 预测变量表

在目标变量选择器下方，一个表格列出其余每一列，包含以下字段：

列	说明
Variable （变量）	列名（若被截断，完整名称显示在工具提示中），前面带有一个 ×/+ 切换按钮（见下文）。在估价/市场模式下，蓝色徽章显示所分配的特殊类型。
Type （类型）	数据类型下拉框：numeric、integer、character、logical、factor、Date、POSIXct
Include （纳入）	复选框——将此列作为预测变量纳入模型
Factor （因子）	复选框——将此列视为分类变量
Linear （线性）	复选框——强制仅以线性方式进入模型（不生成铰链函数）
Latent （潜变量）	复选框——将此列完全排除在模型之外（既不进入 earth，也不包含在导出到 glmnetUI / mgcvUI 的数据中），保留用于潜变量分析
Special （特殊类型）	下拉框（仅估价/市场模式）——见下文“特殊列类型参考”
T/F/NA	每列三个计数：对逻辑列为 TRUE/FALSE/NA；对数值列为 ≠ 0/= 0/NA。当缺失值超过 30% 时以红色显示

表格上方的提示行解释了这些缩写：“Type = 列数据类型，Inc = 作为预测变量纳入，Factor = 视为分类变量，Linear = 仅线性（无铰链），Special = 列角色（例如 contract_date）。”

软禁用而非删除。每个变量名旁的 × 切换按钮会将该行变灰，并使该列不参与模型，但不会将其从数据中删除；按钮随之变为绿色的 +，可重新启用该列。列永远不会被销毁。

6.3 数据类型检测与覆盖

earthUI 在导入时自动检测数据类型。数值、整数、逻辑、因子和日期列都会被识别。看起来像日期的字符列（至少 80% 的值能按常见日期格式解析）会被归类为 Date。

你可以通过更改 **Type** 下拉框来覆盖任何检测结果。当你将某列改为 character 或 factor 时，Factor 复选框会自动勾选。更改类型会影响该列传递给 earth() 函数的方式。

6.4 Factor 与 Linear 标志

- **Factor**——勾选后，该列被视为分类变量。这适用于 style、area_id 或 grade 等表示离散组别而非连续测量值的列。因子列以指示（哑）变量的形式进入模型。
- **Linear**——勾选后，该列被强制以线性方式进入模型——不会为其创建铰链（分段线性）函数。当你确定某个变量与目标变量之间存在严格的线性关系时，此选项很有用。

6.5 特殊列类型参考

在估价和市场模式下，**Special** 下拉框提供以下选项。每种类型最多只能分配给一列（**display_only**、**remarks** 和 **list** 除外，它们允许多列）：

日期与时间类型：

- **contract_date**——触发自动计算 **sale_age**。原始日期列被替换为一个整数列，度量销售日期与 Effective Date（评估基准日）之间的天数。
- **listing_date**——当未指定明确的 **dom** 列时，用作计算在市天数（DOM = 合同日期 - 挂牌日期）的后备来源。
- **dom**——标识在市天数（Days on Market）列。显示在 Sales Grid（销售比较网格）的 APN 行和 Date of Sale 行中。

货币类型：

- **concessions**——标识销售让步（卖方信贷、买方激励等）。在 Sales Grid 中用于计算净售价：Net SP = 售价 - 让步。

面积与位置类型：

- **latitude**——值自动四舍五入到小数点后 3 位以防止过拟合。用于 Haversine 邻近度计算（估价对象（subject）到每个可比销售案例的距离），并归入 Sales Grid 的 Location 行。
- **longitude**——与 latitude 相同的舍入处理。用于邻近度计算和 Location 组。
- **area**——通常为社区或区域标识符。与 latitude、longitude 一起归入 Sales Grid 的 “Loc: Long | Lat | Area” 行。
- **living_area**——启用下载输出中的每平方英尺残差计算（**residual_sf** 和 **cqa_sf**）。
- **lot_size**——归入 Sales Grid 的 “Site Size | Dimensions” 行。
- **site_dimensions**——与 lot size 一起归入 Sales Grid（例如 “75x120”）。

房龄类型：

- **actual_age**——归入 Sales Grid 的 “Actual Age | Effective Age” 行。
- **effective_age**——与 actual age 一起归入 Sales Grid。

显示类型：

- **display_only**——该列包含在 Excel 导出中，但完全不参与模型拟合。适用于地址字段、MLS 编号、宗地编号或其他不应作为预测变量的参考数据。允许多列使用此指定。

文本提取类型（实验性）：

- **remarks**——一个自由文本列（例如 **public_remarks**），由可选的 Prolog 步骤解析为特征列。在 Settings（设置）中启用该功能之前不起任何作用。允许多列。参见第 15 章。
- **list**——一个逗号分隔的列（例如 **existing_structures**），由同一步骤拆分为独热编码的 **<col>_<value>** TRUE/FALSE 列。允许多列。参见第 15 章。

6.6 多个目标变量

选择多个目标变量会拟合一个多响应 Earth 模型。选择多个目标变量时：

- 模型同时预测所有响应，各目标变量共享基函数
- **wp (response weights)**（响应权重）按钮被激活，允许为每个目标变量分配数值权重
- 方差模型（**varmod.method**）被禁用（多响应模型不支持）
- 结果标签页会显示每个响应的指标、方程和诊断图

提示：指定为 **display_only** 的列保留在数据集中并出现在 Excel 导出中，但完全不出现在预测变量表中。适用于 ID 列、地址或其他参考字段。

7 参数选择

侧边栏第 4 节——**Earth Call Parameters** (Earth 调用参数)——提供对 `earth()` 函数所接受的全部参数的访问。每个参数都带有一个蓝色帮助图标 (?) 及工具提示说明。参数按可折叠的小节组织。

7.1 基本参数

参数	默认值	说明
subset	(空)	行过滤表达式。见下文“子集过滤”。
weights	NULL	用于个案 (行) 权重的列选择器。仅列出数值列。
wp	NULL	多目标模型的响应权重。按钮会打开一个对话框，为每个目标变量提供一个数值输入 (默认各为 1.0)。
keepxy	关闭	在模型对象中保留 x、y、subset 和 weights。
trace	0	拟合输出的跟踪级别 (0 到 5)。
glm	无	可选的 GLM 族: gaussian、binomial 或 poisson。
degree	1	最大交互阶数。将 degree 设为 ≥ 2 会自动启用交叉验证，并显示 Allowed Interactions (允许的交互) 矩阵。
penalty	2	每个节点的 GCV 惩罚。值越高，模型越简单。

7.2 前向过程

参数	默认值	说明
nk	自动	剪枝前的最大模型项数。自动填充为所加载数据的推荐值 (见下文“推荐参数值”)。
thresh	0.001	前向步骤阈值。值越小，允许的模型项越多。
minspan	自动	节点之间的最小间隔。自动填充为推荐值。负值设置每个预测变量的最大节点数。
endspan	自动	端部间隔——节点到数据边缘的最小距离。自动填充为推荐值。
newvar.penalty	0	引入新变量的惩罚 (鼓励重用已有的预测变量)。
fast.k	20	快速 MARS 算法中考虑的父项数量。
fast.beta	1	控制快速 MARS 的老化因子。

7.3 允许的交互

当 **degree** 设为 2 或更高时，基本参数下方会出现一个交互矩阵。它是一个上三角形的复选框网格，每个预测变量对对应一个复选框。勾选表示允许这两个预测变量交互；取消勾选则禁止该特定交互。

点击预测变量名称（行或列标签）可切换该变量的所有交互。顶部的 **Allow All**（全部允许）和 **Clear All**（全部清除）复选框提供批量控制。矩阵使用固定表头，滚动时行、列标签保持可见。

一条信息提示会提醒你：“交互项会增加过拟合的风险。交叉验证已启用（10 折）。”

7.4 剪枝

参数	默认值	说明
pmethod	backward	剪枝方法：backward、none、exhaustive、forward、seqrep 或 cv。
nprune	NULL	剪枝后的最大模型项数。留空则自动选择。

7.5 交叉验证

参数	默认值	说明
nfold	10	交叉验证折数。设为 0 可禁用交叉验证。
ncross	20	交叉验证重复次数。
stratify	开启	对交叉验证样本进行分层，使每一折具有相似的响应分布。

当 $\text{degree} \geq 2$ 时，交叉验证会自动启用（**nfold** 设为 10），以帮助防范交互项带来的过拟合。

7.6 子集过滤

subset 文本输入框接受一个 R 表达式，用于筛选哪些行参与模型拟合。你可以直接输入表达式（例如 `sale_age < 365 & area_id == 460`），也可以使用 **Build filter...**（构建过滤器…）按钮以可视化方式构建。

Build Filter Dialog（构建过滤器对话框）——点击“Build filter...”打开一个引导式对话框。每个条件行包含一个列下拉框、运算符（<、>、<=、>=、==、!=），以及一个根据列类型自适应的值输入框：数值列为数字输入，日期列为日期选择器，字符/因子列为唯一值下拉框。各条件通过 **AND**（&）或 **OR**（|）连接符相连。底部的预览会显示表达式以及匹配的行数。点击 **Apply**（应用）将表达式插入文本输入框。

日期列在手动表达式中必须使用 `as.Date("...")` 或 `as.POSIXct("...")` 包装。Build Filter 对话框会自动处理这一点。

子集过滤是非破坏性的——被排除的行仍保留在数据集中，并在 Excel 导出中获得预测值。

提示：将 **OR** 条件与 **AND** 组合时，请使用括号进行分组。若不加括号，`A & B | C` 会被解释为 `(A & B) | C`，这可能不符合你的本意。

7.7 推荐参数值

earthUI 会在第 4 节的关键参数下方显示推荐值。这些推荐值会根据拟合行数 (n) 和所选预测变量数 (p) 动态更新。这些公式来源于 Friedman 的 MARS 论文、earth 的内部算法以及经验测试。**nk**、**minspan** 和 **endspan** 字段会在数据加载时预填充其推荐值；手动编辑的值会被保留。

7.7.1 前向过程参数

nk (剪枝前的最大模型项数):

$$\text{nk} = \min(100, \max(21, 2p + 1, \lfloor n/10 \rfloor))$$

n	$p=5$	$p=10$
30	21	21
50	21	21
100	21	21
200	21	21
500	50	50
1000	100	100
1500	100	100

Earth 的默认值 $\min(200, \max(20, 2p)) + 1$ 未考虑数据集规模，对大数据集而言可能过小，会在前向过程充分探索所有预测变量之前就对其加以限制。 $\lfloor n/10 \rfloor$ 一项允许大约每 10 个观测值对应一个模型项——这是避免过度参数化的标准经验法则。100 的上限可避免收益递减。

minspan (节点之间的最小观测数):

$$\text{minspan} = \min(16, \lfloor 5 + n/50 \rfloor)$$

endspan (距数据边界的最小观测数):

$$\text{endspan} = \min(16, \lfloor 5 + n/28 \rfloor)$$

n	minspan	endspan
30	5	6
50	6	6
100	7	8
200	9	12
300	11	15
500	15	16
1500	16	16

Earth 自动计算的 minspan 使用 Friedman 的公式 43: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, 它按 $\ln(np)$ 缩放，对大数据集增长过慢。在 $n=1500$ 时它仅得出 ≈ 7 ，使每个连续预测变量约有 245 个候选节点位置——太多了，会让前向过程去拟合噪声。推荐公式的目标是在大 n 时每个预测变量约有 100 个候选位置，而在小 n 时（此时 earth 的自动计算经过了充分检验）沿用其自动计算结果。endspan 设置得略大于 minspan，以提供额外的边界保护，这在数据尾部稀薄时（房地产数据中很常见）很有帮助。

penalty (每个节点的 GCV 惩罚):

$$\text{penalty} = \begin{cases} 3 & \text{if degree} > 1 \\ 2 & \text{if degree} = 1 \end{cases}$$

这正是 earth 自身的默认值——加性模型为 2，交互模型为 3（更高的惩罚补偿了更大的搜索空间）。

newvar.penalty（引入新预测变量的惩罚）：

推荐值：当预测变量之间存在相关性时（例如居住面积与卧室数、浴室数、地块面积相关）取 **0.1**。这会使前向过程倾向于重用模型中已有的预测变量，而不是引入相关的替代变量。其机制是：每个新预测变量的 RSS 改进量会乘以 $1/(1 + \text{penalty})$ 。当 $\text{newvar.penalty} = 0.1$ 时，新变量的 RSS 改进必须比在已有变量上再加一个节点至少高出 10%。这会产生预测变量更少、更简单的模型，且不影响最终的系数估计（惩罚在选择完成后即被移除）。若预测变量之间不相关，则保持为 0。注意，earth 不支持在使用个案权重的同时设置非零的 **newvar.penalty**——一旦指定了权重列，earthUI 就会禁用该输入（强制为 0）。

7.7.2 剪枝参数

pmethod：推荐值：**backward**。默认的基于 GCV 的后向剪枝是确定性的——相同的数据总是产生相同的模型，不依赖任何随机种子。这对可复现性很重要，在估价工作中尤其如此。cv 方法使用交叉验证进行剪枝，会引入对种子的依赖。

nprune：推荐值：留空（NULL）。让 GCV 选择最优的模型项数。设置 **nprune** 会施加一个硬性上限，凌驾于 GCV 的判断之上。

7.7.3 交叉验证参数

nfold（交叉验证折数）：按样本量区间推荐：

n （拟合行数）	nfold
0–299	5
300–399	6
400–499	7
500–599	8
600–699	9
700+	10

折数越多，每一折的训练数据越多（估计偏差更小），但留出的测试折会随之缩小。该表随着样本增大逐级升至标准的 10 折，同时保持每个测试折足够大（约 50 行以上）以获得稳定的单折指标；对小样本则回退到 5 折——此时 CVR^2 估计仍有噪声，因此应提高 **ncross**（例如 15–20）以平均掉方差。上限 10 反映的是：每增加一折都是一次完整的 earth 拟合，而诊断收益微乎其微。当 **pmethod** = "backward" 时，交叉验证不影响模型——它只计算诊断性的 CVR^2 ，并为方差模型提供残差。

ncross（交叉验证重复次数）：按样本量区间推荐：

n （拟合行数）	ncross
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

重复可以平均掉折划分中的运气成分，从而稳定 CVR^2 ；样本越大所需重复越少，因为单次交叉验证本身已经更稳定。重复交叉验证的残差还会输入方差模型。

7.7.4 方差模型

varmod.method: 推荐值: **lm**——Milborrow 的默认值：用于预测区间的稳健线性方差模型（离散度是从带噪声的残差中估计的，因此简单模型的泛化能力更好）。如果 **lm** 在你的数据上无法收敛，可尝试 **rlm** 或 **const**（同方差，总能收敛），或退回到 **earth**（也能处理非线性的离散度）。它要求 $nfold > 1$ 且 $ncross > 1$ 。

7.7.5 Earth 输出文件

每次成功拟合后，earthUI 会将 `<filename>_earth_output_<timestamp>.txt` 写入输出文件夹。该文件包含模型项、汇总统计量（ R^2 、 $GRSq$ 、 $CVRSq$ ）、方差模型详情和跟踪日志。每次拟合创建一个文件，为比较不同参数配置提供累积记录。

7.8 设置默认值

第 4 节顶部的单选按钮控制加载哪一组默认值：

- **Use last settings for input file**（使用该输入文件的上次设置）——恢复你上次针对这个特定文件使用的设置
- **Use default settings**（使用默认设置）——应用你保存的自定义默认值
- **Earth defaults**（Earth 默认值）——将所有参数重置为出厂值

点击第 4 节底部的 **Save current as default**（将当前设置保存为默认）以存储当前的 **earth** 参数和 Allowed Interactions 矩阵；第 3 节中对应的按钮存储目标变量、评估基准日和预测变量配置。两者都按项目和用途保存在项目数据库中。

8 拟合 Earth 模型

8.1 Fit 按钮

侧边栏第 5 节包含 **Fit Earth Model** (拟合 Earth 模型) 按钮, 以及一个 **Random seed** (随机种子) 字段。该种子使交叉验证折的分配具有可重现性——下拉列表可调出最近使用过的 10 个种子, 拟合完成后所用的种子会记录在 Earth Output (Earth 输出) 选项卡中。在拟合之前, earthUI 会验证您的配置——必须选定一个目标变量, 并且至少包含一个预测变量。在估价/市场模式下, 纬度和经度列会被四舍五入, 销售时距则根据估价基准日和合同日期 (若已指定) 计算得出。

8.2 拟合模态框与跟踪输出

当您点击 Fit 时, 会出现一个深色模态覆盖层, 其中包含:

- 标题栏——显示“Fitting Earth Model”, 附带以秒计数的已用时间计时器, 以及一个可取消后台拟合的 **Abort** (中止) 按钮
- 跟踪日志——一个可滚动的等宽文本区域, 实时显示 earth() 函数的输出: 数据集信息、前向遍历进度、交叉验证折以及完成时间
- 颜色编码——默认行为绿色, 交叉验证行为黄色, 错误为红色

earthUI 使用 callr::r_bg() 异步拟合模型, 将计算放在单独的 R 进程中运行。这使应用程序在长时间拟合期间保持响应。一个 300 毫秒的轮询观察器从后台进程读取输出并将其流式写入跟踪日志。

拟合完成后, 会出现“Done in X.Xs”消息, 模态框上会添加一个关闭按钮 (X), 并移除背景遮罩, 以便结果选项卡可以按需渲染。待可见输出渲染完毕后, 模态框会自动关闭——您也可以提前关闭它。

拟合成功时, Fit 按钮上会附加一个绿色对勾。如果发生错误, 错误消息会显示在跟踪日志中, 并弹出一条通知。

8.3 拟合日志

每次拟合 (无论成功或失败) 都会在您的输出文件夹中写入一个日志文件:

- 文件名: <datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt
- 内容: 时间戳、拟合过程中的全部跟踪输出, 以及所有错误消息
- 位置: 当前活动项目的输出文件夹

提示: 如果未安装 callr, earthUI 会回退到带简单进度条的同步拟合。安装 callr 可获得完整的跟踪输出体验。

9 下载数据

拟合完成后，侧边栏第 6 节——在通用模式下标题为“Download Estimated Target Variable(s) & Residuals”（下载估计的目标变量与残差），在估价和市场模式下标题为“Download Estimated Sale Prices & Residuals”（下载估计的销售价格与残差）——可下载一个包含预测值和诊断信息的 Excel 文件。该输出用于第 7 步（RCA），为估价对象（subject）指定 CQA（状况—质量—吸引力）评分。输出按 `residual_sf` 和 `cqa_sf` 排序，以帮助您评估估价对象在排名中的位置。如果模型质量良好，那么各房产应按照未纳入回归的残差特征，从吸引力最低到吸引力最高进行排序。中间值应约为 0，下半部分为负值，上半部分为正值。您应当在排名底部附近发现质量最差的住宅，即“待修房”（fixers），而最好的住宅位于顶部。通常会有一些异常情形的例外，例如法拍、亏本急售（short sale）、遗嘱认证（与继承相关的）销售，以及因工作变动或其他原因需要的快速出售。对异常案例的调查通常会揭示出价格异常的相关原因。

按钮标签会根据用途模式而变化：

- 通用/市场：**Download Output (Excel)**（下载输出（Excel））
- 估价：**Download Intermediate Output (Excel)**（下载中间输出（Excel））

文件名格式为 `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`。

9.1 输出列

对于每个目标变量，将追加以下各列：

列名	说明
<code>est_<target></code>	模型预测值，例如 <code>est_sale_price</code> （保留 1 位小数）
<code>residual</code>	实际值减去预测值（1 位小数）
<code>cqa</code>	状况—质量—吸引力评分（2 位小数，范围 0–10）
<code>residual_sf</code>	残差除以居住面积（若已指定，1 位小数）
<code>cqa_sf</code>	通过 <code>residual_sf</code> 排名计算得出的 CQA（2 位小数）
<code><variable>_contribution</code>	每个 g 函数的贡献值（1 位小数，每个 g 函数一列）
<code>basis</code>	截距贡献值，对所有房产相同（1 位小数）
<code>calc_residual</code>	实际值减去（ <code>basis</code> + 所有贡献值）——验证列（1 位小数）

对于多目标模型，会添加 `_<i>` 后缀以区分各目标对应的列。

9.2 列排序与 Excel 格式

排名列被放置在输出文件的最左侧位置，顺序为：`residual_sf`、`cqa_sf`、`residual`、`cqa`。这样便于浏览排序后的列表，并评估估价对象在 CQA 分布中的位置。

这些列应用了以下 Excel 数字格式：

列名	格式	示例
<code>residual_sf</code>	数值，2 位小数	12.34
<code>cqa_sf</code>	数值，2 位小数	7.25
<code>residual</code>	数值，0 位小数	15,200
<code>cqa</code>	数值，2 位小数	6.80

9.3 CQA 评分

CQA（状况—质量—吸引力）评分将每一行的残差与所有其他残差进行排名比较。对于给定的行，CQA 是带符号残差小于该行的行所占的百分比乘以 10。由此得到一个 0–10 的评分尺度，其中：

- 高 CQA ($\approx 9-10$) ——该房产的成交价远高于模型预测（较大的正残差）
- 低 CQA ($\approx 0-1$) ——该房产的成交价远低于预测（较大的负残差）
- CQA ≈ 5 ——残差接近所有残差的中位数

当指定了 `living_area` 列时，`cqa_sf` 提供基于每平方英尺残差的同类排名。

9.4 排序与估价对象行

在估价和市场模式下，可比销售案例行按 `residual_sf` 降序排序（若未指定居住面积则按 `residual` 排序）。估价对象所在行（第 1 行）保持在第 1 位，不参与排序。在通用模式下，各行按其原始顺序导出。

这种排序方式与最左侧的排名列相结合，使估价师能够快速浏览从预测偏高最多到预测偏低最多的可比销售案例，并评估估价对象被赋予的 CQA 评分在该分布中的位置。

在调用 `predict()` 之前，预测数据中的因子水平会与训练数据对齐。含有未见过的因子水平的行将产生 NA 预测值。

提示：`calc_residual` 列应始终等于 `residual`（在舍入误差范围内）。如果两者差异显著，则表明存在值得调查的计算问题。

9.5 mgcvUI 自动导出

每次成功拟合且 $\text{degree} \leq 2$ 时，`earthUI` 会自动将完整的结果对象保存为 `.rds` 文件，存入项目的输出文件夹。文件名遵循如下模式：`<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`。

该文件可由 `mgcv` 后处理模块加载——该模块包含在 `earthUI` 包中，通过 `earthUI::launch_mgcv()` 启动——使用 `readRDS()` 读取。GAM 例程使用 `earth` 模型的节点位置和基函数作为 GAM 平滑项的起点，从而实现从 MARS 到 GAM 建模的无缝过渡。

$\text{degree} > 2$ 的模型会被跳过，因为 `mgcvUI` 仅支持两两交互。当 `earthUI` 以 Trilogy 模式（从 Trilogy UI）启动时，额外的 **Lock Model Output**（锁定模型输出）区块可让您将某次拟合的 `.rds`——连同其预测变量基和共享的 CQA 设置——锁定为供 `glmnetUI` 和 `mgcvUI` 导入的规范 `earth` 模型。

9.6 glmnetUI 导入

为 GAM 例程保存的同一 `.rds` 文件也可导入弹性网络后处理模块——同样包含在 `earthUI` 包中，通过 `earthUI::launch_glmnet()` 启动。在 `glmnet` 应用中，导航至 **Section 2: Import from earthUI**（第 2 节：从 `earthUI` 导入），使用 Browse（浏览）按钮选择该 `.rds` 文件。

`glmnetUI` 采用标准的 `earth::model.matrix()` 方法：`earth` 模型的基函数（铰链、交互项和线性项）成为 `glmnet` 设计矩阵的列。`glmnet` 随后对这些基列执行正则化回归，通过 lasso/弹性网络对其进行选择和收缩。这将 `earth` 的自适应基构造与 `glmnet` 的正则化相结合。

9.6.1 主要优势

- 可解释的交互项：`earth` 基于铰链的交互项（包括三阶项）具有清晰的几何含义，而 `glmnet` 原生的叉积交互项在法庭或审计场合难以解释。
- 自动非线性：`earth` 的铰链函数可以捕捉使用原始预测变量的标准 `glmnet` 模型会遗漏的非线性关系。
- 变量选择：`glmnet` 可以进一步修剪 `earth` 基，剔除没有贡献的铰链项。

9.6.2 重要提示：权重列必须一致

警告

如果 earthUI 使用了权重列（例如，用 `weight=0` 排除估价对象所在行），则必须在 glmnetUI 的变量配置中指定相同的权重列。如果权重设置不同，行数将无法对齐，earth 基也就无法附加到 glmnet 的模型矩阵中。

9.6.3 工作流程

1. 在 earthUI 中拟合一个 earth 模型（任意阶数）。
2. `.rds` 结果文件会自动保存到项目的输出文件夹。
3. 在 glmnetUI 中，打开 **Section 2: Import from earthUI**，浏览至该 `.rds` 文件。
4. 确认两个应用中使用的是同一数据文件和同一权重列。
5. 在第 5 节中配置 glmnet 参数（当 earthUI 导入处于活动状态时，交互矩阵会自动禁用）。
6. 点击 **Fit Model**（拟合模型）。earth 基列将出现在 Equation（方程）、Coefficients（系数）和 Contributions（贡献）选项卡中。

10 RCA 计算与下载

Calculate RCA Adjustments & Download (计算 RCA 调整并下载) 按钮位于侧边栏第 7 节, 仅在估价模式下且模型拟合完成后可见。RCA (Residual Constraint Approach, 残差约束法) 生成相对于估价对象 (subject) 的、源自市场的逐个可比案例调整。

10.1 打开 RCA 对话框

点击该按钮会打开一个小型模态对话框, 其中包含:

- **Score type** (评分类型) —— 单选按钮, 用于选择 **CQA** 或 **CQA per SF** (仅当已指定 `living_area` 列时, SF 选项才可用)。如果选择“CQA per SF”, 则根据你为估价对象指定的 CQA 评分, 其残差评分将是其居住面积乘以与给定 CQA_SF 评分匹配的 `residual_sf`。
- **CQA value** (CQA 值) —— 用于输入估价对象的 CQA 评分 (0.00–10.00)。该评分表示你认为估价对象在可比案例质量分布中所处的位置。
- **Generate** (生成) 按钮——计算 RCA 输出并以 Excel 格式下载。在输入有效评分之前, 该按钮保持禁用状态。

10.2 CQA 评分插值

当你点击 Generate 时, earthUI 会根据可比案例的 CQA/残差对来插值估价对象的残差:

1. 将可比案例的 CQA 评分和残差按 CQA 升序排序
2. 通过线性插值 (`stats::approx()`) 将你输入的 CQA 值映射为一个残差
3. 如果使用 CQA per SF, 则将每平方英尺残差乘以估价对象的居住面积, 换算回总残差
4. 估价对象的估计价值计算为: 模型预测值 + 插值残差

10.3 输出列

RCA Excel 文件 (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) 包含所有中间输出列, 另加:

列	说明
<code>subject_value</code>	模型预测值 + 插值残差 (仅第 1 行)
<code>subject_cqa</code>	你输入的 CQA 评分 (仅第 1 行)
<code><variable>_adjustment</code>	估价对象贡献减去可比案例贡献 (按 g 函数逐项计算)
<code>residual_adjustment</code>	估价对象残差减去可比案例残差
<code>net_adjustments</code>	所有调整之和 (贡献 + 残差)
<code>gross_adjustments</code>	所有调整绝对值之和
<code>residual_pct</code>	残差调整占可比案例售价的比例
<code>net_adj_pct</code>	净调整占可比案例售价的比例
<code>gross_adj_pct</code>	总调整占可比案例售价的比例
<code>adjusted_sale_price</code>	可比案例售价 + 净调整

各调整列告诉你, 对于每个可比案例, 模型将差异归因于各变量的金额。`adjusted_sale_price` 是应用所有模型推导调整后的可比案例售价——这组调整后售价应聚集在估价对象的估计价值附近。

10.4 RCA Adjustments 选项卡

RCA 输出计算完成后，主面板中的 **RCA Adjustments** (RCA 调整) 选项卡会显示各可比案例的 Residual Adj %、Net Adj % 和 Gross Adj % 分布的直方图 (多目标模型则每个目标一组)，从而可以快速直观地检查调整的聚集程度。

提示：CQA 评分 5.00 表示估价对象处于可比案例的中位数位置。评分高于 5 表示估价对象质量高于平均水平；低于 5 则表示低于平均水平。

11 销售比较网格

11.1 概述与用途

销售比较网格 (Sales Comparison Grid) 是在估价模式下 (侧边栏第 8 节)、计算 RCA 调整之后生成的 Excel 工作簿。它以结构化的网格形式将估价对象与选定的可比案例并排呈现, 并附带自动计算调整和每个可比案例调整后售价的 Excel 公式。

该网格专为估价师的工作底稿而设计——它将 Earth 模型推导出的回归调整与可编辑单元格相结合, 估价师可在其中将 CQA 残差分配到具体的房产特征上。输出文件名为 SalesGrid_<YYYYMMDD_HHMMSS>.xlsx。

11.2 可比案例选择模态框

点击“Generate Sales Grid & Download”(生成销售网格并下载) 会打开一个模态对话框, 供你选择要包含的可比案例:

- 推荐可比案例已预先勾选。这些可比案例的总调整百分比低于售价的 25%, 按销售时间由近及远排序 (最近的销售在前), 上限为 30 个。
- 其他可比案例列在下方, 默认不勾选。它们的总调整达到或超过 25%, 但仍可供选择 (最多显示 50 个, 按总调整百分比升序排序)。
- 最多可选择 30 个可比案例, 最多生成 10 张工作表 (每张工作表 3 个可比案例)。

确认选择后, 所选可比案例按总调整百分比升序排序, 因此最相似的可比案例会出现在 Sheet 1 上。

11.3 网格布局

每张工作表采用 20 列布局, 容纳估价对象加 3 个可比案例。各实体的列为:

实体	列 (每个实体 5 列)
估价对象	标签、事实值 1、事实值 2、事实值 3、价值贡献 (VC)
每个可比案例	事实值 1、事实值 2、事实值 3、价值贡献 (VC)、调整

各行从上到下依次为:

1. 标题——工作表名称 (例如“Intermediate Sales Comparable Grid — Sheet 1 of 3”)
2. 表头——“Subject”与各可比案例的列标题及地址
3. 地址——完整的房产地址
4. APN | MLS# | DOM | Subj.Prox (宗地号 | 挂牌编号 | 在售天数 | 与估价对象的距离)——宗地号、挂牌编号、在售天数, 以及与估价对象的 Haversine 距离 (英里)
5. Sales Price | Concess. | Net SP (售价 | 让价优惠 | 净售价)——售价、让价优惠, 以及 Net Sale Price (净售价) 公式
6. Regression Features (回归特征) 表头行
7. BASE VALUE (基准值)——模型截距
8. Date of Sale | OffMkt | OnMkt (成交日期 | 下市 | 上市)——合同日期、销售距今时间和 DOM
9. 分组行 (视情况而定)——Location (位置)、Site Size (地块面积) 和/或 Age (房龄) 组
10. 模型变量行——每个预测变量一行 (不含已分组的变量)
11. 空白分隔行
12. Residual Features (残差特征) 表头行
13. CQA|Residual (CQA | 残差)——CQA 评分 + 剩余残差公式
14. 残差特征行——命名行加空白行, 供估价师填写

15. **Total VC / Net Adjustment** (总 VC / 净调整)
16. **Net Adjustment %** (净调整百分比)
17. **Gross Adjustment %** (总调整百分比)
18. **Adjusted Sale Price** (调整后售价) ——公式行
19. 版权页脚

11.4 售价、让价优惠与 Net SP

第 5 行为每个可比案例显示三个值:

- **Sale Price**——来自数据的可比案例售价
- **Concessions**——指定为 **concessions** 的列中的值 (未指定时为 0)
- **Net SP**——一个 Excel 公式: $\text{Sale Price} - \text{Concessions}$

估价对象列的售价显示为“N/A”(因为未知), 若有让价优惠值则一并显示。

11.5 分组行 (位置、地块、房龄)

当指定了某些特殊类型且这些变量出现在模型中时, earthUI 会创建将相关变量合并的分组行:

Loc: Long | Lat | Area (位置: 经度 | 纬度 | 区域) ——当 **longitude**、**latitude** 或 **area** 中的任一变量出现在模型中时出现。显示每个组成变量的事实值、合并的价值贡献 (各 VC 之和), 以及针对可比案例的调整 (估价对象合并 VC - 可比案例合并 VC)。VC 单元格以浅蓝色背景标示。

Site Size | Dimensions (地块面积 | 尺寸) ——当 **lot_size** 或 **site_dimensions** 出现在模型中时出现。结构与 **Location** 组相同。

Actual Age | Effective Age (实际房龄 | 有效房龄) ——当 **actual_age** 或 **effective_age** 出现在模型中时出现。结构相同。

被分组行占用的变量将从下方的单个模型变量行中排除, 以防止调整合计中的重复计算。

11.6 模型变量行

在分组行下方 (若无分组行, 则直接在 **BASE VALUE** 下方), 每个剩余的模型预测变量占一行, 显示:

- **事实值**——估价对象和每个可比案例的实际数据值
- **价值贡献 (VC)** ——该行 **g** 函数对预测值的贡献
- **调整 (仅可比案例)** ——估价对象 VC 减去可比案例 VC

11.7 CQA|Residual 与残差特征行

CQA|Residual 行显示每个房产的 CQA 评分, 并包含剩余残差公式——即尚未分配到具体特征的残差部分。公式为:

$$\text{Remaining Residual} = \text{Total Residual} - \sum (\text{Residual Feature VCs below})$$

其下是残差特征行——5 个命名行 (**Location**、**View/Appeal**、**Condition**、**Quality**、**Other**, 即位置、景观/吸引力、状况、品质、其他) 加 6 个空白行。这些是输入单元格, 供估价师为模型未捕捉到的特征填写价值贡献。随着估价师填入数值, **Remaining Residual** 公式的值会自动减少。

对于每个可比案例, 调整列包含一个公式: 估价对象特征 VC - 可比案例特征 VC。

11.8 调整后售价公式

Adjusted Sale Price 行包含 Excel 公式:

- 估价对象：从 BASE VALUE 到最后一个残差特征行的所有价值贡献单元格之和。这代表模型对估价对象的总预测值，加上估价师分配的残差特征。
- 可比案例：Net SP + 从第一个分组行或变量行到最后一个残差特征行的所有调整单元格之和。这给出应用所有回归推导调整和估价师填写调整后的可比案例售价。

11.9 工作表保护与可编辑单元格

每张工作表均已受保护，以防止公式和数据被意外修改。保护锁定了：

- 所有公式单元格（Net SP、剩余残差、调整、调整后售价等）
- 所有数据单元格（地址、售价、事实值、来自模型的价值贡献）
- 标签和表头

唯一未锁定（可编辑）的单元格是残差特征价值贡献输入格——即 CQA|Residual 行下方供估价师填写分解值的单元格。这些单元格以浅黄色背景标示，表示可编辑。

11.10 在 Excel 中使用网格

1. 打开文件——使用 Excel 或兼容的电子表格应用程序。
2. 审核回归调整——模型推导的调整已预先填入并锁定。
3. 分配残差——在残差特征行（黄色单元格）中，填写你对剩余残差中归因于各特征（位置、景观、状况、品质等）的金额判断。随着分配的进行，Remaining Residual 公式的值会相应减少。
4. 检查 **Adjusted Sale Price**——该公式会随你填入残差特征值而自动更新。
5. 多张工作表——如果选择了 3 个以上的可比案例，请在工作表之间切换。每张工作表的左列均为同一估价对象，并配以 3 个不同的可比案例。

提示：目标是将 CQA 残差分配到各特征行，直至 *Remaining Residual* 接近于零。这表明你能够解释每个可比案例的回归估计值与其实际售价之间的差异。

12 下载报告

模型拟合完成后，报告生成是一个两步式的 Quarto 工作流：**Generate Quarto Report**（生成 Quarto 报告，估价模式下为侧边栏第 9 节，其他模式下为第 7 节）写出一个自包含的报告包，而 **Convert Quarto Report**（转换 Quarto 报告，第 10 节）将其——或任何其他 Quarto 文件——渲染为最终格式。

12.1 生成 Quarto 报告

点击 **Generate Quarto Report** 会在项目输出文件夹下的 `<base>_qmd/` 中写出一个自包含的 Quarto 包，其中包含已填充内容的 `.qmd` 源文件、所有预生成的图表资源（PNG 和 PDF）、报告数据，以及一个 Word 参考模板。此步骤不进行任何渲染——该报告包可以手工编辑、通过 Quarto includes 与其他 Quarto 源文件合并，或使用 `quarto preview` 单独预览。

12.2 转换 Quarto 报告

第 10 节提供一个文件选择器（默认指向当前项目最近生成的报告包——但你可以浏览到任意 `.qmd` 文件）以及格式复选框：

- **HTML**——一个 `.html` 文件，带有 KaTeX 数学公式渲染、内嵌图片和目录。使用 Flatly Bootstrap 主题和 Roboto Condensed 字体。
- **Word**——一个适合编辑和分发的 `.docx` 文件。包含目录、页码，并使用自定义参考模板以保持样式一致。
- **PDF**——使用 LuaLaTeX 排版，输出达到专业质量。纸张尺寸遵循区域设置（Letter 或 A4）。包含用于大型交互矩阵的横向页面，并使用 Roboto Condensed 搭配 Latin Modern Math 排版公式。仅当检测到 LaTeX 发行版时才会提供 PDF 选项。

点击 **Convert**（转换）以渲染所选格式。渲染在后台运行——一个模态对话框显示已用时间和 Quarto 进度，应用保持可响应状态。完成后会有通知确认文件位置。

所有操作（模型拟合、数据下载、RCA 计算、销售网格生成和报告渲染）都会连同起止时间戳和耗时记录到输出文件夹中的 `<datafile>_earthui_log.txt`，以便排查问题和监控性能。

12.3 报告内容

每份生成的报告都包含以下各节：

1. 数据集描述——观测数量、目标变量、所用预测变量、分类预测变量
2. 模型规格——阶数（degree）、交叉验证状态、项数、模型中的预测变量
3. 允许的交互项（仅当 $\text{degree} \geq 2$ 时）——带勾选标记的交互矩阵，按输出类型格式化
4. 结果摘要—— R^2 、GRSq、GCV、RSS，以及 $\text{CV } R^2$ （若启用了交叉验证）
5. 模型方程——以 LaTeX 记号表示的完整 Earth 模型方程
6. 系数与基函数——所有项、其系数以及定义它们的基函数的表格
7. 变量重要性——预测变量重要性得分的条形图和排名表
8. g 函数贡献——每个 g 函数的图：单变量项为折线图，双变量交互项为 3D 透视图和等高线图
9. 相关矩阵——预测变量相关性的热图
10. 诊断——残差对拟合值（Residuals vs Fitted）、正态 Q-Q 图，以及实际值对预测值（Actual vs Predicted）图
11. ANOVA 分解——显示每个基函数对 RSS 贡献的表格
12. Earth 输出——来自 `earth::summary()` 的原始文本输出，包括剪枝过程的细节

对于多响应模型，第 4–8 节和第 10 节会针对每个目标变量重复一次。

提示：生成报告需要 `quarto R` 包和 `Quarto` 安装。若要输出 *PDF*，还必须安装 *LaTeX* 发行版（例如通过 `tinytex::install_tinytex()`）——否则 *PDF* 复选框会被隐藏。

13 后处理模块：弹性网络与 GAM

earthUI 包附带两个后处理模块——一个基于 `glmnet` 的弹性网络 (elastic net) 例程和一个基于 `mgcv` 的广义可加模型 (GAM) 例程。本期各有一篇文章对它们分别做了深入论述；本节说明它们在 earthUI 工作流中的用途，以及三个例程如何协作。

13.1 为什么需要辅助拟合？

earth 模型给出价值结论。后处理模块的存在是为了印证它：由方法学上不同的引擎、针对同一估价对象 (subject)、基于相同项目数据产生的两个额外估计。当对 earth 基函数进行弹性网络重估计与独立的 GAM 光滑拟合得出实质上相同的价值时，估价师的结论便不再仅仅依赖于单一模型的假设——这在复核和诉讼场景中是一项实实在在的优势。当它们不一致时，这种分歧本身就是诊断信息，通常指向数据稀薄、某个过度灵活的项，或某个值得重新审视的离群可比销售案例。两个模块都不凌驾于 earth 模型之上；它们是佐证证据，并有意不做加权呈现。

13.2 启动与共享状态

每个例程都是一个独立的本地 Shiny 应用：

- `earthUI::launch_glmnet()`——弹性网络应用，端口 7879。
- `earthUI::launch_mgcv()`——GAM 应用，端口 7880。

三个应用读取同一项目树和同一设置数据库，因此你在 earthUI 中打开的项目在模块中立即可用——连同 earthUI 为其保存的沿用字段：Response (响应/目标变量)、Effective Date (估价基准日)、RCA 估价对象的 CQA 类型和取值，以及居住面积 (living-area) 指定。这些无需重新输入；模块会读取它们，使同样的取值贯穿全部三次拟合。

13.3 弹性网络例程

`glmnet` 例程将 earth 模型的基函数——其铰链函数、交互项和线性项——作为设计矩阵的列导入，并在套索回归 (lasso) /弹性网络正则化下对其重新估计（其机制及权重列的注意事项参见上文“`glmnetUI Import`”）。其结果保留了 earth 可解释、具有几何意义的项，同时让正则化器收缩或删除那些不能证明自身价值的项。完整论述参见本期的 `glmnetUI` 文章。

13.4 GAM 例程

`mgcv` 例程走的是另一个方向：它不复用 earth 的基函数，而是重新拟合带惩罚的光滑项，以 earth 模型的节点位置作为起点（参见上文“`mgcvUI Auto-Export`”）。由于 GAM 在不同的光滑范式下重新推导函数形式，它与 earth 模型的一致性真正独立的印证，而非回声。支持 $\text{degree} \leq 2$ 的 earth 模型。完整论述参见本期的 `mgcvUI` 文章。

13.5 一个实用的工作流

1. 先在 earthUI 中拟合并打磨 earth 模型，直到你对其满意为止。每次成功的拟合 ($\text{degree} \leq 2$) 都会自动导出供模块使用的 `.rds` 文件。
2. 针对同一项目打开 `glmnet` 和 `mgcv` 例程，导入 earth 结果并进行拟合。一次只在一个例程中工作——拟合耗时较长，且各例程通过共享项目进行协调，因此合理的节奏是顺序进行，而非并行。
3. 自由迭代。在你于各应用之间切换时，每个应用都会保留其结果，没有任何环节强制你在过程中生成报告。
4. 只有当三个结果都令你满意时，再回到各个例程生成其 Quarto 报告。每种方法都会将其拟合和价值结论登记到项目的三部曲 (trilogy) 记录中，使三份报告可追溯到同一次比较运行。

提示：这些模块在任何层面上都是可选的：*earthUI* 自身的工作流——拟合、*RCA*、销售网格、报告——没有它们也是完整的。当某个结论需要印证时再使用它们：棘手的估价对象、数据稀薄的市场细分，或很可能面临复核的委托任务。

14 演示数据集：Appraisal_1.csv

earthUI 包含一个演示 MLS 数据集，用于探索估价 workflow。可通过编程方式加载：

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
df <- import_data(demo_file)
```

或者，在 Shiny 应用中，将其复制到项目的 in/ 文件夹（或在创建项目时将其作为初始数据文件附加），然后在第 2 节（Import Data，导入数据）中选择它。

14.1 描述

该文件包含来自一份模拟 MLS 导出的 1,501 条住宅销售记录，外加第 1 行的 1 个估价对象 (subject)（共 1,502 行）。数据代表一个多区域市场中的独栋住宅销售，涵盖多种房屋面积、房龄和位置。

这不是真实数据，但基于北加州一个真实感很强的社区。所有识别信息均已被更改或删除。

14.2 数据列

列名	类型	特殊类型	描述
weight	numeric	—	观测权重 (0 = 不参与拟合)
id	numeric	display_only	内部记录 ID
property_id	numeric	display_only	MLS 房产标识符
listing_id	character	display_only	MLS 挂牌编号
parcel_number	character	display_only	县税务评估员宗地号 (APN)
street_address	character	display_only	房产地址
city_name	character	display_only	城市
postal_code	character	display_only	邮政编码 (ZIP)
county_name	character	display_only	县
contract_date	Date	contract_date	销售合同日期 (用于计算 sale_age)
sale_age	numeric	—	从合同日期到估价基准日的天数 (预先计算)
coe_date	Date	display_only	过户完成 (close of escrow) 日期
listing_status	character	display_only	挂牌状态 (例如 "Sold")
sale_price	numeric	(target)	销售价格——响应变量
rent	numeric	—	月租金 (用于多目标模型)
list_price	numeric	display_only	挂牌价格
original_list_price	numeric	display_only	原始挂牌价格
living_sqft	numeric	living_area	总居住面积 (平方英尺)
beds_total	integer	—	卧室数量
baths_total	numeric	—	浴室总数 (例如 2.5 = 2 个全卫 + 1 个半卫)
lot_size	numeric	lot_size	地块面积 (平方英尺)
area_id	integer	area	MLS 区域标识符
area_text	character	display_only	区域名称
age	numeric	actual_age	房龄 (年)
year_built	integer	display_only	建成年份
latitude	numeric	latitude	纬度 (建模时四舍五入到小数点后 3 位)
longitude	numeric	longitude	经度 (建模时四舍五入到小数点后 3 位)
latitude6	numeric	display_only	全精度纬度
longitude6	numeric	display_only	全精度经度
garage_spaces	integer	—	车库车位数
fp_count	integer	—	壁炉数量
no_of_stories	numeric	—	楼层数
style	character	—	建筑风格
view	character	—	景观类型 (例如 "Neighborhood"、"Hills")
days_on_market	integer	dom	在市天数
listing_date	Date	listing_date	挂牌日期
sale_concessions	numeric	concessions	卖方让步 (concessions) 金额

14.3 建议的快速上手步骤

1. 启动 earthUI: `earthUI::launch()`
2. 在 New Project (新建项目) 向导中创建一个 Purpose (用途) 为 **For Appraisal** (用于估价) 的项目, 并附加 `Appraisal_1.csv` 作为初始数据文件
3. 在第 2 节 (Import Data) 中选择 `Appraisal_1.csv`
4. 选择 `sale_price` 作为目标变量
5. 按上表所示指定特殊类型
6. 纳入预测变量: `sale_age`、`living_sqft`、`baths_total`、`lot_size`、`area_id` (作为因子)、`age`、`latitude`、`longitude`、`garage_spaces`
7. 将 `degree` 设为 1, 点击 **Fit Earth Model** (拟合 Earth 模型)
8. 下载中间输出 (第 6 步), 审阅 CQA 排名
9. 计算 RCA 调整 (第 7 步), CQA 得分约为 ~5.00
10. 生成销售比较网格 (第 8 步)

提示: 在导入你自己的 *MLS* 数据之前, 演示数据集是探索 *earthUI* 全部功能的好方法。上面建议的预测变量通常能得到 R^2 高于 0.85 的模型。

15 实验性功能：Prolog 文本提取与推导规则

0.10.0 版本引入了一项可选且确属实验性的功能，用于从 MLS 导出数据的描述性文本列中提取可供模型使用的特征——首要的是“Public Remarks”（公开备注）列，此外还包括经纪人备注以及以逗号分隔的特征列表，例如房间特征或既有建筑物。提取引擎是 Prolog，具体而言是确定子句文法（**Definite Clause Grammars, DCG**）：在备注（remarks）文本中识别诸如“move in ready”“3 car garage”或“owned solar”等短语的规则，均以 Prolog 语法规则的形式编写。

该功能仍在开发之中。目前它可以部分用于提取——随附的文法已能识别一组实用的特征——但要将其适配到您所在 MLS 的用语习惯，您需要能够阅读并编写 Prolog，尤其是 DCG。如果这不适合您，可以完全跳过本章：该功能默认关闭，关闭时 earthUI 的其余部分工作方式完全不变。

Warning

这是一项处于积极开发中的实验性功能。其提取文法既不完整，也未经过面向实际估价工作的验证——在将提取出的列用于模型之前，必须将其与源文本核对。要有效使用该功能，需要具备 Prolog 和 DCG 的实际工作知识。它所做的一切都是附加且可逆的；如果出现任何异常，只需关闭 Settings 中的开关并照常继续即可。

15.1 它的作用

MLS 房源在自由文本和分隔列表字段中承载着大量与价值相关的信息。这一可选步骤通过两种特殊的列角色（见第 6 章），将这些文本转化为回归可以使用的普通列：

- **remarks** 列由 Prolog DCG 文法解析，每个提取出的属性生成一个特征列。列前缀由源列名称各单词的首字母构成：public_remarks 生成 pr_* 列，agent_remark 生成 ar_* 列，依此类推。
- **list** 列（例如 existing_structures = "Shed, Workshop"）被拆分为独热编码的 <col>_<value> TRUE/FALSE 列。这一部分是纯 R 实现，完全不需要 Prolog。

随附的文法（inst/prolog/mls_remarks.pl）目前可提取的内容包括：带有 1-5 级 condition_score 的分级状况等级；景观标志与景观类型；厨房与卫浴翻新标志；硬木及其他地板材质；升级项目列表及数量；车库车位数与车库面积；泳池、水疗池及其他室外设施；壁炉数量；附属建筑（棚屋、谷仓、工作间、casita、客房、犬舍等）及从“40x50”这类尺寸中解析出的面积；房车停放与储物空间；太阳能自有与租赁之分、电池储能；地块面积（含英亩到平方英尺的换算）；与主体相关联的交易背景价格金额（降价、HOA 会费、太阳能租赁或买断、Mello-Roos）；区位线索；以及取值范围为 [-1, 1] 的亮点 / 顾虑情感评分。

15.2 环境要求：vProlog (Trealla)，而非 SWI-Prolog

备注解析运行于 vProlog R 包之上——一个自包含的 Trealla Prolog 引擎。它不在 CRAN 上；请按照本文开头“Installing earthUI and vProlog”（安装 earthUI 与 vProlog）一节所示，从作者的 r-universe 仓库安装（提供 Windows 和 macOS 的预编译二进制包，源代码位于 github.com/wcraytor/vProlog）。无需也不会使用任何系统级 Prolog 安装（无论是 SWI-Prolog 还是其他实现）。如果缺少 vProlog，备注解析会被跳过并给出提示，而 list 列拆分仍照常运行。辅助函数 vProlog_available() 可报告引擎与随附文法是否均已就绪。

15.3 启用与运行

1. 打开 **Settings**（设置，齿轮图标），勾选 **Enable Prolog / list column processing**（启用 Prolog / 列表列处理）。该开关默认关闭，附有“高级功能——面向具备 Prolog 专业知识用户”的说明，按项目记忆设置，并显示在 Project Information（项目信息）对话框中。

2. 在预测变量表中，将相关列的 **Special**（特殊）角色设为 `remarks` 或 `list`（每种角色都可以指定给多个列）。
3. 使用位于侧边栏第 4 节与第 5 节之间的 **Execute Prolog Processing**（执行 Prolog 处理）区域。它提供两个步骤：

1. **Expand lists & remarks**（展开列表与备注）——对每个 `remarks` 列运行 DCG 文法，并对每个 `list` 列进行独热拆分。展开是增量式的：只处理自上次运行以来新指定的列，一旦所有已指定的列都已展开，按钮即被禁用。解析过程由进度条跟踪。

2. **Execute derivation rules**（执行推导规则）——在展开后的列上应用您自己的 Prolog `derive/2` 规则（见下文）。规则在运行前会经过语法检查和试运行验证。

15.4 推导规则

对于每一数据行，earthUI 将该行的各列断言为 `cell(Name, Value)` 事实，并收集您的 `derive(Col, Val)` 规则的所有解——每个不同的 `Col` 都成为一个新列，且每列以找到的第一个值为准，因此可以在后面附加一条兜底子句。辅助谓词 (`count_true/2`、`count_false/2`、`count_true_prefix/2`、`cell_or/3`、`is_na/1`) 支持聚合与 NA 处理，而 `drop/1` 事实会在规则运行后丢弃相应列。例如：

```
% Merge a collinear cluster of solar flags into one boolean.
derive(solar_any, B) :-
    ( cell(pr_has_solar, true)
      ; cell(pr_solar_owned, true)
      ; cell(pr_has_battery, true) ) -> B = true ; B = false.

% Count amenities instead of OR-ing them (numeric column).
derive(amenity_count, N) :- count_true_prefix(pr_has_, N).

% Threshold -> boolean, with a fallback clause.
derive(big_outbuilding, true) :- cell(pr_outbuilding_sqft, S), S >= 1000.
derive(big_outbuilding, false).

% Discard a raw granular column once consolidated.
drop(pr_price_solar_lease).
```

规则存放在每个项目独立的 `<project>_rules.pl` 文件中，该文件由一份带完整注释的模板初始化，并通过 **Edit derivation rules...**（编辑推导规则……）在一个可浮动、可拖动、也可停靠到侧边栏的代码编辑器中编辑。**Check**（检查）按钮可在不执行规则的情况下对其进行验证（语法扫描、加载检查，以及针对数据第 1 行的试运行）。

15.5 带版本号的的活动数据文件

任一步骤完成后，增补后的数据框都会以 `<original_name>_<YYMMDD_HHMMSS>.xlsx` 的形式写入项目的输入文件夹，并成为该项目的活动数据文件——重新打开项目时加载的即是增补后的数据，因此展开结果永远不会被重复计算，而 `glmnetUI` 和 `mgcvUI` 使用的也正是同一个文件。原始的备注源列会被自动排除在模型之外；生成的列则不会被自动纳入预测变量。文本解析可能产生大量列，因此请将您（暂时）不希望进入回归的列标记为 **Latent**（潜在），或用 `×` 将其禁用（见第 6 章）。

15.6 编程接口与随附文件

整个流水线均已导出，可供批处理使用：`execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` 负责编排各个处理阶段，其底层基于 `remarks_to_columns()`、`split_list_column()`、`apply_user_rules()` 和 `validate_rules()`。位于 `inst/prolog/`

下随附的 Prolog 源文件包括: `mls_remarks.pl` (DCG 文法与特征聚合)、`mls_analyze.pl` (用于批量解析的文件 IO 驱动)、`mls_rules.pl` (可供推导规则使用的辅助谓词), 以及 `climate_regions_ca.pl` (见下文)。

15.7 相关实验: 气候区先验

一项配套的实验性功能针对数据稀薄的特征——即那些因仅出现在极少数可比销售案例中而无法由回归估计的特征 (例如蒸发式冷风机)。三个 R 辅助函数——`climate_region_for()`、`market_area_profile()` 和 `climate_feature_priors()`——将项目所在位置 (县+城市) 归入七个购房者认知层面的加利福尼亚气候区之一 (这是对加州能源委员会 16 个建筑气候区的实用性归并), 并按区域返回 HVAC 与制冷特征 (供暖、空调、蒸发式冷风机、全屋换气扇、空气过滤) 的序数化贡献价值先验。Prolog 模块 `climate_regions_ca.pl` 镜像了同一套分类, 可在推导规则内部使用; 两侧均由同一个规范的 R 源生成, 因此不会产生偏差。这些先验旨在为特征估值提供参考, 并作为对估计系数的合理性核查——区域归属 (尤其是横跨多个区域的县) 应由估价师复核。

15.7.1 超越加利福尼亚: 辖区包与市场区域模块

加利福尼亚是第一个辖区, 而非边界。`earthUI` 在国际范围内被使用, 气候先验知识正被重构为辖区包 (**jurisdiction packs**): 自包含、数据化的事实表 (分区定义、行政单元归属、特征先验), 以项目系统的国家代码和行政层级代码为键, 因此一个包可以描述一个美国的州、一个德国的联邦州 (Land), 乃至直接描述一个小国。与之配套的市场区域模块 (**market-area modules**) 层则捕捉某一具名市场区域的本地用语与更精细的先验。

二者的设计目标都是由 Valuation Engineer 和估价师在一线现场、用他们自己的语言编写——并且, 从现实出发, 可借助 LLM 完成: 世界上并没有多少 Prolog 程序员, 因此现场产出的成果被有意设计为表格而非程序。包格式是一个有文档的模式 (schema), 并附有完整的成品示例 (即加利福尼亚本身), 其明确目的就是让 Claude 这样的助手能够根据对某一市场的自然语言描述起草出一个有效的包; `earthUI` 会从包生成底层的 Prolog, 并且一个具有阻断性的验证器 (模式、覆盖度、引用完整性和试运行检查) 必须通过后, 包才能被启用。每个现场编写的包都必须携带出处信息——署名作者、日期以及所引用的权威依据——因此为这些知识背书的是负有责任的专业人员而非起草工具, 报告也能准确引用所使用的包及其版本。

提示: 一个实用的模式: 先展开备注, 再用 *derive/2* 规则将细粒度的 *pr_** 标志整合为少数几个派生的计数或布尔列, 用 *drop/1* 丢弃已整合的原始标志, 只保留派生列作为候选预测变量。这样可以保持预测变量列表精简, 并使共线性可控。

参考文献

- Friedman, Jerome H. (1991). "Multivariate Adaptive Regression Splines". In: *The Annals of Statistics* 19.1, pp. 1–141. DOI: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Milborrow, Stephen (Oct. 1, 2024). *earth: Multivariate Adaptive Regression Splines*. R package. URL: <https://CRAN.R-project.org/package=earth>.
- (n.d.[a]). *Notes on the earth Package*. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (n.d.[b]). *Variance Models in earth*. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.

glmnetUI: glmnet (Lasso 与弹性网络) 软件包的交互式界面

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

William Bert Craytor

第 1 卷，第 1 期 · 2026 年夏季

发表日期：2026 年 7 月 1 日

DOI: [10.67330/ja8s2k39](https://doi.org/10.67330/ja8s2k39)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

本开放获取文章依据 [知识共享署名 4.0 国际许可协议 \(CC BY 4.0\)](#) 的条款发布。

摘要。 glmnetUI 是 R 语言 glmnet 软件包的图形用户界面 (graphical user interface)，该软件包通过 Lasso、岭回归与弹性网络 (elastic net) 惩罚来拟合正则化的广义线性模型 (generalized linear model)。它提供三种用途模式——通用预测建模、房地产估价与市场区域分析——并引导用户完成数据导入、模型配置、交叉验证拟合、系数路径查看以及报告下载。本文记录了 glmnetUI 的数据格式要求、建模工作流程、输出显示界面以及完整的功能参考。

关键词： Lasso，岭回归，弹性网络，正则化回归，glmnet，R，图形用户界面，房地产估价

出版者按 (2026 年 7 月)：付印前不久，glmnetUI 已并入 earthUI (0.11.0 及以后版本)，本文所述的弹性网络工作流程现与 MARS 和 GAM 工作流程一同由 earthUI 提供。glmnetUI 仍以其最终发布形态 (0.5.0) 保留在 <https://github.com/wcraytor/glmnetUI>；后续开发将在 earthUI 中继续。本文适用于已发布的 glmnetUI。

目录

1	引言	3
1.1	历史与背景	3
1.1.1	Lasso	3
1.1.2	岭回归与弹性网络	4
1.1.3	glmnet 软件包	4
1.1.4	与 MARS (earth) 和 GAM (mgcv) 的比较	4
1.1.5	推荐的工作流程	5
1.2	什么是 glmnetUI?	5
1.3	三种用途模式	5
1.4	房地产专用功能	6
1.5	快速上手	6
2	MLS 输入数据要求	7
2.1	文件格式与结构	7
2.2	估价模式所需的列	7
2.3	特殊列命名约定	8
2.4	数据质量与完整性	8
2.5	标的行的位置	8
3	通用模式	9
3.1	概述	9
3.1.1	跳过第一行	9
3.1.2	按用途分别保存设置	9
3.2	侧边栏工作流程	9

3.3	项目	10
3.4	从 earthUI 导入 (可选)	10
3.5	主面板选项卡	10
3.6	设置的持久化	11
3.7	深色模式	11
3.8	区域与本地化设置	11
3.8.1	保存默认值	12
4	估价模式	13
4.1	标的行的处理	13
4.2	基准日与成交时距	13
4.3	特殊列指定	13
4.4	RCA 调整概览	13
5	市场区域分析模式	14
5.1	与估价模式的差异	14
5.2	何时使用市场模式	14
6	变量选择	15
6.1	目标变量	15
6.2	预测变量表	15
6.3	数据类型检测与手动覆盖	15
6.4	预期符号与约束的施加	15
6.5	特殊列类型参考	16
7	参数选择	17
7.1	Alpha (混合参数)	17
7.2	Lambda 的选择	17
7.3	Family (分布族)	17
7.4	Standardize (标准化)	17
7.5	符号约束	18
7.6	松弛 Lasso	18
7.7	随机种子	18
7.8	交互项矩阵	18
7.9	禁止进入主效应 (仅作交互项)	19
7.10	高级参数	19
7.11	设置默认值	19
8	拟合模型	20
8.1	拟合按钮	20
8.2	拟合过程中发生了什么	20
9	结果选项卡	21
9.1	Data Preview (数据预览)	21
9.2	Equation (方程)	21
9.3	Correlation (相关性)	21
9.4	Summary (摘要)	21
9.5	Coefficients (系数)	21
9.6	Variable Importance (变量重要性)	22
9.7	Contributions (贡献)	22
9.8	ANOVA (方差分析)	22
9.9	Diagnostics (诊断)	22

9.10	Glmnet Output (Glmnet 输出)	22
9.11	Report (报告)	23
10	下载数据	24
10.1	输出列	24
10.2	CQA 分数	24
11	RCA 计算与下载	25
11.1	打开 RCA 对话框	25
11.2	CQA 分数插值	25
11.3	输出列	25
12	销售比较表格	26
12.1	可比实例选择对话框	26
12.2	工作簿版式	26
12.3	工作表保护与剩余特征单元格	26
13	下载报告	27
13.1	两步式 Quarto 工作流程	27
13.2	报告格式	27
13.3	报告内容	27
14	与 earthUI 的比较	29
14.1	主要差异	29
14.2	共有的功能	29
14.3	何时该用哪一个	29
15	演示数据集: Appraisal_1.csv	30
15.1	说明	30
15.2	各列	30
15.3	建议的快速上手流程	30
16	系统要求与故障排查	32
16.1	受支持的平台	32
16.2	可选依赖	32
16.3	各平台说明	32
16.4	降级运行	32
16.5	故障排查	32

1 引言

1.1 历史与背景

1.1.1 Lasso

Lasso (Least Absolute Shrinkage and Selection Operator, 最小绝对收缩与选择算子) 由 Robert Tibshirani 于 1996 年提出。这个名称是对牛仔套索 (lasso) 的有意双关——套索是一种既能约束又能选择的绳索。Tibshirani 之所以选用这个名字, 是因为该方法真的会“套住”系数, 约束它们的绝对值, 并把其中一部分拉到恰好为零, 从而选出哪些变量留在模型之中。

Tibshirani 的洞见在于给最小二乘目标函数加上一个 L_1 惩罚 (penalty, 系数绝对值之和)。岭回归 (ridge regression, Hoerl 与 Kennard, 1970) 所用的 L_2 惩罚只会把系数向零收缩而永

远达不到零；与之不同， L_1 惩罚产生的是稀疏解，其中许多系数恰好等于零。这使得 Lasso 既是一种正则化（regularization）方法，也是一种变量选择方法。

1.1.2 岭回归与弹性网络

岭回归（ridge regression）由 Arthur Hoerl 与 Robert Kennard 于 1970 年提出。它添加的是 L_2 惩罚（系数平方之和），会按比例收缩所有系数，但把所有变量都保留在模型中。当预测变量之间存在相关性（多重共线性）时，岭回归很有效，但它不会简化模型。

弹性网络（elastic net）由 Hui Zou 与 Trevor Hastie 于 2005 年提出，它把两种惩罚结合起来。混合参数 α 控制两者的配比： $\alpha = 1$ 为纯 Lasso， $\alpha = 0$ 为纯岭回归，介于其间的取值则给出折中方案。弹性网络克服了 Lasso 的一个局限：当预测变量高度相关时，Lasso 往往只选中其中一个而忽略其余，而弹性网络则会把相关的预测变量成组保留。

1.1.3 glmnet 软件包

R 软件包 `glmnet` 由斯坦福大学的 Jerome Friedman、Trevor Hastie 与 Robert Tibshirani 开发。它使用极其高效的坐标下降算法为广义线性模型（generalized linear model, GLM）实现弹性网络正则化（Friedman, Hastie, and Tibshirani 2010），其后又扩展到 Cox 比例风险模型（Simon et al. 2011）以及全部广义线性模型族（Tay, Narasimhan, and Hastie 2023）。该软件包可处理 gaussian（线性）、binomial（逻辑）、poisson 及其他 GLM 族。

`glmnet` 的主要特性包括：

- 正则化路径：一次运算即可在整个 lambda 网格上拟合模型，比在每个 lambda 处分别拟合快得多。
- 交叉验证（cross-validation）：`cv.glmnet()` 通过 k 折交叉验证自动选出最优的正则化强度。
- 系数界限：支持对单个系数设定上限和下限（符号约束）。
- 松弛 lasso（relaxed lasso）：在选中的变量上以减弱的惩罚或完全不加惩罚重新拟合模型，以减小偏差。

1.1.4 与 MARS (earth) 和 GAM (mgcv) 的比较

`earth` 应用程序及其两个后处理模块（`glmnet` 与 `mgcv`）使用不同的建模引擎。下表总结了它们的主要差异：

特性	<code>glmnet</code> （弹性网络）	<code>earth</code> （MARS）	<code>mgcv</code> （GAM）
模型类型	线性（带正则化）	分段线性（自适应样条）	光滑非线性（惩罚样条）
非线性	仅能通过交互项或基函数展开实现	通过带数据驱动节点的铰链函数自动实现	通过光滑函数自动实现（薄板、三次样条等）
变量选择	内置，经由 L_1 惩罚（Lasso）	内置，经由前向 / 后向逐步法配合 GCV 剪枝	可选，经由双惩罚收缩（ <code>select=TRUE</code> ）
交互项	两两交叉乘积（不易解释）	基于铰链函数的乘积（可解释，最高 3 阶）	张量积光滑项（ <code>te()</code> 、 <code>ti()</code> ）
可解释性	系数即线性权重，逐个易于说明	g 函数按变量展示分段线性偏效应	光滑的偏效应曲线，非常直观
最适用于	高维数据、变量选择、小样本	自动识别非线性与交互作用	光滑的非线性关系、灵活建模
RCA 调整	逐变量的线性调整	经由 g 函数的分段线性调整	经由偏效应的光滑调整
过拟合控制	交叉验证选定的 lambda	基于 GCV 的剪枝	REML/GCV 光滑参数估计

MARS (Multivariate Adaptive Regression Splines, 多元自适应回归样条) 由 Jerome Friedman 于 1991 年提出。它通过从数据中自适应地选择铰链函数 $\max(0, x - k)$ 及其节点位置来构建分段线性模型。其 R 实现是 Stephen Milborrow 编写的 **earth** 软件包。MARS 的长处在于无需用户指定即可自动发现非线性关系与交互作用。它的 g 函数 (按变量归组的基项) 提供了高度可解释的偏效应曲线。

GAM (Generalized Additive Models, 广义可加模型) 由 Hastie 与 Tibshirani 于 1990 年正式确立。其 R 实现 **mgcv** 由 Simon Wood 编写, 提供带自动光滑度选择 (经由 REML 或 GCV) 的惩罚回归样条。GAM 产生的光滑偏效应曲线便于可视化和解释。当 **earthUI** 把节点位置导出到 **mgcvUI** 时, **earth** 得出的节点会作为 GAM 光滑项的起始点, 从而把 MARS 的自适应节点定位与 GAM 的光滑估计结合起来。

1.1.5 推荐的工作流程

对于房地产估价 (real estate appraisal) 以及其他需要可解释、可辩护模型的类似应用:

1. 从 **earthUI** 入手, 发现数据中的重要变量、非线性关系与交互作用。**earth** 的自动基函数选择与 g 函数能提供极佳的初步洞察。
2. 用 **mgcvUI** 加以细化, 如果您希望得到更光滑的偏效应。把 **earthUI** 的节点导入 **mgcvUI**, 即可从分段线性模型平滑过渡到光滑模型。
3. 使用 **glmnetUI**, 当您需要力度更大的变量选择 (Lasso)、预测变量数量接近或超过观测数量, 或者希望得到一个带正则化、便于辩护的纯线性模型时。

1.2 什么是 glmnetUI?

glmnetUI 是 R 语言 **glmnet** 软件包的图形用户界面 (graphical user interface)。它以本地 Shiny 应用程序的形式运行——无需登录、无需服务器、也没有账户。您从 R 中启动它, 导入数据集 (CSV 或 Excel), 配置模型, 然后交互式地进行拟合。

从当前版本起, **glmnetUI** 不再是一个单独安装的软件包: 它作为弹性网络后处理模块随 **earthUI** 软件包一起发布, 用 `earthUI::launch_glmnet()` 启动。应用程序本身——以及本文所述的一切——并无改变; 不同的只是安装方式和启动命令。三个例程 (**earth**、**glmnet**、**mgcv**) 共享同一个项目树、同一个设置数据库以及同一套支撑基础设施。

该应用程序提供了完整的工作流程: 数据导入、变量配置、模型拟合、诊断图、变量重要性、模型方程, 以及可下载的 Word、PDF 或 HTML 格式报告。

弹性网络回归结合了两种正则化技术:

- 岭回归 ($\alpha = 0$): 按比例把所有系数向零收缩。当您认为所有变量都相关时较为合适。
- Lasso 回归 ($\alpha = 1$): 可以把系数收缩到恰好为零, 从而实现自动变量选择。产生的模型更简洁、更易解释。
- 弹性网络 ($0 < \alpha < 1$): 两种做法的混合。当预测变量之间相关时很有用。

1.3 三种用途模式

每个 **glmnetUI** 项目都有一个 **Purpose** (用途), 在创建项目时选定 (第 1 节, *Project* (项目))。当前项目的用途决定哪些工具和界面元素会出现; 若要在另一种模式下工作, 请创建或打开一个具有该用途的项目。三种模式为:

- **General** (通用) ——适用于任何类型的总体或数据集的弹性网络回归。这是默认模式。它提供完整的正则化回归工作流程, 不含任何领域专用的附加功能。
- **For Appraisal** (用于估价) ——为房地产估价定制的弹性网络回归。增加了针对单一物业估值的功能, 包括标的物业处理、特殊列指定以及残差约束法 (Residual Constraint Approach, RCA)。
- **Market Area Analysis** (市场区域分析) ——为市场区域研究定制的弹性网络回归。增加了分析既定市场中成组物业的功能, 并可选择排除标的行。

在这三种模式下，核心建模引擎完全相同——您始终是在拟合一个弹性网络 (glmnet) 模型。用途设置只控制哪些附加工具和界面元素可用。

1.4 房地产专用功能

当选择 **For Appraisal** 或 **Market Area Analysis** 时，glmnetUI 会启用若干为房地产分析设计的功能：

- **特殊列指定**——每个预测变量都可以标记一个特殊角色，例如 `contract_date`、`dom`、`concessions`、`latitude`、`longitude`、`living_area`、`lot_size`、`actual_age`、`effective_age`、`area`、`site_dimensions` 或 `display_only`。这些指定控制该列在拟合与输出过程中的处理方式。
- **经纬度取整**——被指定为纬度或经度的列会自动四舍五入到小数点后 3 位，以防止过拟合。
- **Sale Age 列（成交时距）**——当某列被指定为 `contract_date` 且提供了 Effective Date（估价基准日）时，glmnetUI 会计算一个 `sale_age` 列（成交日期与基准日之间的天数），并用它替代原列作为预测变量。
- **RCA 计算（仅限估价模式）**——在估价模式下，拟合模型之后 glmnetUI 可以计算残差约束法（RCA）输出，得出逐个可比实例的调整值、净 / 毛调整汇总，以及标的物业的调整后成交价格。

提示：General 用途模式适用于任何数据集——金融、科学、工程或房地产。估价模式与市场模式只是为房地产专业人士增加了一些便利功能。

1.5 快速上手

使用 glmnetUI 的步骤：

1. 在 R 中安装 **earthUI**：`install.packages("earthUI")`——glmnet 例程已包含在内。（关于开发版和可选组件，参见本期 earthUI 一文中的“Installing earthUI and vProlog”。）
2. 启动应用程序：在 R 中运行 `earthUI::launch_glmnet()`。应用程序会在您的网页浏览器中以端口 7879 打开。您也可以在浏览器中直接访问 <http://localhost:7879>。应用程序会记住您上次使用的用途模式并自动恢复。
3. 在侧边栏第 1 节创建或打开项目。项目设定用途和工作地点（国家、州 / 省、县、市）及其输入 / 输出文件夹。项目存放在共享的 `regProj` 根目录下，并与姊妹应用程序 earthUI 和 mgcvUI 共用。打开一个具有不同用途的项目，就是切换模式的方式。
4. 在第 2 节导入数据，从当前项目的输入文件夹中选择一个 CSV 或 Excel 文件。把文件放入该文件夹后，点击 **Refresh**（刷新）即可列出它们。
5. 从 **earthUI** 导入（可选）——第 3 节允许您导入一个 earthUI 结果 `.rds` 文件，以便把 earth 的铰链基函数与 glmnet 的正则化结合使用。若没有 earthUI 结果，可跳过此步。
6. 配置变量——选择目标变量和预测变量，设置数据类型、预期符号，并指定任何特殊列角色。
7. 设定 **glmnet** 参数——`alpha`、`lambda`、`family`、符号约束、交互项及其他选项。
8. 拟合模型——点击“Fit Glmnet Model”（拟合 Glmnet 模型），并在主面板中查看结果。
9. 导出——将预测结果下载为 Excel，生成并转换 Quarto 报告，或（在估价模式下）计算 RCA 调整值与销售比较表格（Sales Comparison Grid）。

设置会自动保存在浏览器的本地存储中，并在您重新载入同一个输入文件时恢复。

2 MLS 输入数据要求

对于房地产估价与市场分析工作流程，您的输入数据通常来自多重上市服务（Multiple Listing Service, MLS）导出文件。本章说明预期的文件结构以及 glmnetUI 可以使用的列。

2.1 文件格式与结构

glmnetUI 接受 CSV 与 Excel (.xlsx、.xls) 文件。导入时，列名会自动转换为 snake_case——例如，“Living SqFt”变为 living_sqft, “Contract Date”变为 contract_date, “Sale Price”变为 sale_price。这种规范化确保了整个工作流程中列引用的一致性。导入时使用的 CSV 分隔符和小数点标记由区域设置决定（参见第 3 章“Locale & Regional Settings”）。

您的数据文件应当是一张扁平表格，每行一处物业，每列一个属性。文件的第一行必须包含列标题。

2.2 估价模式所需的列

虽然 glmnetUI 可以处理任意一组列，但完整的估价工作流程若具备以下各列会更为得力：

列	特殊类型	用途
Sale Price（成交价格）	（目标变量）	模型的响应变量
Contract Date（合同日期）	contract_date	用于计算 sale_age（距基准日的天数）
Listing Date（挂牌日期）	listing_date	在没有 DOM 列时，与合同日期一起计算 DOM
Days on Market（在市天数）	dom	在市天数；显示于导出结果中
Concessions（成交让利）	concessions	成交让利；净成交价 = 成交价格 - 让利
Living Area (SF)（居住面积）	living_area	启用每平方英尺残差（residual_sf、cqa_sf）
Lot Size（地块面积）	lot_size	宗地面积列
Site Dimensions（宗地尺寸）	site_dimensions	与地块面积归为一组
Latitude（纬度）	latitude	四舍五入到小数点后 3 位
Longitude（经度）	longitude	四舍五入到小数点后 3 位
Area ID（区域编号）	area	市场区域 / 街区标识
Actual Age（实际年龄）	actual_age	物业年龄
Effective Age（有效年龄）	effective_age	物业的有效年龄
Address（地址）	display_only	显示于导出结果中；不进入模型

电子表格的列名可以使用外语——“特殊”名称之所以用英文，是为了让 R 程序能对它们作特殊处理。除此之外，您所给定的列名会出现在回归模型、图形中，以及（若在做估价）输出报告中。

并非所有列都是必需的。glmnetUI 会自行适应——若缺少某一列，相应的功能就直接省略。不过，对于房地产定价模型，为达到可接受的拟合效果，强烈建议具备以下几列：

1. **Sale Age** (成交时距) ——合同成交日期与估价或分析基准日之间的天数。若使用跨年度的成交历史,尤其是超过 5 年的期间, `sale_age` 在估计成交价格时往往起核心作用。
2. **Living Area** (居住面积) ——也常被称为“Living Sqft”、“GLA”(gross living area, 总居住面积)等。这是成交价格的又一主要决定因素。
3. **Total Bath Count** (卫生间总数) ——全卫、四分之一卫、半卫和四分之三卫的总数。例如,两个全卫加一个半卫记为 2.5。
4. **Garage Bays** (车位数) 或 **Garage Area** (车库面积) ——车库停车位数量或车库面积。
5. **Lot Size** (地块面积) ——物业的土地面积,通常以平方英尺或英亩计。
6. **Longitude** (经度)、**Latitude** (纬度),以及(若有) **Area ID** (区域编号)。这些位置变量有助于模型解释地理上的价格差异。

2.3 特殊列命名约定

glmnetUI 通过特殊类型指定而非列名来识别列。在 MLS 导出文件中,您可以随意命名各列——关键在于您在变量配置表(第 6 章)中指定了正确的特殊类型。

例如,您的 MLS 可能把居住面积导出为“GLA”、“Living SqFt”、“liv_area”或“gross_living_area”。导入之后(列名转为 snake_case),您只需在 Special(特殊)下拉框中把它指定为 living_area。此后无论其原始名称如何,glmnetUI 都会用它来计算每平方英尺的残差。

2.4 数据质量与完整性

- **缺失值 (NA)**: 任何预测变量列或目标列中含 NA 的行,会在拟合前自动剔除。
- **日期列**: 符合常见日期格式的字符列在导入时会自动解析。当检测不到 4 位年份时,优先按 2 位年份格式解析。
- **数值列**: 成交价格、居住面积、地块面积、让利等字段必须为数值。若您的 MLS 导出的价格带有货币符号或千位分隔符(例如“\$350,000”或“350.000,00”),可能需要在导入前先行清洗。
- **因子列**: 文本列、TRUE/FALSE 列和 R 因子列会自动按分类变量处理——回归只能把它们当作因子。而数值列(例如数字形式的区域码或风格码)只有在您勾选其 **Factor** (因子)复选框时才会变成因子;glmnetUI 绝不会根据数值的取值范围推断其为分类变量,因此像 bath_count 这样的离散数值预测变量仍保持连续。Special 角色与是否作为因子无关。

提示: 导入后请查看预测变量表中的 NA 一列。缺失值很多的列可能导致大量行被剔除。可以考虑排除高 NA 列,或在导入前清洗数据。

2.5 标的行的位置

在 **Appraisal** (估价) 模式下,第 1 行必须是标的物业。其余各行均为可比成交实例。标的行不参与模型拟合。拟合完成后,模型仍会为标的行生成预测值。

在 **Market Area Analysis** (市场区域分析) 模式下,把标的放在第 1 行是可选的。

在 **General** (通用) 模式下没有特殊的行处理——所有行一视同仁。

3 通用模式

3.1 概述

启动 glmnetUI 时的默认模式即为通用 (General Purpose) 模式。它为任何数据集——而不仅仅是房地产数据——提供完整的弹性网络回归工作流程。您可以把 glmnetUI 用于科学数据、金融分析、工程研究, 或任何回归问题。

在通用模式下, 界面会略去房地产专用功能 (特殊列、成交时距、坐标取整、RCA)。侧边栏经过精简, 专注于变量选择、参数配置、模型拟合与导出。

3.1.1 跳过第一行

打开通用或市场项目后, 侧边栏顶部会出现一个 **Skip first row** (跳过第一行) 复选框。勾选后, 第 1 行将不参与模型拟合。当第 1 行包含一个不应影响模型的标的或参照观测值时, 这一选项很有用。在估价模式下, 第 1 行 (标的物业) 始终自动排除。

3.1.2 按用途分别保存设置

设置 (预测变量选择、模型参数、交互项) 会按输入文件与用途模式的每种组合分别保存。在通用、估价和市场模式之间切换时, 各模式的设置互不影响、各自保留。

3.2 侧边栏工作流程

侧边栏按编号、可折叠的小节组织, 引导您从项目选择一路走到报告导出。多数小节只在打开活动项目后才出现, 其中几个还与项目的用途有关。输出文件夹由活动项目推导得出——没有单独的输出文件夹字段。

1. Project (项目) ——创建或打开项目。项目设定用途和工作地点 (国家、州 / 省、县、市) 及其输入 (`<os>_in/`) 和输出文件夹。项目存放在共享的 `regProj` 根目录下 (默认 `~/regProj`, 可用 `REGPROJ_ROOT` 环境变量覆盖), 并与姊妹应用程序 `earthUI` 和 `mgcvUI` 共用, 因此三者的模型可以并存。点击 **Close Project** (关闭项目) 即可切换。

2. Import Data (导入数据) ——从活动项目的输入文件夹中选择一个 CSV 或 Excel 文件。通过访达 / 资源管理器把文件放入该文件夹, 然后点击 **Refresh** (刷新) 列出它们。对于含多个工作表的 Excel 文件, 会出现工作表选择器。列名会自动转换为 `snake_case`。

3. Import from earthUI (从 earthUI 导入, 可选) ——导入一个 `earthUI` 结果 `.rds` 文件, 以便把 `earth` 的铰链基函数与 `glmnet` 的弹性网络正则化结合使用。浏览按钮的样式与第 2 节的文件选择器一致。此步骤为可选——若没有可导入的 `earthUI` 结果, 可以跳过。

4. Variable Configuration (变量配置) ——目标变量选择器, 带 Include (纳入)、Factor (因子) 和 Force (强制保留) 复选框的预测变量表, 以及 Sign (符号) 下拉框。Special (特殊) 列——用于指定合同日期、坐标等——只在估价与市场模式下出现。详见第 6 章。

5. glmnet Call Parameters (glmnet 调用参数) ——全部模型配置: `alpha`、`lambda`、`family`、`standardize`、符号约束、松弛 `lasso`、随机种子、交互项矩阵, 以及高级参数 (`lambda.min.ratio`、`nlambda`、CV 损失度量、收敛阈值、最大迭代次数、截距)。完整的参数参考见第 7 章。

6. Fit Glmnet Model (拟合 Glmnet 模型) ——运行模型的按钮。结果会显示在右侧的各个选项卡中。

7. Download Output (下载输出) ——把预测值、残差、CQA 分数和逐变量贡献导出为 Excel 文件, 存入项目的输出文件夹。模型拟合后, 在所有用途模式下均可用。

8. Calculate RCA Adjustments & Download (计算 RCA 调整值并下载) ——为每个可比实例计算逐变量的 RCA 调整值, 并通过 CQA 插值得出标的残差。仅在估价与市场模式下、拟合之后出现。详见第 11 章。

9. Generate Sales Grid & Download (生成销售比较表格并下载) ——生成中间的销售比较表格, 按毛调整百分比对可比实例排序。仅在估价模式下、计算完 RCA 调整值之后出现。详见第 12 章。

10. Generate Quarto Report (生成 Quarto 报告)——把一个自包含的 Quarto (.qmd) 报告包——含全部图形与表格——写入项目的输出文件夹。模型拟合后即可使用。

11. Convert Quarto Report (转换 Quarto 报告)——把生成的 .qmd 报告包渲染为 HTML、Word 或 PDF。PDF 需要安装 LaTeX；若未检测到，则隐藏 PDF 选项。详见第 13 章。

在市场模式与通用模式下，仅限估价的步骤会被跳过，两个报告小节的编号也相应调整。

3.3 项目

glmnetUI 把工作组织为共享 regProj 根文件夹下的项目 (earthUI 与 mgcvUI 使用的是同一棵目录树)。每个项目拥有一个输入文件夹 (<os>_in) 和按引擎划分的输出文件夹 (例如 <os>_out_glmnet)。

要创建项目，请点击第 1 节中的 + **New...** (新建)，并提供：

- **Purpose (用途)**——General、Appraisal 或 Market Area。
- **Location (地点)**——项目的政治 / 行政层级。需要填写哪些层级取决于您选择的国家，因此各国的命名惯例都能自动支持——美国项目使用州 / 县 / 市，法国项目使用 region/département/commune，日本项目使用都道府县 / 市或区 / 町丁，依此类推。
- **Project Name (项目名称)**——一个简短标签，限 8 个字符 (字母、数字、_ 和 -)。

应用程序不会原样使用您输入的名称：它会在前面加上地点代码和创建日期时间，构造出一个唯一的带时间戳的文件夹名，形式为 <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name>——例如，一个名为 burl 的美国项目会变成 us_ca_081_burlin_20260604-131500_burl。8 个字符的限制只适用于您键入的那一部分。只有新建的项目才会带上时间戳前缀；磁盘上已有的项目保留原名，并且照常完全可用。

3.4 从 earthUI 导入 (可选)

侧边栏的第 3 节——**Import from earthUI** (从 earthUI 导入)——允许您导入一个 earthUI 结果 .rds 文件。这使得 glmnetUI 能够把 earth 的铰链基函数 (分段线性项) 与 glmnet 的弹性网络正则化结合起来，把 earth 的自动非线性识别与 glmnet 的变量选择和系数约束合二为一。

导入 earthUI 结果后，驱动 glmnet 拟合的是 earth 模型的预测变量，而不是 Include 复选框，因为 glmnet 是在 earth 的基函数展开上拟合的。变量配置表会明确体现这一点：**Include** 与 **Type** 被锁定，earth 模型未使用的预测变量所在行被禁用，并有一条提示说明预测变量集来自 earth 模型。**Force** 与 **Sign** 对 earth 的预测变量仍可编辑，因为它们在拟合时仍会作用于 earth 基函数。

浏览按钮的样式与第 2 节的文件选择器一致。此步骤完全可选——若没有 earthUI 结果文件，直接跳过第 3 节进入第 4 节即可。

3.5 主面板选项卡

拟合完成后，主面板提供以下选项卡：

选项卡	内容
Data Preview (数据预览)	导入数据的预览。在估价模式下，分为标的物业表与可比成交实例表。
Equation (方程)	以 LaTeX/MathJax 显示的拟合模型方程，展示每个非零系数与交互项。
Correlation (相关性)	数值型预测变量相关系数的热力图（拟合前即可查看）。
Summary (摘要)	关键指标 (R^2 、调整 R^2 、 GR^2 、 $CV R^2$ 、RMSE、MAE) 与估价指标 (COD、PRD、中位数比率)。
Coefficients (系数)	非零系数表，附符号违例警示。
Variable Importance (变量重要性)	预测变量重要性 ($ \beta \times sd(x)$) 的条形图与排序表。可用时支持 plotly 交互式悬停。
Contributions (贡献)	带斜率标注的逐变量偏效应图。数值型预测变量为散点图 + 拟合线，因子为箱线图，交互项为三维曲面 / 散点 / 热力图。
ANOVA (方差分析)	方差分解表：逐变量的平方和、占模型平方和的百分比、系数。
Diagnostics (诊断)	四幅图：系数路径、CV 误差、实际值对预测值、残差对拟合值。
Glmnet Output (Glmnet 输出)	原始模型输出：所用随机种子、模型打印结果、选定的 λ 、 $\lambda_{min}/1se$ 、 γ (若使用松弛 lasso)，以及完整的系数向量。
Report (报告)	报告配置与导出字段。
RCA Adjustments (RCA 调整)	RCA 分析结果（估价模式，计算完成后）。

3.6 设置的持久化

glmnetUI 会自动把您的配置保存到浏览器的本地存储中，以输入文件名作为索引键。当您重新载入同一个文件时，所有设置都会恢复：目标变量选择、预测变量复选框、数据类型、预期符号、glmnet 参数以及交互项矩阵。上次使用的用途模式也会全局保存，并在重新启动应用程序时恢复。有三个选项可供选择：

- **Use last settings for input file** (使用该输入文件的上次设置) ——恢复为此文件保存的设置
- **Use default settings** (使用默认设置) ——应用您保存的全局默认值
- **glmnet defaults** (glmnet 默认值) ——重置为出厂默认值 ($\alpha = 1$ 、 $CV \lambda_{1se}$ 、10 折、gaussian、开启标准化)

Save current as default (将当前设置存为默认) 按钮可把当前的全部设置保存为全局默认值。

3.7 深色模式

点击右上角的月亮 / 太阳图标，即可在浅色与深色主题之间切换。主题偏好保存在本地存储中，跨会话保持。所有界面元素（表格、图形、卡片、按钮）都会通过 CSS 变量适配所选主题。

3.8 区域与本地化设置

glmnetUI 通过基于国家的区域设置系统支持各国的数字与 CSV 格式惯例。标题栏中的 **Settings** (设置) 下拉菜单提供面向 31 个受支持国家的 **Country** (国家) 与 **Paper** (纸张) 选择器。每个预设会配置：

- **CSV 分隔符**——美国 / 英国 / 日本用逗号 (,)，欧洲多数国家用分号 (;)，因为那里逗号被用作小数点。
- **小数点标记**——句点 (.) 或逗号 (,)。
- **千位分隔符**——逗号 (美国 / 英国 / 日本)、句点 (德国 / 意大利 / 西班牙)、空格 (芬兰 / 法国 / 波兰 / 波罗的海国家 / 乌克兰 / 俄罗斯)，或撇号 (瑞士)。
- **纸张尺寸**——Letter (美国 / 加拿大 / 墨西哥) 或 A4 (其他所有地区)。

3.8.1 保存默认值

点击 **Save as my default** (存为我的默认值)，即可全局保存您的区域设置偏好。这些默认值将适用于今后所有会话，无论您载入哪个数据文件。

提示：图形坐标轴和斜率标注上的数字格式会自动适配您的区域设置。德语区域用句点作千位分隔符 (200.000)，芬兰用空格 (200 000)，瑞士用撇号 (200'000)。不显示任何货币符号——*glmnetUI* 与币种无关。

4 估价模式

当您把 Purpose 选为 **For Appraisal**（用于估价）时，glmnetUI 会针对单一物业估值进行自我配置。第 3 章所述的全部功能依然可用；本章只讲估价模式特有的附加内容。

4.1 标的行的处理

在估价模式下，数据集的第 1 行即标的物业，其余各行均为可比成交实例。您的输入文件必须据此组织（参见第 2 章）。标的的成交价格可以留空或填任意值——glmnetUI 在拟合时会自动把它当作 NA 处理。

导入之后，Data Preview 选项卡会分为两部分：**Subject Property**（标的物业，第 1 行）与 **Comparable Sales**（可比成交实例，第 2 行及以后）。第 1 行始终不参与模型拟合。拟合完成后，模型仍会为标的行生成预测值。

4.2 基准日与成交时距

在估价与市场模式下，变量配置小节中会出现一个 **Effective Date**（估价基准日）字段（默认为当天日期）。如果您在 Special 列下拉框中把某列指定为 `contract_date`，glmnetUI 就会计算一个 `sale_age` 列——即每笔成交的合同日期与基准日之间的整数天数。该列会作为预测变量加入模型。

当基准日发生变化时，`sale_age` 会自动重新计算。

4.3 特殊列指定

在估价与市场模式下，变量配置表中每个预测变量都会出现一个 **Special**（特殊）下拉框。特殊类型及其作用的完整参考见第 6 章。

4.4 RCA 调整概览

Calculate RCA Adjustments & Download（计算 RCA 调整值并下载）按钮（侧边栏第 8 节，在估价与市场模式下拟合之后可见）会为每个可比实例计算相对于标的的市场推导调整值。完整的 RCA 工作流程见第 11 章。

提示：您赋予标的的 *CQA* 分数决定了残差分布中有多大比例归属于标的。分数为 5.00 时，标的位于各可比实例的中位数处。

5 市场区域分析模式

当您把 Purpose 选为 **Market Area Analysis**（市场区域分析）时，glmnetUI 提供与估价模式相同的房地产专用功能（特殊列、成交时距、坐标取整），但其取向是分析一组物业，而非为单一标的估值。

5.1 与估价模式的差异

- 标的行处理更为灵活——在市场模式下，第 1 行不会被自动当作标的物业。
- 没有销售比较表格——“Calculate RCA Adjustments & Download”（第 8 节）步骤在市场模式下可用，但销售比较表格步骤（第 9 节）仅限估价模式。

5.2 何时使用市场模式

在以下情形下适合采用市场区域分析模式：

- 为某个街区或市场区域建立回归模型，以了解价值驱动因素
- 分析面积、年龄、地块面积和位置等变量如何影响一组物业的成交价格
- 为市场状况分析或街区界定准备支撑材料
- 处理某个数据集时，您需要特殊列功能（成交时距、坐标取整），但不需要 RCA 调整工作流程

提示：当您需要特殊列功能但不需要 RCA 调整工作流程时，市场模式同样适用于一般性的房地产回归分析。

6 变量选择

侧边栏的第 4 节——**Variable Configuration**（变量配置）——是您选择哪些列参与模型以及如何处理它们的地方。

6.1 目标变量

第 4 节顶部的 **Target (response) variable**（目标 / 响应变量）下拉框会列出数据集中的每一列。请选择一列作为响应变量（例如成交价格）。目标列会自动从预测变量列表中排除。

6.2 预测变量表

在目标变量选择器下方，有一张表列出其余每一列，包含以下字段：

列	说明
Variable （变量）	列名
Type （类型）	数据类型下拉框：numeric、integer、character、Date、POSIXct
Include （纳入）	复选框——把该列作为预测变量纳入模型
Factor （因子）	复选框——把该变量当作分类因子处理（在模型矩阵中生成哑变量列）
Force （强制保留）	复选框——强制进入模型（惩罚因子 = 0，绝不会被 lasso 剔除）
Special （特殊）	下拉框（仅估价 / 市场模式）——参见下文的特殊列类型参考
Sign （符号）	预期的系数符号：正、负或不限
NAs	缺失值计数

6.3 数据类型检测与手动覆盖

glmnetUI 在导入时会自动检测数据类型，能够识别数值型、整型和日期型的列。看起来像日期的字符列（符合常见日期格式模式）会被归类为 Date。

您可以通过修改 **Type** 下拉框来覆盖任何检测结果。更改类型会影响该列在模型矩阵中的编码方式。

文本列、TRUE/FALSE 列和 R 因子列会自动按分类变量处理。数值列只有在您勾选其 **Factor** 复选框时才会变成因子——glmnetUI 绝不会根据数值的取值范围推断其为分类变量，因此像卫生间数量这样的离散数值预测变量除非您另行指定，否则保持连续。Factor 设置通过 localStorage 持久保存。

6.4 预期符号与约束的施加

对每个预测变量，**Sign** 下拉框用于指定您预期其系数为正、为负还是不限方向。拟合之后：

- 符号与预期不符的系数会在系数表中被标记出来
- 符号警示会显示在 Summary 选项卡中

当 glmnet 参数（第 7 章）中启用了 **Enforce Sign Constraints**（施加符号约束）复选框时，模型会被强制产生与预期符号一致的系数。这是通过 glmnet 中的 upper.limits 与 lower.limits 实现的。

6.5 特殊列类型参考

在估价与市场模式下，**Special** 下拉框提供以下选项：

Weighting (加权):

- **weight** ——观测权重列（所有模式下均可用；只允许指定一列；权重为 0 的行不参与拟合）

Date & Time Types (日期与时间类型):

- **contract_date** ——触发依据基准日自动计算 **sale_age**
- **listing_date** ——在计算在市天数时作为备用来源
- **dom** ——标识在市天数 (Days on Market) 列
- **sale_age** ——指定一个已有的成交时距列，这样任何一列都可以充当成交时距，而不必让 glmnetUI 从 **contract_date** 计算

Monetary Types (货币类型):

- **concessions** ——标识成交让利（卖方补贴、买方激励等）

Transaction Types (交易类型):

- **sale_type** ——标识成交 / 交易类型列

Size & Location Types (面积与位置类型):

- **latitude** ——数值自动四舍五入到小数点后 3 位
- **longitude** ——与纬度作相同的取整处理
- **area** ——市场区域或街区标识（例如 **area_id**、**neighborhood**、**market_area**）
- **living_area** ——启用按平方英尺计的残差计算 (**residual_sf** 与 **cqa_sf**)
- **lot_size** ——宗地面积列
- **site_dimensions** ——与地块面积归为一组（例如“75x120”）

Age Types (年龄类型):

- **actual_age** ——物业年龄列
- **effective_age** ——物业的有效年龄

Display Types (仅显示类型):

- **display_only** ——该列会包含在 Excel 导出结果中，但完全不参与模型拟合。可用于地址字段、MLS 编号、地块编号或其他参考数据。可以有多列采用此指定。

提示：被指定为 *display_only* 的列仍保留在数据集中并出现在 Excel 导出结果里，但完全不进入预测变量表。可用于 ID 列、地址或其他参考字段。

7 参数选择

侧边栏的第 5 节——**glmnet Call Parameters** (glmnet 调用参数)——提供弹性网络模型的全部配置选项。每个参数旁都有一个蓝色帮助图标 (?), 悬停可看到说明。

7.1 Alpha (混合参数)

alpha 参数控制正则化的类型:

Alpha 取值	类型	行为
0	岭回归	按比例收缩所有系数; 绝不会把任何系数置为零
1	Lasso	可以把系数置为恰好为零 (变量选择)
0 至 1	弹性网络	岭回归与 lasso 的混合

有两种 alpha 选择方法:

- **Fixed** (固定) ——通过滑块设定单一 alpha 值 (默认: 1.0)
- **Grid Search** (网格搜索) ——测试多个 alpha 值, 选出 CV 误差最小的那个。可配置取值范围与取值个数。

7.2 Lambda 的选择

lambda 控制正则化的强度:

方法	说明
交叉验证	推荐。使用 k 折 CV 测试多个 lambda 值。
手动	若您有充分理由, 可输入指定的 lambda 值。

使用交叉验证时, 有两种 lambda 取法:

- **lambda.1se** (默认) ——在最小值一个标准误范围内的最大 lambda。产生更简洁、正则化更强的模型。估价工作推荐使用。
- **lambda.min** ——使交叉验证误差最小的 lambda。产生拟合最好但可能更复杂的模型。

Number of CV Folds (CV 折数) 参数控制交叉验证时数据的划分方式 (默认: 10 折)。

7.3 Family (分布族)

为您的响应变量选择分布族:

分布族	适用情形
gaussian	连续型响应 (例如成交价格)。最常用。
binomial	二值结果 (例如已售 / 未售)。
poisson	计数数据 (例如成交笔数)。

7.4 Standardize (标准化)

勾选时 (默认), 所有预测变量在拟合前都会被缩放为均值 0、标准差 1。这确保惩罚对所有预测变量一视同仁, 而不受其原始量纲影响。系数会按原始量纲返回。通常应保持开启。

7.5 符号约束

启用 **Enforce Sign Constraints** 后，系数会被约束为与变量表（第 6 章）中设定的预期符号一致：

- “positive”（正）符号强制系数 ≥ 0
- “negative”（负）符号强制系数 ≤ 0
- 设为“either”（不限）的变量不受约束

这是通过 `glmnet()` 中的 `upper.limits` 与 `lower.limits` 参数实现的。

7.6 松弛 Lasso

常规 lasso 对变量选择和系数估计使用同一个惩罚，这可能把重要系数过度收缩到零。松弛 lasso (relaxed lasso) 把这两步分开：

1. 首先，通过 lasso 选择变量（由惩罚决定哪些系数得以保留）
2. 然后，对保留下来的变量以更小的惩罚或不加惩罚重新拟合，得到偏差更小的估计

Gamma 滑块控制松弛的程度：

Gamma	效果
0	完全松弛（OLS 重新拟合，完全不作收缩）
1	不作松弛（与常规 glmnet 相同）
介于其间	在松弛拟合与带惩罚拟合之间取折中

在使用交叉验证时，gamma 会被自动选取以获得最优表现。

7.7 随机种子

Random seed（随机种子）文本框预填了一个随机整数，用于使交叉验证的折分配可复现——拟合前会调用 `set.seed()`，所用的种子会显示在 **Glmnet Output** 选项卡以及拟合状态消息中。每次拟合之后会自动生成一个新种子（使用系统时间，而非 R 的随机数发生器）。当前文件最近使用的 5 个种子会以可点击链接的形式列出，便于快速调回；载入新文件时种子历史会重置。

7.8 交互项矩阵

一个可交互的上三角矩阵让您控制哪些变量对之间允许产生交互作用。每个单元格内有一个复选框——勾选表示交互项 $x_1:x_2$ 会被纳入模型矩阵。

- **Allow All**（全部允许）——勾选所有交互变量对
- **Clear All**（全部清除）——取消勾选所有变量对（新文件的默认状态是全部不勾选）
- 点击变量名——切换该变量的所有交互项

交互项的选择会按输入文件分别保存到浏览器 `localStorage`，并在您重新载入同一文件时恢复。

只要启用了任何交互项，导出的报告中就会包含一张 **Enabled Interactions**（已启用交互项）矩阵（两个坐标轴都是预测变量，纳入 $x_1:x_2$ 乘积项之处打勾），以横向页面呈现，与 `earthUI` 的 **Allowed Interactions** 矩阵相对应。请注意，lasso 仍可能把某个已启用项的系数收缩为零。

提示：当预测变量很多时，lasso 会自动把不需要的交互项归零。先用 “Allow All” 起步，再让模型自行选择，是一种合理的做法。

警告

启用 glmnet 交互项会产生在法庭或审计场合难以解释和说明的调整值。请谨慎使用交互项，并核实所得调整值是否合理。

7.9 禁止进入主效应（仅作交互项）

在交互项矩阵中右键点击某个变量名，即可禁止它以主效应的身份进入模型。变量名后会出现一个粗体的“1”标记。被禁止时：

- 该变量会从主公式项中移除
- 该变量仍可出现在交互项（x1:x2）中
- 再次右键点击即可解除禁止
- 该状态按输入文件通过 localStorage 持久保存

一个典型用例：在为应随物业面积按比例变化的时间调整建模时，禁止 sale_age 进入主效应，只允许它与 living_area 产生交互。这样得到的是 sale_age:living_area 交互项，而不是一个平坦的 sale_age 系数——即时间增值率随房屋面积而异。

7.10 高级参数

glmnet 调用参数底部有一个可折叠的“Advanced”（高级）小节，其中是一些附加设置。它们采用合理的默认值，但仍然可见，以便所有模型设置都能为法庭或审计留下书面记录。

- **Lambda Min Ratio** ——正则化路径中最小 lambda 与最大 lambda 之比。默认值：0.00001。对于小数据集（观测数 <50），取较大的值（例如 0.01）可能表现更好。
- **Number of Lambda Values** ——路径中 lambda 值的个数。默认值：100。
- **CV Loss Metric** ——交叉验证所用的损失函数：MSE（gaussian 的默认值）、MAE（对离群值稳健）或 Deviance（偏差）。
- **Convergence Threshold** ——坐标下降的收敛阈值。默认值：1e-07。
- **Max Iterations** ——坐标下降的最大迭代次数。默认值：100,000。若出现收敛警告，可调大此值。
- **Fit Intercept** ——是否包含截距（常数）项。默认：开启。在估价 RCA 中，截距即“基准值”。

7.11 设置默认值

有三个选项控制载入文件时参数如何初始化：

- **Use last settings for input file** ——恢复上次处理此文件时所用的确切设置
- **Use default settings** ——应用您保存的全局默认值
- **glmnet defaults** ——重置为出厂设置（alpha = 1、CV lambda.1se、10 折、gaussian、开启标准化）

点击 **Save current as default**，即可把当前全部设置存为今后文件所用的全局默认值。

8 拟合模型

8.1 拟合按钮

侧边栏的第 6 节包含 **Fit Glmnet Model** (拟合 Glmnet 模型) 按钮。点击它即以您当前的配置运行模型。

8.2 拟合过程中发生了什么

当您点击 Fit 时, glmnetUI 会:

1. 准备数据——剔除含 NA 的行, 施加权重 (若已指定), 排除标的行 (估价模式下), 并把因子变量编码为哑变量列。
2. 构建模型矩阵——生成设计矩阵, 含交互项矩阵中所有被允许的交互项。
3. 施加约束——若启用了符号约束, 则为系数设定 `upper.limits` 与 `lower.limits`。
4. 运行交叉验证——若选择了 CV, 则调用 `cv.glmnet()` 寻找最优 `lambda`。若启用了松弛 lasso, 则调用 `cv.glmnet(..., relax = TRUE)`。
5. 拟合最终模型——保存拟合好的模型、系数、预测值与残差。
6. 更新所有结果选项卡——填充 Summary、Coefficients、Equation、Variable Importance、Contributions、ANOVA 与 Diagnostics。

Fit 按钮下方的状态消息会显示观测数量、被排除的行数以及选定的 `lambda`。

9 结果选项卡

9.1 Data Preview (数据预览)

以交互式 DataTable 显示导入的数据。在估价 / 市场模式下, 预览分为两张表: **Subject Property** (标的物业, 第 1 行) 与 **Comparable Sales** (可比成交实例, 第 2 行及以后)。

每个单元格都保持单行显示, 并按列宽截断, 因此较宽的自由文本字段 (如物业备注) 不会把行撑得很高。双击任一单元格可弹出窗口, 显示其完整、未截断的内容。

9.2 Equation (方程)

显示通过 MathJax 以 LaTeX 渲染的拟合模型方程。展示所有非零系数, 并采用规范的数学排版:

- 正系数以 + 号显示
- 负系数以 - 号显示
- 交互项显示为 $x_1 \times x_2$
- 数字带千位逗号, 并保留适当的小数位

9.3 Correlation (相关性)

一张热力图矩阵, 展示所有数值型预测变量与响应变量之间的 Pearson 相关系数。数据导入后立即可用——无需拟合模型。

- 颜色梯度: 蓝色 (-1) → 白色 (0) → 红色 (+1)
- 每个单元格中打印相关系数值
- 文字与坐标轴尺寸随变量数量自适应
- 有助于在拟合前识别多重共线性

9.4 Summary (摘要)

以卡片形式显示的模型拟合统计量:

- R^2 ——被解释方差的比例
- $\text{Adj } R^2$ ——按预测变量个数调整后的 R^2
- GR^2 ——广义 R^2 , 由 glmnet 在选定 lambda 处的偏差比得出
- $\text{CV } R^2$ ——交叉验证 R^2 (使用 CV 时)
- **RMSE** ——均方根误差
- **MAE** ——平均绝对误差

在估价 / 市场模式下, 还会显示以下卡片:

- **Median Ratio** —— (预测值 / 实际值) 的中位数
- **COD** ——离散系数 (Coefficient of Dispersion, 越低越好)
- **PRD** ——价格相关差异 (Price-Related Differential)

当训练 R^2 超出 $\text{CV } R^2$ 达 0.1 以上时, 会出现过拟合警示。

卡片下方有一张带符号警示的系数表, 为方便查阅而重复了 Coefficients 选项卡的内容。

9.5 Coefficients (系数)

一张列出所有非零系数的表, 显示:

- 变量名
- 系数值
- 预期符号 (来自变量配置)
- 若系数方向与预期不符, 则给出符号警示标记

9.6 Variable Importance (变量重要性)

标准化的系数量级：对每个预测变量计算 $|\beta| \times \text{sd}(x)$ ，因子变量则跨其哑变量列汇总。有两种显示模式：

- 交互式 **plotly** (安装了 **plotly** 软件包时)：水平条形图，悬停提示显示精确的重要性、系数与相对百分比。
- 静态 **ggplot2** (备用方案)：水平条形图，坐标轴标签带千位逗号。

图表下方有一张 **DataTable**，显示 **Variable**、**Importance**、**Coefficient** 与 **Relative %**。

9.7 Contributions (贡献)

逐变量的偏效应图，展示每个预测变量对模型预测值的贡献。请从下拉框中选择一个变量。

对数值型预测变量：一张该变量的 x 值对其贡献 ($\beta \times x$) 的散点图，附一条红色线段和一个斜率标注。斜率标注显示每单位的边际效应 (例如 $+1,234.56/\text{unit}$)，对于纬度 / 经度这类取值范围很小的变量，单位会自适应调整 (例如 $+0.12/0.001$)。副标题显示截距 (基准) 值。

对因子变量：一张箱线图，展示各因子水平上贡献值的分布。

对双变量交互项：三幅上下叠放的可视化图——联合贡献的三维曲面图 (可用时为交互式 **plotly**，否则为静态 **persp()**)、按贡献着色的散点图，以及跨两个变量取值范围的平均贡献热力图。

9.8 ANOVA (方差分析)

方差分解表，为每个预测变量显示：

- **SS** ——平方和： $\sum (\text{contribution}_j - \overline{\text{contribution}_j})^2$
- **% of Model SS** ——占模型总平方和的百分比
- **Coefficient** ——该预测变量的主导系数

表中含一行截距和一行 **TOTAL MODEL** (模型合计)。交互项 (名称中含“:”者) 作为单独的组显示。

9.9 Diagnostics (诊断)

四幅诊断图，采用大字号 (基准字号 16pt)、15 个坐标轴刻度，标签带千位逗号 (不使用科学计数法)：

1. 系数路径 (Coefficient Path) ——展示随着 $\log(\lambda)$ 增大，所有系数如何变化。曲线按变量着色。有助于直观理解正则化路径与变量选择过程。
2. **CV 误差** (CV Error) ——交叉验证误差随 $\log(\lambda)$ 的变化。散点为 CV 误差均值；误差棒为 ± 1 个标准误。虚线竖线标出 $\lambda_{\min}$ (蓝色) 与 λ_{1se} (绿色)。
3. 实际值对预测值 (Actual vs Predicted) ——观测值与预测值的散点图。散点应聚集在 45 度虚线参照线附近。散布越宽，说明预测误差越大。
4. 残差对拟合值 (Residuals vs Fitted) ——残差与拟合值的散点图。应呈现围绕零线 (虚线) 的随机散布。若出现某种模式，则提示模型设定有误 (例如非线性、异方差)。

每幅图都有一个 **Download PNG** 按钮，可按 150 DPI 保存。

9.10 Glmnet Output (Glmnet 输出)

原始模型输出，显示：所用随机种子、模型打印结果、选定的 λ 、 $\lambda_{\min}$ / λ_{1se} 、 γ (若启用松弛 **lasso**)，以及完整的系数向量。

9.11 Report (报告)

配置报告字段（估价师姓名、物业地址、报告日期、档案编号），并可直接从该选项卡导出为 Word (.docx) 或 PDF。完整的基于 Quarto 的报告工作流程（HTML、Word 或 PDF）见第 13 章。

10 下载数据

拟合之后，**Download Output (Excel)**（下载输出，Excel）按钮（侧边栏第 7 节，在所有用途模式下均可用）会把含预测值与诊断信息的 Excel 文件导出到项目的输出文件夹。该小节标题在估价与市场模式下为“Download Estimated Sale Prices & Residuals”（下载估计成交价格与残差），在通用模式下为“Download Estimated Target Variable(s) & Residuals”（下载估计目标变量与残差）。

10.1 输出列

列	说明
est_<target>	模型预测值（例如 est_sale_price）
residual	实际值 - 预测值
cqa	状况-品质-吸引力分数（Condition-Quality-Appeal，0-10 分制）
residual_sf	残差 / 居住面积（若已指定 living_area）
cqa_sf	依据 residual_sf 排名计算出的 CQA
<var>_contribution	各变量对预测值的贡献
basis	截距值（所有行相同）

成功完成后，下载按钮上会出现一个白色对勾。

10.2 CQA 分数

CQA 把每一行的残差与其余各行相比较，按 0-10 分制排名：

- **CQA 高（~9-10）**：成交价远高于预测值——很可能状况 / 品质 / 吸引力更优
- **CQA 低（~0-1）**：成交价远低于预测值——很可能状况较差或属不良资产处置
- **CQA ~5**：接近中位数

在估价 / 市场模式下，若指定了 living_area 列，各行会按 residual_sf 降序排列。

11 RCA 计算与下载

RCA (Residual Constraint Approach, 残差约束法) 工作流程在 **Appraisal** (估价) 与 **Market Area Analysis** (市场区域分析) 模式下、模型拟合之后可用。后续的销售比较表格 (第 12 章) 仅限估价模式。

11.1 打开 RCA 对话框

点击侧边栏第 8 节的 **Calculate RCA Adjustments & Download** 按钮。然后选择：

- 分数类型：**CQA** 或 **CQA per SF** (每平方英尺 CQA)。若选择“CQA per SF”，标的的残差会按居住面积缩放。
- 输入标的的 CQA 分数 (0.00–9.99，默认 5.00)

11.2 CQA 分数插值

1. 对各可比实例的 CQA 分数与残差进行排序
2. 用线性插值把您给定的 CQA 值映射到一个残差
3. 标的价值 = 模型预测值 + 插值所得残差

11.3 输出列

列	说明
subject_value	模型预测值 + 插值所得残差
<var>_adjustment	标的贡献 – 可比实例贡献
residual_adjustment	标的残差 – 可比实例残差
net_adjustments	所有调整值之和
gross_adjustments	各调整值绝对值之和
adjusted_sale_price	可比实例成交价格 + 净调整额

计算成功后，按钮上会出现一个白色对勾。

12 销售比较表格

销售比较表格 (Sales Comparison Grid) 仅在 **Appraisal** (估价) 模式下、计算完 RCA 调整值 (第 8 步) 之后可用。它会生成一份格式化的 Excel 工作簿, 可直接编入估价报告。

12.1 可比实例选择对话框

点击侧边栏第 9 节的 **Generate Sales Grid & Download** 按钮, 会弹出一个模态对话框, 列出 RCA 输出中的所有可比成交实例。可比实例分为两组:

- **Recommended comps** (推荐的可比实例, 预先勾选): 毛调整额 < 成交价格的 25%, 按成交时距排序 (最新在前)。
- **Other comps** (其他可比实例, 默认不勾选): 其余的可比实例, 按成交时距排序。

最多可选 30 个可比实例, 然后点击 **Generate Sales Grid** 生成工作簿。

12.2 工作簿版式

生成的工作簿最多含 10 个工作表, 每个工作表并排放置 3 个可比实例:

- 表头行: 标的与可比实例的地址、成交价格、成交日期以及在市天数
- 分组行: 位置 (纬度、经度、区域)、宗地面积 (地块面积、宗地尺寸) 与年龄 (实际年龄、有效年龄), 带组标题与子项目
- 模型变量行: 每个非零模型系数一行, 显示标的贡献、各可比实例贡献以及调整百分比
- 剩余特征行: 留空且未锁定的单元格, 供估价师填写模型未涵盖的特征 (例如状况、品质、景观)
- 汇总行: 净调整百分比、毛调整百分比与调整后成交价格, 均带有实时的 Excel 公式

12.3 工作表保护与剩余特征单元格

每个工作表都设有密码保护以防误改, 但剩余特征的取值单元格是明确解锁的。这样估价师既能填入模型之外的特征取值, 又能保留由公式驱动的调整计算。

13 下载报告

13.1 两步式 Quarto 工作流程

从侧边栏导出报告采用两步式的 Quarto 工作流程：

1. **Generate Quarto Report** (生成 Quarto 报告, 侧边栏第 10 节) —— 把一个自包含的 Quarto 报告包 (.qmd 源文件加上图形、report_data.rds 与 reference.docx) 写入项目的 glmnet 输出文件夹。
2. **Convert Quarto Report** (转换 Quarto 报告, 侧边栏第 11 节) —— 把 .qmd 文件渲染为您选择的输出格式。路径默认指向最近生成的报告包, 但您也可以浏览选择任何一个 .qmd。

13.2 报告格式

转换步骤中有三种格式可选：

- **HTML** —— 自包含的 HTML 文档。在所有平台上都能使用, 除 pandoc 外无需额外软件。推荐的默认选项。
- **Word (.docx)** —— 格式化文档, 便于编辑与分发。在所有平台上均可使用。
- **PDF** —— 通过 Quarto 配合 LaTeX 生成。纸张尺寸遵循区域设置 (Letter 或 A4)。需要安装 LaTeX (参见系统要求附录)。若未检测到 LaTeX, 格式复选框中会自动隐藏 PDF 选项。

有 Quarto 可用时使用 Quarto 生成报告, 否则回退到 rmarkdown。Report 选项卡另外还提供直接导出为 Word 和导出为 PDF 的按钮。

13.3 报告内容

报告涵盖所有选项卡的内容：

- 数据集说明 (文件名、观测数、响应变量、预测变量、用途)
- 模型设定 (alpha、lambda、family、standardize、是否松弛、非零系数)
- 结果摘要 (R^2 、调整 R^2 、广义 R^2 、CV R^2 、RMSE、MAE)
- 模型方程 (以 LaTeX 渲染)
- 系数表
- 变量重要性 (条形图与表格)
- 与 Contributions 选项卡一致的逐变量贡献图：
 - 数值型预测变量：带线性拟合线的散点图
 - 因子变量：按水平分组的箱线图
 - 交互项：按贡献着色的散点图、平均贡献热力图, 以及静态三维曲面快照 (persp)
- Enabled Interactions (已启用交互项) 矩阵 (启用了任何交互项时) —— 两个坐标轴都是预测变量, 纳入 $x_1:x_2$ 乘积项之处打勾, 以横向页面呈现
- 相关系数矩阵热力图
- ANOVA 分解表
- 诊断图：
 - 系数路径
 - 交叉验证误差
 - 实际值对预测值
 - 残差对拟合值
 - 正态 Q-Q 图
- 模型输出

所有坐标轴标签与颜色图例中的数字均带千位逗号（不使用科学计数法）。
报告生成成功后，报告按钮上会出现一个白色对勾；渲染过程中会显示已用时间计时器。

14 与 earthUI 的比较

earth 应用程序及其 glmnet 后处理模块同属一套回归建模工具。它们共用相同的数据格式、特殊列类型、RCA 工作流程和演示数据集，但使用不同的建模引擎。

14.1 主要差异

特性	glmnetUI	earthUI
方法	弹性网络 (glmnet)	MARS/Earth (earth)
变量选择	经由 lasso 惩罚自动进行	经由前向 / 后向剪枝自动进行
系数关系	仅限线性效应	分段线性 (铰链函数)
交互项	经由交互项矩阵两两组合	最高 3 阶
系数符号	可通过约束强制施加	无法直接强制施加
松弛拟合	可使用松弛 lasso	不适用
Alpha 调优	可用网格搜索	不适用 (无混合参数)
销售比较表格	可用 (第 9 步)	可用 (第 8 步)
g 函数图	不适用	归组项可视化
报告格式	HTML、Word、PDF	HTML、Word、PDF

14.2 共有的功能

两个工具都提供：

- CSV 与 Excel 导入，并把列名转换为 snake_case
- 支持 31 个国家的区域设置，含 CSV 分隔符、小数点标记与纸张尺寸
- 特殊列指定 (contract_date、living_area、area、latitude、longitude 等)
- 依据基准日自动计算 sale_age
- 深色 / 浅色模式切换
- 通过 localStorage 持久保存设置
- 相关性热力图、变量重要性、贡献图、ANOVA、诊断图
- CQA 评分与 RCA 调整工作流程
- 带可比实例选择和格式化 Excel 输出的销售比较表格
- 兼容同一批演示数据集 (Appraisal_1.csv)

14.3 何时该用哪一个

- 使用 **glmnetUI**，当您想要一个简洁、可解释、并能自动选择变量的线性模型时。当预测变量很多而您希望由模型挑出最重要的那些时，弹性网络尤其合适。
- 使用 **earthUI**，当您预期存在非线性关系（例如居住面积对价格的影响在不同面积区间有所不同），或需要用铰链函数刻画阈值效应时。earth 模型还能刻画更高阶的交互作用（2 阶和 3 阶）。此外，earthUI 还提供一项基于 Prolog/DCG 的实验性备注抽取功能，这在 glmnetUI 中没有对应项。
- 两者并用，以在同一份数据上比较线性模型与非线性模型。若两者给出的 R^2 相近，出于可解释性考虑，可能更宜采用较简洁的 glmnet 模型。

15 演示数据集：Appraisal_1.csv

15.1 说明

演示数据集随 glmnetUI 一同发布（并与 earthUI 共用）。可用以下方式定位：

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "glmnetUI")
```

该文件包含 1,502 笔住宅成交记录（外加第 1 行的 1 处标的物业），取自一份模拟的 MLS 导出文件。数据代表一个多区域市场中的独栋住宅成交，涵盖不同的物业面积、房龄与位置。

这并非真实数据，但以北加州一个真实街区为原型。所有身份识别信息均已更改或移除。

15.2 各列

列	类型	特殊类型	说明
weight	numeric	weight	观测权重（0 = 不参与拟合）
id	numeric	display_only	内部记录 ID
property_id	numeric	display_only	MLS 物业标识
listing_id	character	display_only	MLS 挂牌编号
parcel_number	character	display_only	县估税官地块编号（APN）
street_address	character	display_only	物业地址
city_name	character	display_only	城市
postal_code	character	display_only	邮政编码
county_name	character	display_only	县
contract_date	Date	contract_date	成交合同日期（用于计算 sale_age）
sale_age	numeric	—	从合同日期到基准日的天数
sale_price	numeric	（目标变量）	成交价格——响应变量
living_sqft	numeric	living_area	总居住面积（平方英尺）
beds_total	integer	—	卧室数量
baths_total	numeric	—	卫生间总数（例如 2.5 = 2 全卫 + 1 半卫）
lot_size	numeric	lot_size	地块面积（平方英尺）
area_id	integer	area	MLS 区域标识
age	numeric	actual_age	物业年龄（年）
latitude	numeric	latitude	纬度（入模前四舍五入到小数点后 3 位）
longitude	numeric	longitude	经度（入模前四舍五入到小数点后 3 位）
garage_spaces	integer	—	车库车位数
days_on_market	integer	dom	在市天数
listing_date	Date	listing_date	挂牌日期
sale_concessions	numeric	concessions	卖方让利

15.3 建议的快速上手流程

1. 启动 glmnet 例程：earthUI::launch_glmnet()
2. 新建一个用途为 **For Appraisal** 的项目（第 1 节）
3. 把 Appraisal_1.csv 放入项目的输入文件夹，并在第 2 节中选中它（必要时点击 **Re-fresh**）
4. 选择 sale_price 作为目标变量
5. 按上表所示指定各列的特殊类型
6. 纳入以下预测变量：sale_age、living_sqft、baths_total、lot_size、area_id（作为因子）、age、latitude、longitude、garage_spaces
7. 保持 alpha = 1（Lasso），使用 CV 配 lambda.1se

8. 点击 **Fit Glmnet Model**
9. 查看 Summary、Coefficients 与 Diagnostics 选项卡
10. 下载输出 (第 7 步), 查看 CQA 排名
11. 以约 ~5.00 的 CQA 分数计算 RCA 调整值 (第 8 步)
12. 用推荐的可比实例生成销售比较表格 (第 9 步)

16 系统要求与故障排查

16.1 受支持的平台

glmnetUI 可在 **macOS**、**Windows** 与 **Linux** 上运行。强烈建议使用 RStudio Desktop (2023.06 及以上版本) ——它自带生成报告所需的 **pandoc**。

16.2 可选依赖

- **LaTeX** (TinyTeX、MiKTeX 或 MacTeX): 仅用于生成 PDF 报告。若未检测到, PDF 选项会自动隐藏。安装方式: `tinytex::install_tinytex()`
- **sysfonts + showtext**: 为图形提供 Roboto Condensed 字体。若不可用或处于离线状态, 会自动回退到系统无衬线字体。
- **earth** ($\geq 5.3.0$): 仅在使用 earthUI 导入通道时需要。

16.3 各平台说明

macOS: 问题最少。安装 TinyTeX 以生成 PDF: `tinytex::install_tinytex()`

Windows: 与 RStudio 配合良好。公司配发或权限受限的机器可能对临时目录有限制——应用程序会在控制台给出警告, 但仍会继续运行, 只是设置不能持久保存。

Linux: 可能需要安装系统库: `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

16.4 降级运行

缺失的组件	表现
没有 LaTeX	隐藏 PDF 选项。HTML 与 Word 仍可使用。
没有网络 / 字体加载失败	使用系统无衬线字体。控制台记录相应消息。
没有 RSQLite / 文件系统只读	设置不能持久保存。应用程序照常运行。
临时目录不可写	报告生成失败并给出明确错误。请把 TMPDIR 设为可写位置。

16.5 故障排查

“**PDF option not available**”(PDF 选项不可用): 在 R 中运行 `tinytex::install_tinytex()`, 然后重启应用程序。

“**Settings will not persist**”(设置无法保存): 检查用户数据目录的权限 (macOS: `~/Library/Application Support/R/glmnetUI/`, Linux: `~/.local/share/R/glmnetUI/`, Windows: `%APPDATA%/R/data/glmnetUI/`)。

端口 **7879** 已被占用: 运行 `lsof -ti:7879 | xargs kill(macOS/Linux)` 或 `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell)。

参考文献

- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1, pp. 1–22. DOI: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01).
- Simon, Noah et al. (2011). “Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent”. In: *Journal of Statistical Software* 39.5, pp. 1–13. DOI: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).

Tay, J. Kenneth, Balasubramanian Narasimhan, and Trevor Hastie (2023). “Elastic Net Regularization Paths for All Generalized Linear Models”. In: *Journal of Statistical Software* 106.1, pp. 1–31. DOI: [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).

mgcvUI: mgcv (广义可加模型) 软件包的交互式界面

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

William Bert Craytor

第 1 卷，第 1 期 · 2026 年夏季

发表日期: 2026 年 7 月 1 日

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

本开放获取文章依据 [知识共享署名 4.0 国际许可协议 \(CC BY 4.0\)](#) 的条款发布。

摘要。 mgcvUI 是 R 语言 mgcv 软件包的图形用户界面 (graphical user interface)，该软件包使用带自动光滑度选择 (smoothness selection) 的惩罚回归样条 (penalized regression spline) 来拟合广义可加模型 (generalized additive model, GAM)。它提供三种用途模式——通用预测建模、房地产估价与市场区域分析——并引导用户完成数据导入、光滑项设定、模型拟合、诊断图与效应图查看以及报告下载。本文记录了 mgcvUI 的数据格式要求、建模工作流程、输出显示界面以及完整的功能参考。

关键词： 广义可加模型，GAM，惩罚样条，光滑化，mgcv，R，图形用户界面，房地产估价

出版者按 (2026 年 7 月): 付印前不久，mgcvUI 已并入 earthUI (0.11.0 及以后版本)，本文所述的 GAM 工作流程现与 MARS 和弹性网络工作流程一同由 earthUI 提供。mgcvUI 仍以其最终发布形态 (0.3.0) 保留在 <https://github.com/wcraytor/mgcvUI>；后续开发将在 earthUI 中继续。本文适用于已发布的 mgcvUI。

目录

1	引言	4
1.1	历史与背景	4
1.1.1	广义可加模型	4
1.1.2	mgcv 软件包	4
1.1.3	与 MARS (earth) 和弹性网络 (glmnet) 的比较	4
1.1.4	earthUI-mgcvUI 流水线	5
1.1.5	推荐的工作流程	5
1.2	什么是 mgcvUI?	5
1.3	三种用途模式	6
1.4	房地产专用功能	6
1.5	入门	6
2	MLS 输入数据要求	8
2.1	文件格式与结构	8
2.2	估价模式的必需列	8
2.3	特殊列命名约定	9
2.4	数据质量与完整性	9
2.5	标的物行的位置	9
3	通用模式	10
3.1	概述	10
3.2	侧边栏工作流程	10
3.3	从 earthUI 导入 (可选)	11
3.4	主面板标签页	11
3.5	设置的持久化	12

3.6	暗色模式	12
3.7	区域与地区设置	12
4	估价模式	14
4.1	标的物行的处理	14
4.2	估价基准日与销售时距	14
4.3	特殊列指定	14
4.4	RCA 调整概览	14
5	市场区域分析模式	15
5.1	与估价模式的差异	15
5.2	何时使用市场模式	15
6	变量选择	16
6.1	目标变量	16
6.2	响应变量变换	16
6.3	预测变量表	16
6.4	数据类型检测与覆盖	16
6.5	特殊列类型参考	17
7	参数选择	18
7.1	参数预设	18
7.2	Family (分布族)	18
7.3	Method (方法)	18
7.4	Gamma (光滑度乘子)	18
7.5	交叉验证	19
7.6	Select (收缩)	19
7.7	默认基	19
7.8	默认 k (基维数)	19
7.9	张量类型	19
7.10	允许的交互	20
7.11	因子-光滑交互	20
7.12	高级参数	20
8	拟合模型	21
8.1	Fit 按钮	21
8.2	拟合过程中发生了什么	21
8.3	后台拟合与进度显示	21
9	结果标签页	22
9.1	Data (数据)	22
9.2	Equation (方程)	22
9.3	Summary (摘要)	22
9.4	Variable Importance (变量重要性)	22
9.5	贡献图	22
9.5.1	单变量光滑项 — $s(x)$	23
9.5.2	因子-光滑项 — $s(x, \text{by} = \text{factor})$	23
9.5.3	交互项 — $ti(x_1, x_2)$ 与 $te(x_1, x_2)$	23
9.5.4	参数项 — 线性项与因子项	23
9.6	Correlation (相关性)	23
9.7	Diagnostics (诊断)	23
9.8	RCA Adjustments (RCA 调整)	23

9.9 Anova (方差分析)	24
9.10 Mgcvc Output (mgcv 输出)	24
9.11 Sign Check (符号检查)	24
9.12 Concurvity (共曲性)	24
9.13 Basis Check (基检验)	24
10 下载数据	25
10.1 输出列	25
10.2 CQA 得分	25
11 RCA 计算与下载	26
11.1 打开 RCA 对话框	26
11.2 CQA 得分插值	26
11.3 输出列	26
12 销售比较网络	27
12.1 可比实例选择对话框	27
12.2 工作簿版式	27
12.3 工作表保护与残差单元格	27
13 下载报告	28
13.1 第 1 步: 生成 Quarto 报告	28
13.2 第 2 步: 转换 Quarto 报告	28
13.3 报告内容	28
14 导出模型函数	29
14.1 R 函数	29
14.2 Python 函数	29
14.3 C++ 头文件	29
14.4 SQLite 数据库	29
15 与 earthUI 和 glmnetUI 的比较	30
15.1 主要差异	30
15.2 共有功能	30
15.3 推荐的工作流程	30
16 演示数据集: Appraisal_1.csv	31
16.1 说明	31
16.2 各列	31
16.3 建议的快速上手步骤	32
17 系统要求与故障排除	33
17.1 受支持的平台	33
17.2 R 版本	33
17.3 必需的 R 软件包	33
17.4 可选依赖	33
17.5 与平台相关的说明	33
17.5.1 macOS	33
17.5.2 Windows	34
17.5.3 Linux	34
17.6 优雅降级	34
17.7 故障排除	34
17.7.1 “PDF option is not available” (PDF 选项不可用)	34

17.7.2 “Settings will not persist” (设置将不会保存)	35
17.7.3 “Diagnostics plot unavailable” (诊断图不可用)	35
17.7.4 “Report error: number of items to replace...” (报告错误: 待替换的项数)	35
17.7.5 7880 端口已被占用	35
17.7.6 在 Linux 上安装软件包失败	35

1 引言

1.1 历史与背景

1.1.1 广义可加模型

广义可加模型 (generalized additive model, GAM) 由 Trevor Hastie 与 Robert Tibshirani 在其 1990 年的专著中正式提出。GAM 对线性模型作了推广: 把每一个线性项 $\beta_j x_j$ 替换为一个光滑函数 (smooth function) $f_j(x_j)$, 于是模型成为:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

每个 f_j 都用惩罚回归样条 (penalized regression spline) 从数据中估计得到。这些光滑函数使模型无需用户事先指定函数形式即可刻画非线性关系。惩罚项控制每个光滑项的摆动程度 (wiggliness), 在保留足够灵活性以追踪真实模式的同时防止过拟合。

1.1.2 mgcv 软件包

R 软件包 mgcv (Mixed GAM Computation Vehicle) 由巴斯大学的 Simon Wood 开发 (Wood 2017)。它于 2001 年首次发布, 如今已是 GAM 在 R 中的标准实现, 并随每一份 R 安装一同发行。其底层方法见于一系列论文, 涵盖稳定的多光滑参数估计 (Wood 2004)、薄板回归样条 (Wood 2003)、快速稳定的 REML 估计 (Wood 2011) 以及一般光滑模型的选择 (Wood, Pya, and Säfken 2016)。主要特性包括:

- 自动光滑度选择 (smoothness selection): 每一项的光滑参数都作为模型拟合的一部分加以估计, 可采用 REML (受限极大似然)、GCV (广义交叉验证) 或 ML (极大似然)。
- 多种基类型: 薄板回归样条 (tp)、三次回归样条 (cr)、P 样条 (ps) 与 B 样条 (bs), 各具不同性质。
- 张量积光滑: `te()` 与 `ti()` 可实现变量间的光滑交互, 而不受维数灾难之困。
- 变量选择: `select = TRUE` 选项会附加一个额外惩罚, 可将光滑项收缩至零, 从而实际完成变量选择。
- 共曲性诊断 (concurvity diagnostics): 度量光滑项中与多重共线性相对应的现象。

1.1.3 与 MARS (earth) 和弹性网络 (glmnet) 的比较

earth 应用程序及其两个后处理模块 (glmnet 与 mgcv) 使用不同的建模引擎。下表概括了主要差异:

特性	mgcv (GAM)	earth (MARS)	glmnet (弹性网络)
模型类型	光滑非线性 (惩罚样条)	分段线性 (自适应样条)	线性 (带正则化)
非线性	通过光滑函数自动实现 (薄板、三次等)	通过带数据驱动节点的铰链函数自动实现	仅能通过交互项或基展开实现
变量选择	可选, 通过双重惩罚收缩实现 (select=TRUE)	内置, 通过前向/后向逐步法配合 GCV 剪枝实现	内置, 通过 L_1 惩罚 (Lasso) 实现
交互项	张量积光滑 (te()、ti())	基于铰链函数的乘积 (可解释, 最高三阶)	两两交叉乘积 (难以解释)
可解释性	光滑的偏效应曲线; 非常直观	g 函数按变量给出分段线性偏效应	系数即线性权重; 易于逐一说明
最适用于	光滑的非线性关系、灵活建模	自动检测非线性与交互	高维数据、变量选择、小样本
RCA 调整	通过偏效应给出光滑调整	通过 g 函数给出分段线性调整	按变量给出线性调整
过拟合控制	REML/GCV 光滑参数估计	基于 GCV 的剪枝	交叉验证选取的 lambda

MARS (Multivariate Adaptive Regression Splines, 多元自适应回归样条) 由 Jerome Friedman 于 1991 年提出。它通过从数据中自适应地选择铰链函数及其节点 (knot) 位置来构建分段线性模型。其 R 实现是 Stephen Milborrow 编写的 **earth** 软件包。

弹性网络 (elastic net) 回归由 Friedman、Hastie 与 Tibshirani (2010) 实现于 **glmnet** 软件包中, 它把 Lasso (L_1) 惩罚与岭回归 (L_2) 惩罚结合起来, 同时完成变量选择与系数收缩。

1.1.4 earthUI-mgcvUI 流水线

一套强有力的工作流程把 earthUI 的自动节点发现与 mgcvUI 的光滑估计结合起来。当 earthUI 导出一个 .rds 结果文件而 mgcvUI 将其导入时:

1. **earth** 发现节点 ——MARS 为每个变量找出由数据驱动的铰链位置 (变点)。
2. **mgcvUI** 以其为锚点 ——earth 的节点成为 GAM 中三次回归样条 (cr 基) 的基础, 使光滑项一开始就拥有反映数据真实结构的节点位置。
3. **mgcv** 精化拟合 ——惩罚样条估计把 earth 的分段线性拟合加以光滑化, 在保留 earth 所发现的总体形状的同时, 给出光滑的偏效应曲线。

这条流水线把探索性建模 (earth) 与验证性建模 (GAM) 衔接起来, 兼取两种方法之长。

1.1.5 推荐的工作流程

对于房地产估价及其他要求模型可解释、可辩护的类似应用:

1. 先用 **earthUI** 发现数据中的重要变量、非线性关系与交互作用。earth 的自动基选择与 g 函数能提供极好的初步洞见。
2. 再用 **mgcvUI** 精化, 如果你希望得到更光滑的偏效应。把 earthUI 的节点导入 mgcvUI, 即可从分段线性模型平滑过渡到光滑模型。
3. 使用 **glmnetUI**, 当你需要激进的变量选择 (Lasso)、预测变量数目接近或超过观测数目, 或者你希望得到一个纯线性且带正则化、便于辩护的模型时。

1.2 什么是 mgcvUI?

mgcvUI 是 R 语言 mgcv 软件包的图形用户界面 (graphical user interface)。它以本地 Shiny 应用程序的形式运行——无需登录、无需服务器、也没有账户。你从 R 中启动它, 导入数据集 (CSV 或 Excel), 配置模型, 然后交互式地拟合 GAM。

自当前版本起, mgcvUI 不再是单独安装的软件包: 它随 earthUI 软件包一同发行, 作为其中的 GAM 后处理模块 (Post Processing Module), 用 `earthUI::launch_mgcv()` 启动。应用程序本身——以及本文所述的一切——并无改变; 改变的只是安装与启动命令。三个例程 (`earth`、`glmnet`、`mgcv`) 共用一棵项目树、一个设置数据库以及同一份支撑基础设施。

该应用程序提供完整的工作流程: 数据导入、含光滑项设定的变量配置、带后台处理的模型拟合、诊断图、光滑偏效应曲线, 以及可下载的 Word、PDF 或 HTML 格式报告。

广义可加模型把每个预测变量的线性关系 βx 替换为光滑函数 $f(x)$ 。这意味着:

- 无需指定函数形式——每种关系的形状由数据本身决定。
- 光滑的偏效应——每个预测变量的贡献是一条光滑曲线而非直线, 便于看清某变量在不同取值处如何影响响应变量。
- 自动光滑度选择——mgcv 通过 REML 或 GCV 确定每条曲线应有多摆动, 在拟合优度与过拟合之间取得平衡。

1.3 三种用途模式

每个项目都有一个用途 (purpose), 在侧边栏第 1 节 (**Project**, 项目) 中创建项目时选定, 此后即被锁定。当前活动项目的用途决定哪些工具与界面元素出现; 若要在别的模式下工作, 请创建或打开一个具有该用途的项目。三种模式为:

- **General** (通用)——适用于任何类型总体或数据集的 GAM 回归。这是默认模式, 提供完整的 GAM 工作流程, 不含任何领域专用的附加功能。
- **For Appraisal** (用于估价)——为房地产估价定制的 GAM 回归。增加了单一物业估价所特有的功能, 包括标的物业处理、特殊列指定以及残差约束法 (Residual Constraint Approach, RCA)。
- **Market Area Analysis** (市场区域分析)——为市场区域研究定制的 GAM 回归。增加了在既定市场中分析成组物业的功能, 并可选择处理标的物行。

在三种模式下, 核心建模引擎完全相同——你始终是在通过 `mgcv::gam()` 拟合 GAM。用途设置只控制哪些附加工具与界面元素可用。

1.4 房地产专用功能

当选定 **For Appraisal** 或 **Market Area Analysis** 时, mgcvUI 会激活若干为房地产分析而设计的功能:

- 特殊列指定——每个预测变量都可以被标记上某种特殊角色, 例如 `contract_date`、`dom`、`concessions`、`latitude`、`longitude`、`living_area`、`lot_size`、`actual_age`、`effective_age`、`area`、`site_dimensions` 或 `display_only`。这些指定决定该列在拟合与输出过程中的处理方式。
- **Sale Age** 列 (销售时距)——当某列被指定为 `contract_date` 且提供了 Effective Date (估价基准日) 时, mgcvUI 会计算一个 `sale_age` 列 (成交日期与基准日之间的天数), 并以它作为预测变量代入模型。
- **RCA** 计算 (仅估价模式)——在估价模式下, 模型拟合完成后 mgcvUI 可计算残差约束法 (RCA) 输出, 给出逐个可比实例的调整额、净调整与总调整汇总, 以及标的物业的调整后销售价格。

提示: *General* 用途模式适用于任何数据集——金融、科学、工程或房地产。估价模式与市场模式只是为房地产专业人士增添了一些便利功能。

1.5 入门

使用 mgcvUI 的步骤:

1. 在 R 中安装 **earthUI**: `install.packages("earthUI")` ——mgcv 例程已包含在内。(开发版与可选组件请参见本期 earthUI 一文中的“Installing earthUI and vProlog”一节。)
2. 启动应用程序: 在 R 中运行 `earthUI::launch_mgcv()`。应用程序会在你的网页浏览器中通过 7880 端口打开。
3. 在侧边栏第 1 节创建或打开项目。项目设定用途 (General、For Appraisal 或 Market Area Analysis) 与工作地点 (国家、州、县、市) 及其输入/输出文件夹。项目存放在共享的 `regProj` 根目录之下, 并与同族应用 earthUI 和 glmnetUI 共享。打开一个用途不同的项目, 正是切换模式的方式。
4. 在第 2 节导入数据: 从当前活动项目的输入文件夹中选择一个 CSV 或 Excel 文件。把文件放入该文件夹, 然后点击 **Refresh** (刷新) 以列出它们。
5. 从 **earthUI** 导入 (可选) ——第 3 节允许你导入一个 earthUI 结果 `.rds` 文件, 用 earth 发现的节点为 GAM 提供初值。若你没有 earthUI 结果, 跳过此步即可。
6. 配置变量 ——选择目标变量与预测变量, 设定数据类型, 指派特殊列角色, 并将变量标记为因子型或线性。
7. 设定 **mgcv** 参数 ——family、method、gamma、基类型、k、张量类型、交互项以及其他选项。
8. 拟合模型 ——点击“Fit Mgcvm GAM Model”(拟合 mgcv GAM 模型), 并在主面板中查看结果。
9. 导出 ——把预测结果下载为 Excel, 生成并转换 Quarto 报告, 或 (在估价模式下) 计算 RCA 调整额与销售比较网格。

设置会自动保存在一个 SQLite 数据库中, 并在你重新加载同一输入文件时恢复。

2 MLS 输入数据要求

对于房地产估价与市场分析工作流程而言，你的输入数据通常来自多重上市服务（Multiple Listing Service, MLS）的导出文件。本章说明预期的文件结构以及 mgcvUI 可以使用的各列。

2.1 文件格式与结构

mgcvUI 接受 CSV 与 Excel (.xlsx、.xls) 文件。导入时，列名会自动转换为 snake_case——例如，“Living SqFt”变为 living_sqft，“Contract Date”变为 contract_date，“Sale Price”变为 sale_price。这一规范化确保整个工作流程中列引用的一致性。导入时使用的 CSV 分隔符与小数点符号由区域设置决定（参见第 3 章“Locale & Regional Settings”，区域与地区设置）。

你的数据文件应是一张扁平表格，每个物业占一行，每个属性占一列。文件的第一行必须是列标题。

2.2 估价模式的必需列

虽然 mgcvUI 可以处理任意一组列，但完整的估价工作流程若具备以下各列将获益更多：

列	特殊类型	用途
Sale Price（销售价格）	（目标变量）	模型的响应变量
Contract Date（成交日期）	contract_date	用于计算 sale_age（距估价基准日的天数）
Listing Date（挂牌日期）	listing_date	在缺少 DOM 列时，与成交日期一起用于计算 DOM
Days on Market（市场停留天数）	dom	市场停留天数；在导出文件中显示
Concessions（成交让价）	concessions	销售让价；净售价 = 销售价格 - 让价
Living Area (SF)（居住面积）	living_area	启用按平方英尺计的残差 (residual_sf、cqa_sf)
Lot Size（地块面积）	lot_size	宗地面积列
Site Dimensions（宗地尺寸）	site_dimensions	与地块面积归为一组
Latitude（纬度）	latitude	四舍五入至 3 位小数
Longitude（经度）	longitude	四舍五入至 3 位小数
Area ID（区域标识）	area	市场区域 / 街区标识符
Actual Age（实际年龄）	actual_age	物业年龄
Effective Age（有效年龄）	effective_age	物业的有效年龄
Address（地址）	display_only	在导出文件中显示；不进入模型

电子表格的列名可以使用外语——而“特殊”名称使用英文，以便 R 程序对其作特殊处理。除此之外，你所给定的列名会出现在回归模型、图形中，若是做估价，还会出现在输出报告中。

并非所有列都是必需的。mgcvUI 会自行适应——若某列缺失，相应功能便自动略去。不过，对于房地产定价模型，以下几列强烈推荐：

1. **Sale Age**（销售时距）——成交日期与估价基准日之间的天数。若使用跨越数年的销售历史，sale_age 往往起着核心作用。

2. **Living Area**（居住面积）——也常被称作“Living Sqft”“GLA”（gross living area，总居住面积）。
3. **Total Bath Count**（卫生间总数）——全卫、四分之一卫、半卫与四分之三卫的总数。
4. **Garage Bays**（车库车位）或 **Garage Area**（车库面积）——车库泊位数或建筑面积。
5. **Lot Size**（地块面积）——物业的土地面积。
6. **Longitude, Latitude**（经度、纬度），以及若有的话 **Area ID**（区域标识），用于刻画地理上的价格变化。

2.3 特殊列命名约定

mgcvUI 依据列的特殊类型指定而非列名来识别各列。你可以在 MLS 导出文件中随意命名列——关键在于你要在变量配置表（第 6 章）中指派正确的特殊类型。

2.4 数据质量与完整性

- **缺失值 (NA)**：任一预测变量列或目标列中含 NA 的行会在拟合前自动删除。mgcvUI 会把数据子集化为模型实际使用的那些列，因此未被使用的列中的 NA 不会导致丢行。
- **数值列**：销售价格、居住面积、地块面积等类似字段必须为数值型。
- **因子列**：文本列、TRUE/FALSE 列以及 R 因子列会被自动视为分类变量——它们只能以因子形式进入模型。一个数值列（例如数值型的区域码或风格码）只有在你勾选其 **Factor**（因子）复选框时才会成为因子；mgcvUI 绝不会根据数值的取值范围推断其为分类变量，因此像 `bath_count` 这样的离散数值预测变量仍保持连续（并被光滑化）。Special（特殊）角色与因子的判定彼此独立。

提示：导入后请查看预测变量表中的 NA 列。缺失值很多的列可能导致丢行。NA 占比 $\geq 50\%$ 的列会被标红并自动排除。

2.5 标的物行的位置

在 **Appraisal**（估价）模式下，第 1 行必须是标的物。其余各行均为可比销售实例。标的物行不参与模型拟合。拟合完成后，模型仍会为标的物行生成预测值。

在 **Market Area Analysis**（市场区域分析）模式下，是否把标的物置于第 1 行可自行决定。

在 **General**（通用）模式下没有特殊的行处理——所有行一视同仁。

3 通用模式

3.1 概述

启动 mgcvUI 时, 通用 (General Purpose) 模式是默认模式。它为任何数据集——而不仅仅是房地产数据——提供完整的 GAM 工作流程。你可以把 mgcvUI 用于科学数据、金融分析、工程研究, 或任何预期存在光滑非线性关系的回归问题。

在 General 模式下, 界面略去房地产专用功能 (特殊列、销售时距、RCA)。侧边栏被精简, 专注于变量选择、参数配置、模型拟合与导出。

3.2 侧边栏工作流程

侧边栏按编号组织为若干可折叠小节, 引导你从项目选择一路走到报告导出。多数小节只有在打开一个活动项目后才出现, 其中若干还与项目的用途相关。输出文件夹由活动项目推导得出——并没有单独的输出文件夹字段。

1. Project (项目) ——创建或打开一个项目。项目设定用途与工作地点 (国家、州、县市), 以及它的输入 (`<os>_in/`) 与输出文件夹。项目存放在共享的 `regProj` 根目录下 (默认为 `~/regProj`, 可通过 `REGPROJ_ROOT` 环境变量或 Settings 下拉菜单覆盖), 并与同族应用 `earthUI` 和 `glmnetUI` 共享, 因此三者的模型可以并列存放。点击 **Close Project** (关闭项目) 即可切换。

2. Import Data (导入数据) ——从活动项目的输入文件夹中选择一个 CSV 或 Excel 文件。先通过 Finder/资源管理器把文件放入该文件夹, 然后点击 **Refresh** (刷新) 以列出它们。对于含多个工作表的 Excel 文件, 会出现一个工作表选择器。列名会自动转换为 `snake_case`。

3. Import from earthUI (optional) (从 earthUI 导入, 可选) ——导入一个 `earthUI` 结果 `.rds` 文件, 用 `earth` 发现的节点位置为 GAM 提供初值 (同时也为变量选择提供初值)。此步骤是可选的——若你没有可导入的 `earthUI` 结果, 跳过即可。

4. Variable Configuration (变量配置) ——目标变量选择器、响应变量变换 (无、`log`、`log10`), 以及带 Include (纳入)、Factor (因子) 和 Linear (线性) 复选框的预测变量表。Special (特殊) 列只在估价模式与市场模式下出现。详见第 6 章。

5. Mgcvcall Parameters (mgcv 调用参数) ——全部模型配置: 参数预设、`family`、`method`、`gamma`、交叉验证、`select`、基类型、`k`、张量类型、交互矩阵以及高级参数。完整的参数参考见第 7 章。

6. Fit Mgcvcall GAM Model (拟合 mgcv GAM 模型) ——运行模型的按钮。结果会显示在右侧的各个标签页中。

7. Download Output (下载输出) ——把预测值、残差、CQA 得分以及各变量的贡献导出为 Excel 文件, 存入项目的输出文件夹。模型拟合之后, 在所有用途模式下均可使用。

8. Calculate RCA Adjustments & Download (计算 RCA 调整额并下载) ——为每个可比实例计算逐变量的 RCA 调整额, 并通过 CQA 对标的物残差作插值。仅在估价模式与市场模式下、且已完成拟合后出现。

9. Generate Sales Grid & Download (生成销售网格并下载) ——生成销售比较网格 (Sales Comparison Grid), 按总调整百分比推荐可比实例。仅在估价模式下出现。

10. Generate Quarto Report (生成 Quarto 报告) ——把一个自包含的 Quarto (`.qmd`) 报告包——含全部图形与表格——写入项目的输出文件夹。模型拟合之后即可使用。

11. Convert Quarto Report (转换 Quarto 报告) ——把生成的 `.qmd` 包渲染为 HTML、Word 或 PDF。PDF 需要安装 LaTeX; 若未检测到 LaTeX, PDF 选项会被隐藏。参见第 13 章。

在其他模式下, 与用途相关的步骤会被略去 (其余步骤重新编号): 市场模式下没有销售网格步骤, 通用模式下 RCA 与销售网格两个步骤都没有。

3.3 从 earthUI 导入 (可选)

侧边栏的第 3 节——**Import from earthUI**——允许你导入一个 earthUI 结果 .rds 文件。这就启用了 earth-mgcv 流水线：earth 由数据驱动得出的节点位置成为 GAM 光滑项的锚点。

导入 earthUI 结果后，mgcvUI 会：

- 显示一份摘要，列出目标变量、R 平方、项数以及每个变量的节点位置。
- 自动选中“Earth Pipeline” (earth 流水线) 参数预设 (三次回归样条基, $\gamma = 1.4$)。
- 把目标变量设为与 earth 的目标变量一致。
- 为 earth 的预测变量预先勾选 Include。
- 为 earth 的分类变量预先勾选 Factor，为 earth 的线性预测变量预先勾选 Linear。
- 预先填入 earth 检测到的交互项 (在交互矩阵中以黄色高亮显示，并锁定不允许手工更改)。
- 提供一个 **Export Knots CSV** (导出节点 CSV) 按钮，可把 earth 的节点下载为 CSV 文件。

随后，变量配置表会显示一条说明，指出变量选择 (Include、Linear、Factor) 与样条节点是由 earth 模型提供初值的。与 glmnetUI 不同——在 glmnetUI 中 glmnet 的拟合被绑定到 earth 的基上——mgcvUI 是依据你当前的选择来拟合 GAM 的，earth 只是提供了一个起步优势，而每一个控件都仍可编辑。

一个 **Clear** (清除) 按钮可移除已导入的 earth 数据，并重置为独立运行模式。

3.4 主面板标签页

数据导入之后，主面板提供以下标签页 (依赖模型的标签页在拟合完成后才有内容)：

标签页	内容
Data (数据)	已导入数据的交互式 DataTable。在估价模式与市场模式下，分为标的物业表与可比销售表。
Equation (方程)	用 MathJax 渲染的完整 GAM 公式，以及一张光滑函数定义表。
Summary (摘要)	关键指标 (R^2 、 $CV R^2$ 、已解释离差、AIC、n、光滑项个数) 以及光滑项 / 参数项表格。
Variable Importance (变量重要性)	F 统计量 (光滑项) 与 $ t $ 值 (参数项) 的水平条形图，以及排序表。
Contribution (贡献)	交互式 plotly 光滑偏效应曲线，带置信带与 earth 节点标记。
Correlation (相关性)	数值型预测变量相关系数的热力图 (拟合前即可查看)。
Diagnostics (诊断)	残差对拟合值、Q-Q 图、直方图、响应值对拟合值 (通过 gratia 生成)、实际值对预测值散点图。
RCA Adjustments (RCA 调整)	RCA 分析直方图 (估价 / 市场模式，计算完成后)。
Anova (方差分析)	参数项与光滑项合并的方差分析表。
Mgcv Output (mgcv 输出)	原始文本：计时、公式、模型摘要、family、共曲性。
Sign Check (符号检查)	earth 与 GAM 的方向一致性检查 (当已导入 earth 结果时)。
Concurvity (共曲性)	整体与两两共曲性矩阵，带颜色编码。
Basis Check (基检验)	<code>gam.check()</code> 的基维数充分性检验。

3.5 设置的持久化

mgcvUI 会自动把你的配置保存到一个 SQLite 数据库中，并以输入文件名为键。当你重新加载同一文件时，所有设置都会恢复：目标变量选择、预测变量复选框、数据类型、mgcv 参数以及交互矩阵。用途是活动项目的属性而非全局设置，最近使用的输入文件按项目分别记住，并在项目重新打开时自动加载。

3.6 暗色模式

点击右上角的月亮 / 太阳图标，可在 Nord Light 与 Nord Dark 两种主题之间切换。主题偏好保存在 localStorage 中，跨会话保持。所有界面元素都会随所选主题相应变化。

3.7 区域与地区设置

mgcvUI 通过一套基于国家的区域设置 (locale) 系统支持国际化的数字与 CSV 格式约定。标题栏中的 **Settings** (设置) 下拉菜单为 30 多个受支持国家提供 **Country** (国家) 与 **Paper** (纸张) 选择器。每种预设配置以下各项：

- **CSV 分隔符** —— 美国 / 英国 / 日本用逗号 (,)，欧洲大部分地区用分号 (;)。
- **小数点符号** —— 句点 (.) 或逗号 (,)。
- **千位分隔符** —— 逗号 (美国 / 英国 / 日本)、句点 (德国 / 意大利 / 西班牙)、空格 (芬兰 / 法国 / 波兰 / 波罗的海国家)，或撇号 (瑞士)。
- **纸张尺寸** —— Letter (美国 / 加拿大 / 墨西哥) 或 A4 (其他所有地区)。

点击 **Save** (保存) 即可把你的区域偏好全局存储。Import Data 小节还提供一个按文件生效的 **Locale** (区域) 选择器, 供采用不同约定的一次性导入使用。同一个 Settings 下拉菜单中还有 **regProj root folder** (regProj 根文件夹) 字段, 用于重新指定 mgcvUI、earthUI 与 glmnetUI 共用的项目树位置。

提示: 图形坐标轴与斜率标签上的数字格式会自动适应你的区域设置。德语区域用句点作千位分隔 (200.000), 芬兰用空格 (200 000), 瑞士用撇号 (200'000)。界面不显示任何货币符号——*mgcvUI* 与货币无关。

4 估价模式

当活动项目的用途为 **For Appraisal**（用于估价）时，mgcvUI 会把自身配置为面向单一物业估价。第 3 章所述的全部功能依然可用；本章只讲估价模式特有的附加内容。

4.1 标的物行的处理

在估价模式下，数据集的第 1 行即标的物，其余各行为可比销售实例。你的输入文件必须据此组织（参见第 2 章）。标的物的销售价格可以留空，也可以填任意值——mgcvUI 在拟合时会自动把它当作 NA 处理。

拟合完成后，模型会为标的物行生成预测值，这是 RCA 调整工作流程的基础。

4.2 估价基准日与销售时距

在估价模式与市场模式下，变量配置小节中会出现一个 **Effective Date**（估价基准日）字段（默认为当天日期）。若你在 Special 下拉菜单中把某列指定为 `contract_date`，mgcvUI 就会计算一个 `sale_age` 列——即每笔交易的成交日期与基准日之间的整天数。该列会作为预测变量加入模型。

当估价基准日改变时，`sale_age` 会自动重新计算。

4.3 特殊列指定

在估价模式与市场模式下，变量配置表中的每个预测变量都会出现一个 **Special**（特殊）下拉菜单。特殊类型及其作用的完整参考见第 6 章。

4.4 RCA 调整概览

Calculate RCA Adjustments & Download（计算 RCA 调整额并下载）按钮（在估价模式与市场模式下、拟合完成后可见）会为每个可比实例相对于标的物计算由市场推导出的调整额。完整的 RCA 工作流程见第 11 章。

提示：你为标的物指定的 *CQA* 得分决定了残差分布中有多大比例被归属于标的物。得分 5.00 会把标的物置于各可比实例的中位数位置。

5 市场区域分析模式

当活动项目的用途为 **Market Area Analysis**（市场区域分析）时，mgcvUI 提供与估价模式相同的房地产专用功能（特殊列、销售时距），但其取向是分析一组物业，而非评估单一标的物的价值。

5.1 与估价模式的差异

- 标的物行的处理更为灵活 —— 在市场模式下，第 1 行不会被自动当作标的物。
- 没有销售网格 —— “Generate Sales Grid & Download” 步骤不可用。

5.2 何时使用市场模式

在以下情形下适合使用市场区域分析模式：

- 为某个街区或市场区域建立 GAM 模型，以理解非线性的价值驱动因素
- 分析建筑面积、年龄、地块面积与区位等变量如何影响一组物业的销售价格
- 为市场状况分析或街区界定准备支撑材料
- 处理这样一份数据集：你想要特殊列功能（销售时距），但不需要销售网格工作流程

提示：市场模式对一般性的房地产回归也很有用——当你想要特殊列功能，但不需要单一物业估价工作流程时。

6 变量选择

侧边栏的第 4 节——**Variable Configuration**（变量配置）——是你选定哪些列参与模型、以及它们如何被处理的地方。

6.1 目标变量

第 4 节顶部的 **Target (response) variable**（目标 / 响应变量）下拉菜单列出数据集中的每一个数值列。选择其中一列作为响应变量（例如销售价格）。目标列会自动从预测变量列表中排除。

6.2 响应变量变换

在目标变量选择器下方，**Response Transform**（响应变量变换）下拉菜单提供三个选项：

- **None (raw)**（无，原值）——直接使用响应变量。当预测变量与响应变量在原始尺度上近似可加时，此选项合适。
- **Log (natural)**（自然对数）——在拟合前施加自然对数变换。反变换之后，这会使调整额成为比例式（%）而非绝对额（\$）。这对销售价格尤其有用，因为一笔 \$10,000 的调整在不同价格水平上含义大不相同。
- **Log10**（以 10 为底的对数）——施加以 10 为底的对数变换。

提示：对数变换会把时间调整变为比例式（%）而非绝对额（\$）。对于价格跨度很大的房地产模型，对数变换往往能产生性态更好的残差。

选择对数变换后，响应变量中 ≤ 0 的取值会被自动滤除。所有输出（预测值、贡献、残差）都会自动反变换回原始尺度。

6.3 预测变量表

在目标变量选择器下方，一张可滚动的表格列出其余所有列。复选框列使用旋转的竖排表头以求紧凑。列的顺序为：

列	说明
Variable	列名（等宽字体，过长时以省略号截断）。带边框单元格。
Type	数据类型下拉菜单：numeric、integer、character、factor、logical、Date、POSIXct。带边框下拉菜单。
Include	复选框——把该列作为预测变量纳入模型。竖排表头。
Factor	复选框——按分类因子处理（为每个水平生成一个单独的系数）。竖排表头。
Linear	复选框——强制使用线性关系（不用光滑项，只用 βx ）。竖排表头。
Special	下拉菜单（仅估价 / 市场模式）——参见下文的特殊列类型参考。带边框下拉菜单。
NAs	缺失值的个数与百分比，带颜色编码：红色（ $\geq 50\%$ ）、橙色（ $\geq 20\%$ ）、灰色（ > 0 ）。带边框单元格。

因子、光滑与线性之别：默认情况下，被纳入的数值变量会得到一个光滑项 $f(x)$ 。勾选 **Factor** 会生成一个分类项（每个水平一个系数）。勾选 **Linear** 会生成一个简单线性项 βx 以取代光滑项。同时被标记为 Factor 与 Linear 的变量按 Factor 处理。

6.4 数据类型检测与覆盖

mgcvUI 在导入时会自动检测数据类型，可识别数值、整数、逻辑、因子与日期列。你可以通过更改 **Type** 下拉菜单来覆盖任何一项检测结果。更改类型会影响该列在模型中的处理方式。

6.5 特殊列类型参考

在估价模式与市场模式下，**Special** 下拉菜单提供以下选项：

加权：

- **weight** ——观测权重列（只允许一列；权重为 0 的行不参与拟合）

日期与时间类型：

- **contract_date** ——触发根据估价基准日自动计算 **sale_age**
- **sale_age** ——标识一个已预先计算好的销售时距列（距基准日的天数），按原样使用
- **listing_date** ——在计算市场停留天数时作为备选来源
- **dom** ——标识市场停留天数列

记录类型：

- **sale_type** ——标识用于区分标的物与可比销售实例的记录类型列

货币类型：

- **concessions** ——标识销售让价（卖方补贴、买方优惠等）

面积与区位类型：

- **latitude** ——取值自动四舍五入至 3 位小数
- **longitude** ——与纬度同样的舍入处理
- **area** ——市场区域或街区标识符
- **living_area** ——启用按平方英尺计的残差计算（**residual_sf** 与 **cqa_sf**）
- **lot_size** ——宗地面积列
- **site_dimensions** ——与地块面积归为一组

年龄类型：

- **actual_age** ——物业年龄列
- **effective_age** ——物业的有效年龄

显示类型：

- **display_only** ——该列会包含在 Excel 导出中，但完全不参与模型拟合。可用于地址字段、MLS 编号、宗地编号或其他参考性数据。

提示：被指定为 *display_only* 的列仍保留在数据集中并出现在 Excel 导出里，但会完全从预测变量表中排除。可用于 ID 列、地址或其他参考字段。

7 参数选择

侧边栏的第 5 节——**Mgcv Call Parameters**（mgcv 调用参数）——提供 GAM 的全部配置选项。每个参数都带有一个蓝色帮助图标（?），其提示框给出说明。

7.1 参数预设

顶部的一个下拉菜单提供两种预设，为常见工作流程配置合理的默认值：

预设	基	设置
Standalone (discovery)（独立运行，探索）	tp（薄板）	k = 自动（10），gamma = 1.2，select = TRUE
Earth Pipeline (refinement)（earth 流水线，精化）	cr（三次回归）	k = 自动（10），gamma = 1.4，select = TRUE

导入 earthUI 节点时会自动选中 **Earth Pipeline** 预设。earth 节点的整合必须使用三次回归样条（cr）基，因为只有它允许指定确切的节点位置。

7.2 Family（分布族）

为你的响应变量选择分布族：

Family	适用情形
gaussian	连续型响应变量（例如销售价格）。最为常用。
Gamma	方差与均值成比例的正值连续型响应变量。
poisson	计数数据（例如成交笔数）。
binomial	二元结果。
inverse.gaussian	严重右偏的正值数据。

7.3 Method（方法）

光滑参数的估计方法：

Method	说明
REML	受限极大似然（默认，推荐）。对过拟合最为稳健。
GCV.Cp	广义交叉验证。倾向于选出略为复杂的模型。
ML	极大似然。与 REML 相似，但可能欠光滑。
P-REML	Pearson REML 变体。
P-ML	Pearson ML 变体。
GACV.Cp	广义近似交叉验证。

7.4 Gamma（光滑度乘子）

gamma 参数会对光滑准则中的有效自由度作乘法放大，从而促成更光滑（摆动更少）的拟合。默认值为 1.2。

取值	效果
1.0	标准光滑度
1.2–1.4	略比默认更光滑（推荐）
2.0+	光滑得多——适用于噪声较大的数据或小样本

较大的 `gamma` 值通过更重地惩罚复杂度来防范过拟合。Earth Pipeline 预设采用 1.4，以便在精化 `earth` 的节点时施加额外的光滑。

7.5 交叉验证

勾选时（默认勾选），`mgcvUI` 会在拟合后运行 10 折交叉验证（cross-validation）以计算 CV R 平方。这给出了样本外预测能力的一个诚实估计。

7.6 Select（收缩）

勾选时（默认勾选），会为每个光滑项附加一个额外惩罚，可将其收缩至零。这就实现了自动变量选择——不重要的光滑项实际上被移出模型。其实现方式是在 `gam()` 中设定 `select = TRUE`。

7.7 默认基

光滑项所用的样条基函数（basis function）类型：

基	全称	说明
tp	薄板回归样条	默认。各向同性（无需安放节点）。是 很好的通用选择。
cr	三次回归样条	允许显式安放节点。整合 <code>earth</code> 节点 时必须使用。
ps	P 样条	节点等距的 B 样条基，配以差分惩 罚。计算效率高。
bs	B 样条	灵活的 B 样条基。

7.8 默认 k（基维数）

基维数 `k` 控制每个光滑项摆动程度的上限。取值为 0（默认）表示“自动”——`mgcvUI` 会视情形取 `k = 10` 或 `earth` 节点的个数。

- 较小的 `k`（3–5）：曲线非常光滑，接近直线。
- 中等的 `k`（8–15）：灵活度适中（通常已足够）。
- 较大的 `k`（20 以上）：非常灵活——若无强正则化则有过拟合之虞。

提示： *Summary* 标签页中的有效自由度（EDF）告诉你每个光滑项实际用掉了多少自由度。若 *EDF* 接近 `k - 1`，说明基可能太小——应加大 `k`。若 *EDF* 远小于 `k`，说明该光滑项受到了良好的惩罚。

7.9 张量类型

对于连续变量之间的交互，有两种张量积类型可用：

类型	函数	说明
ti	ti(x1, x2)	张量交互——只对交互效应建模，主效应由单变量 s() 项分别处理。就可解释性而言更受推荐。
te	te(x1, x2)	张量积——对包含主效应在内的整个联合效应建模。在 RCA 调整中较难分解。

7.10 允许的交互

一个可折叠的 **Allowed Interactions**（允许的交互）小节会为所有被纳入的光滑（非线性、非因子）预测变量显示一个上三角复选框矩阵。每个复选框启用相应变量对之间的张量积光滑。

- **Allow All**（全部允许） / **Clear All**（全部清除）——批量切换按钮。
- 点击变量名 ——切换该变量的全部交互。
- 右键点击变量名 ——“Block from main effect”（从主效应中屏蔽）——把该变量标记为仅用于交互（不生成单变量光滑项）。

导入 earthUI 节点时，earth 检测到的交互会被预先勾选并锁定（以黄色背景高亮）。若 earth 模型使用了 degree = 1（无交互），则会显示一条提示信息。

警告

启用张量交互会大幅提高模型复杂度。请在你拥有领域知识、或有 earth 检测到的证据支持时才使用交互项。务必核实所得调整额是合理且可以解释的。

7.11 因子-光滑交互

当因子变量与光滑变量同时被纳入时，会出现一个单独的 **Factor-by-Smooth Interactions**（因子-光滑交互）矩阵。每个复选框会生成一个 s(x, by = factor) 项——即为每个因子水平生成一条单独的光滑曲线。

7.12 高级参数

一个可折叠的“Advanced”（高级）小节展示若干附加设置：

- **Optimizer**（优化器）——光滑参数估计所用的算法：outer/newton（默认）、outer/bfgs 或 efs（扩展 Fellner-Schall 法）。
- **Scale**（尺度参数）——尺度参数。0（默认）表示由数据估计。对于尺度已知的模型可设为一个确定值。
- **Discrete covariate method**（离散化协变量方法）——针对大型数据集的快速离散化拟合（discrete = TRUE）。
- **Threads**（线程数）——并行线程的数目（1-32）。仅在 discrete = TRUE 时有效。

8 拟合模型

8.1 Fit 按钮

侧边栏第 6 节中有 **Fit Mgcgv GAM Model** (拟合 mgcv GAM 模型) 按钮。点击它即以你当前的配置运行模型。

8.2 拟合过程中发生了什么

当你点击 Fit 时, mgcvUI 会:

1. 校验配置 —— 检查是否至少有一个预测变量、以及至少 10 个完整行。若完整行不足 50%, 会给出警告 (并列出造成大部分 NA 的那些列)。
2. 准备数据 —— 只保留与模型相关的列, 删除含 NA 的行, 应用权重 (若已指定), 排除标的物行 (在估价模式下), 并施加响应变量变换。
3. 构建 GAM 公式 —— 为数值预测变量构造光滑项 $s(x, bs="cr", k=10)$, 为分类变量构造因子项, 为被标记为 Linear 的变量构造线性项, 并为所选交互构造张量积。earth 的节点会作为显式的节点位置嵌入到 cr 基的项中。
4. 处理边界情形 —— 剔除常数变量, 把二值预测变量转换为因子, 把 k 限制在不同取值个数减 1 以内, 并在需要更多节点时用数据分位数补充 earth 的节点。
5. 拟合 GAM —— 以构造好的公式、family、method、gamma、select 及其他参数调用 `mgcv::gam()`。
6. 运行交叉验证 (若已启用) —— 10 折交叉验证, 计算样本外 R 平方。
7. 更新所有结果标签页 —— Summary、Equation、Contribution 图、Diagnostics 以及其他所有标签页都会被填充。

8.3 后台拟合与进度显示

当 callr 软件包可用时, 模型拟合会在一个后台进程中运行, 以保持应用程序的响应性。一个暗色主题的终端式进度浮层会显示:

- 计时器, 显示已耗费的秒数 / 分钟数。
- 按颜色编码的跟踪日志: 蓝色表示状态, 黄色表示交叉验证折的进度, 红色表示错误, 绿色表示结果。
- 一个 **Abort** (中止) 按钮, 可在需要时终止拟合进程。
- 一个 **Close** (关闭) 按钮, 在拟合完成后出现。

若 callr 不可用, 拟合将同步运行 (应用程序会冻结, 直到拟合完成)。

拟合成功后, Fit 按钮上会出现一个白色对勾, 同时状态行显示: “R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X.”

9 结果标签页

9.1 Data (数据)

以交互式 DataTable 显示已导入的数据，支持横向滚动，每页 15 行。在估价模式与市场模式下，该标签页分为一张 **Subject Property**（标的物业，第 1 行）表和一张 **Comparable Sales**（可比销售实例，其余各行）表；在 General 模式下只显示一张表。

每个单元格都保持单行，并按列宽截断，这样很宽的自由文本字段（例如物业备注）就不会把整行拉得很长。单击可选中一行；双击任一单元格可弹出窗口，显示其完整、未经截断的内容。

9.2 Equation (方程)

分两部分显示已拟合的 GAM 模型：

模型方程（由 MathJax 渲染）：

- 左端显示响应变量及其相应的变换（log、log10）与联系函数。
- 右端显示截距值、以 f_i (变量) 形式给出的光滑项、按变量分组并标明水平数的因子项，以及带系数值的线性项。
- family 与 method 显示在方程上方。

Smooth Function Definitions（光滑函数定义）表：

- 项的公式字符串（例如 `s(living_sqft, bs="cr", k=8)`）
- 变量名
- 类型（s、te、ti 或 linear）
- 基函数类型
- k 值
- earth 节点位置（若已导入）

9.3 Summary (摘要)

顶部横向排列六张指标卡片：

- **R-squared** ——已解释离差的比例
- **CV R-squared** ——交叉验证 R 平方（若已启用交叉验证）
- **Dev. Explained %** ——已解释离差的百分比
- **AIC** ——赤池信息准则
- **n** ——所用观测的个数
- **Smooths** ——光滑项的个数

卡片下方：

- **Smooth Terms** 表（光滑项）——项、EDF（有效自由度）、Ref.df、F 统计量、p 值
- **Parametric Terms** 表（参数项）——项、估计值、标准误、t 值、p 值

9.4 Variable Importance (变量重要性)

一张水平条形图，显示各模型项的相对重要性：

- 光滑项：按 F 统计量排序（蓝色条）
- 参数项：按 $|t|$ 值排序（绿色条）

图下方的 DataTable 列出项、类型、EDF、统计量与 p 值，按统计量降序排列。

9.5 贡献图

每一个对预测值有贡献的模型项都有一张交互式 **plotly** 贡献图，显示在一张带边框的卡片中。该标签页会为所有类型的项给出图形：

9.5.1 单变量光滑项 — $s(x)$

- 蓝色曲线，带 95% 置信带
- 灰色散点，显示偏残差（用于未作对数变换的模型）
- 红色竖直虚线，位于 earth 节点处（若已导入 earth 结果）
- 悬停提示：变量取值、贡献（\$）、95% 置信区间、变化率（每单位的斜率）

对于经对数变换的模型，纵轴会用公式 $\bar{y} \times (e^{f(x)} - 1)$ 反变换为以美元计的贡献，使曲线可以直接按美元数额来解读。

对于纬度/经度变量，斜率按每 0.001 度而非每 1 度给出，因为前者是更有意义的地理增量。

9.5.2 因子-光滑项 — $s(x, \text{by} = \text{factor})$

每个因子水平一条彩色曲线，叠绘在同一张图上。这有助于看清某个预测变量的效应如何随组别（例如不同街区）而变化。

9.5.3 交互项 — $ti(x_1, x_2)$ 与 $te(x_1, x_2)$

用热力图显示二维贡献曲面。色标为：红色（负）→ 白色（零）→ 蓝色（正）。悬停时同时显示两个变量的取值与贡献数额。

9.5.4 参数项 — 线性项与因子项

- 数值型线性项：数据点的散点图，叠加拟合直线
- 因子项：条形图，显示每个因子水平的平均贡献

9.6 Correlation（相关性）

一张热力图矩阵，显示所有数值型预测变量之间的 Pearson 相关系数。数据导入后即可查看——无需拟合模型。

- 颜色渐变：蓝色（-1）→ 白色（0）→ 红色（+1）
- 每个单元格中标出相关系数值
- 有助于在拟合之前识别共曲性（concurvity，多重共线性在光滑项中的对应现象）

9.7 Diagnostics（诊断）

两组诊断可视化：

诊断面板（通过 `gratia::appraise` 生成）：四张图——残差对拟合值、Q-Q 图、残差直方图，以及响应值对拟合值。

实际值对预测值散点图：观测值对预测值，附一条 45 度虚线参考线。对于经对数变换的模型，取值会被反变换回原始尺度。

9.8 RCA Adjustments（RCA 调整）

RCA 计算完成后显示三张直方图（仅限估价 / 市场模式）：

- **Residual Adjustment %**（残差调整百分比）——残差调整额占销售价格百分比的分布
- **Net Adjustment %**（净调整百分比）——净调整额的分布
- **Gross Adjustment %**（总调整百分比）——总调整额的分布

每张直方图的副标题中给出均值、中位数与标准差，并以虚线标出均值与中位数的参考位置。

9.9 Anova (方差分析)

一张合并的方差分析表，把 GAM 摘要中的参数项与光滑项汇合在一起。显示项、类型（参数项或光滑项）以及所有相关统计量。

9.10 Mgcvc Output (mgcv 输出)

原始文本输出，内容包括：

- 拟合计时
- GAM 公式
- 模型的 print 输出
- 完整的 `summary(model)` 输出
- family 与联系函数
- 整体共曲性度量

9.11 Sign Check (符号检查)

当已导入 earthUI 节点时，该标签页会把每个变量在 earth 模型中的方向与其在 GAM 中的方向作比较：

- **earth_direction**: 递增、递减或混合（依据铰链函数的符号判定）
- **gam_direction**: 递增、递减或混合（由光滑项在 100 点网格上的导数估计得出）
- **consistent**: TRUE/FALSE
- **warning**: 对任何不一致之处的说明

状态行会显示“All directions consistent”（所有方向一致）或存在不一致的变量个数。

9.12 Concurvity (共曲性)

共曲性是多重共线性在光滑项中的对应现象。该标签页显示：

- 整体共曲性表：各行分别对应 worst / observed / estimate 三种度量，各列对应每个光滑项。带颜色编码：白色 (< 0.5)、黄色 ($0.5-0.8$)、红色 (> 0.8)。
- 两两共曲性表：最坏情形的两两矩阵，采用相同的颜色编码。

“worst”一行中高于约 0.8 的取值表明，该光滑项可以被其他光滑项很好地逼近，提示存在冗余。

9.13 Basis Check (基检验)

`mgcv::gam.check()` 的输出：针对每个光滑项的基维数充分性检验。若某个光滑项的 k 过小，该检验会以一个较小的 p 值将其标示出来，提示你应当加大 k 。

10 下载数据

拟合完成后，**Download Output (Excel)**（下载输出，Excel）按钮（侧边栏第 7 节）会把含预测值与诊断信息的 Excel 文件导出到活动项目的 mgcv 输出文件夹。

10.1 输出列

列	说明
basis	反变换后的截距值（所有行相同）
<var>_contribution	各光滑项对预测值的贡献。对于对数模型，会按比例分配为美元数额。
est_<target>	模型预测值（例如 est_sale_price），若适用则已从对数尺度反变换
residual	实际值 - 预测值
calc_residual	校验用：实际值 - 各项贡献之和（反变换后）
cqa	状况-品质-吸引力（Condition-Quality-Appeal, CQA）得分（0-10 分制）
residual_sf	残差 / 居住面积（若已指定 living_area）
cqa_sf	由 residual_sf 排序计算得到的 CQA

在估价模式下，第 1 行（标的物）的 residual/cqa/cqa_sf 均置为 NA。当指定了 living_area 列时，各行按 residual_sf 降序排列。
成功完成后，下载按钮上会出现一个白色对勾。

10.2 CQA 得分

CQA 把每一行的残差与其余所有行相比较，按 0-10 分制排序打分：

- 高 CQA（约 9-10）：成交价远高于预测值——很可能状况 / 品质 / 吸引力较优
- 低 CQA（约 0-1）：成交价远低于预测值——很可能状况较差或属于困境出售
- CQA 约为 5：接近中位数

11 RCA 计算与下载

RCA (Residual Constraint Approach, 残差约束法) 工作流程在 **Appraisal** (估价) 与 **Market Area Analysis** (市场区域分析) 模式下、模型拟合之后可用。标的物业必须位于第 1 行。

11.1 打开 RCA 对话框

点击侧边栏第 8 节中的 **Calculate RCA Adjustments & Download** 按钮，会弹出一个模态对话框，其中包含：

- **Score type** (得分类型)：CQA 或 CQA per SF (若已指定 living_area)。按平方英尺计的选项会用居住面积对标的物的残差作缩放。
- **Subject CQA Score** (标的物 CQA 得分)：接受 0.00 至 10.00 之间的取值。该字段初始为空；若已导入过 earthUI 模型，你在 earthUI 中使用的得分类型与取值会从共享的项目设置中沿用过来。

点击 **Generate** (生成) 即可计算并下载。Excel 文件会写入活动项目的 mgcv 输出文件夹 (<os>_out_mgcv)。

11.2 CQA 得分插值

1. 先对各可比实例的 CQA 得分与残差排序
2. 用线性插值 (approx()) 把你给定的 CQA 值映射为一个残差
3. 标的物价值 = 模型预测值 + 插值所得残差
4. 权重为 0 的行按与标的物具有相同残差来处理

11.3 输出列

列	说明
subject_value	模型预测值 + 插值所得残差
subject_cqa	用户输入的 CQA 得分
<var>_adjustment	标的物贡献 - 可比实例贡献 (以美元计)
residual_adjustment	标的物残差 - 可比实例残差
net_adjustments	全部调整额之和
gross_adjustments	全部调整额绝对值之和
residual_adj_pct	残差调整额占销售价格的百分比
net_adj_pct	净调整额占销售价格的百分比
gross_adj_pct	总调整额占销售价格的百分比
adjusted_sale_price	实际销售价格 + 净调整额

计算成功完成后，按钮上会出现一个白色对勾。

12 销售比较网格

销售比较网格 (Sales Comparison Grid) 仅在 **Appraisal** (估价) 模式下、且已计算 RCA 调整额之后可用。它会生成一份格式化的 Excel 工作簿, 适合收入估价报告之中。

12.1 可比实例选择对话框

点击 **Generate Sales Grid & Download** 按钮会打开一个模态对话框, 列出所有可比销售实例。可比实例分为两组:

- **Recommended comps** (推荐的可比实例, 已预先勾选): 总调整额 < 销售价格的 25%, 按销售时距排序 (最近的在前)。最多 30 个。
- **Other comps** (其他可比实例, 默认不勾选): 其余的可比实例, 按总调整百分比排序。

选定你的可比实例后, 点击 **Generate Sales Grid** 即可创建工作簿。

必需的特殊类型: 必须已指定 `contract_date` 与 `living_area`。推荐的特殊类型: `latitude`、`longitude` 与 `lot_size`——没有它们网格也能生成, 但某些字段会留空。

12.2 工作簿版式

生成的工作簿含多个工作表, 每张表并排容纳 3 个可比实例:

- 表头行: 标的物与可比实例的地址、销售价格、成交日期以及市场停留天数
- 分组行: 区位 (纬度、经度、区域)、宗地面积 (地块面积、宗地尺寸) 与年龄 (实际年龄、有效年龄), 带组标题与子条目
- 模型变量行: 每个光滑项一行, 显示标的物贡献、各可比实例贡献以及调整额
- 残差特征行: 空白且未锁定的单元格, 供估价师填写模型未能刻画的特征 (例如状况、品质、景观)
- 汇总行: 净调整百分比、总调整百分比与调整后销售价格, 均带有可实时计算的 Excel 公式

12.3 工作表保护与残差单元格

每张工作表都受到保护以防误改, 但残差特征的取值单元格被显式解锁。这既让估价师能够为模型之外的特征输入取值, 又保留了由公式驱动的调整额计算。

13 下载报告

报告生成采用两步式 Quarto 工作流程，取代了早期版本中的一步式报告。

13.1 第 1 步：生成 Quarto 报告

Generate Quarto Report (生成 Quarto 报告) 按钮 (模型拟合后可用) 会把一个自包含的 .qmd 包写入活动项目的 mgcv 输出文件夹。该包含有 Quarto 源文件、预先渲染好的图形文件 (PNG 与 PDF)、一个装载了全部表格与模型元数据的 report_data.rds, 以及用于 Word 输出的 reference.docx 样式模板。由于该包是自包含的, 你可以在渲染之前编辑 .qmd 源文件, 或把它与其他 Quarto 文档组合起来。

图形在生成该包时即已预先渲染; 若某张图渲染失败, 程序会写入一张占位图, 包中其余部分照常生成。

13.2 第 2 步：转换 Quarto 报告

Convert Quarto Report (转换 Quarto 报告) 小节通过 quarto::quarto_render() 把一个 .qmd 文件渲染为一种或多种输出格式。它默认指向最近生成的那个包, 但 **Browse** (浏览) 按钮允许你选择任意 .qmd 文件。可用格式有三种:

- **HTML** —— 自包含的 HTML 文档。在所有平台上均可使用, 无需额外软件。推荐作为默认选项。
- **Word (.docx)** —— 使用随包提供的 reference.docx 样式模板渲染。适合编辑与分发。
- **PDF** —— 需要安装 LaTeX (参见下文的系统要求)。若未检测到 LaTeX, PDF 选项会自动从格式复选框中隐去。

早期的一步式渲染器 render_gam_report() 与 export_gam_docx() 仍保留, 可供程序化调用。

13.3 报告内容

生成的报告包括:

- **数据集与模型:** 数据文件、样本量、family、method 以及拟合信息
- **结果摘要:** R^2 、 $CV R^2$ 、已解释离差、AIC 及相关的拟合指标
- **模型方程:** 模型方程, 其中光滑项用 f_i 记号表示、线性项给出显式系数, 并附光滑函数定义。较长的方程在 PDF 中会跨多行折行。
- **摘要表:** 光滑项 (EDF、F、p 值) 与参数项 (估计值、标准误、t、p 值)
- **已启用的交互:** 模型中所含张量交互项与因子-光滑交互项的矩阵 (仅在存在交互项时显示)
- **变量重要性:** 按 F 统计量 (光滑项) 或 $|t|$ (参数项) 排序的条形图
- **光滑效应 (所有类型的项):**
 - 单变量光滑图 (交互式图形的 ggplot2 静态版本)
 - 因子-光滑的多线条图
 - 交互热力图 (t_i 与 t_e 项)
 - 参数项图 (数值型用散点图, 因子型用条形图)
- **预测变量相关性热力图**
- **诊断:** 残差诊断以及实际值对预测值的散点图
- **mgcv 摘要:** 已拟合模型的完整文本输出
- **记号附录:** 一个不编号的结尾小节, 说明报告中所用的价值贡献函数记号

成功生成后, Generate Quarto Report 按钮上会出现一个白色对勾。

14 导出模型函数

mgcvUI 可以把已拟合 GAM 的光滑函数导出为多种语言的独立代码。每个光滑项都在一个 200 点网格上求值，并以带线性插值的查找表形式导出，从而使该模型可在 R 之外使用，且不依赖 mgcv。

在当前版本中，函数导出是通过程序化方式进行的。导出用的辅助函数接受一个已拟合的结果对象（由 `fit_gam()` 产生），并写出下述文件：

- `generate_r_code()`、`generate_python_code()` 与 `generate_cpp_code()` ——各语言的源文件
- `export_functions_sqlite()` ——SQLite 查找表数据库
- `export_functions_zip()` ——把这四种格式的任意组合打包成一个 zip 压缩包

14.1 R 函数

文件：gam_functions.R

- 头部注释，含响应变量、 R^2 、CV R^2 与模型公式。
- 一个 INTERCEPT 常量。
- 每个光滑变量对应一个 `approxfun()`，带预先算好的 x/y 向量。
- 线性项系数以具名常量形式给出。
- 用法：把截距、各个 `g_var(value)` 以及各个 `coef_var * value` 相加。

14.2 Python 函数

文件：gam_functions.py

- numpy 导入语句。
- INTERCEPT 常量。
- 每个变量对应：_x_var 与 _fx_var 数组，以及使用 `np.interp()` 的 `g_var(x)` 函数。
- 线性项系数以 `COEF_*` 常量形式给出。

14.3 C++ 头文件

文件：gam_functions.hpp

- 仅含头文件，使用 `#pragma once`。
- `gam_functions` 命名空间。
- 使用 `std::lower_bound` 的 `interp()` 辅助函数。
- 每个变量对应：一个带 `static const` 数组的内联函数。
- 线性项系数以 `constexpr double` 形式给出。

14.4 SQLite 数据库

文件：gam_functions.sqlite

- 表 `smooth_grids`：列为 (variable, x, fx) ——即 200 点查找表。
- 表 `model_info`：response、r_squared、cv_rsqr、dev_explained、aic、n_obs、family、method、formula。
- 表 `linear_terms`：variable、coefficient。
- 表 `intercept`：截距值。

15 与 earthUI 和 glmnetUI 的比较

earth 应用程序及其后处理模块构成一套用于回归建模的工具集。它们共用相同的数据格式、特殊列类型、RCA 工作流程与演示数据集，但使用不同的建模引擎。

15.1 主要差异

特性	mgcvUI	earthUI	glmnetUI
方法	GAM (mgcv)	MARS/Earth (earth)	弹性网络 (glmnet)
关系形态	光滑非线性	分段线性	线性
变量选择	select=TRUE 收缩	逐步剪枝	Lasso 惩罚
交互项	张量积 (te/ti)	铰链函数乘积 (2-3 阶)	两两乘积
偏效应	带置信区间的光滑曲线	g 函数 (分段)	线性斜率
earth 流水线	从 earth 导入节点	节点的来源	导入铰链函数基
代码导出	R、Python、C++、SQLite	R、Python、C++、SQLite	—
报告格式	Word、PDF、HTML	Word、PDF、HTML	Word、PDF、HTML

有一项 earthUI 独有的差异值得一提：earthUI 含有一个实验性的 Prolog/DCG 功能，可从自由文本的 MLS 备注中提取结构化的物业属性。

15.2 共有功能

三个工具都提供：

- 共享的 regProj 项目树（默认为 ~/regProj）——项目、输入文件以及按方法分列的输出在同族应用之间共享
- CSV 与 Excel 导入，并作 snake_case 列名转换
- 支持 30 多个国家的区域设置，涵盖 CSV 分隔符、小数点符号与纸张尺寸
- 特殊列指定 (contract_date、living_area、area、latitude、longitude 等)
- 依据估价基准日自动计算 sale_age
- 暗色 / 亮色模式切换 (Nord 主题)
- 设置的持久化 (SQLite 与 / 或 localStorage)
- 相关性热力图、变量重要性、诊断
- CQA 打分与 RCA 调整工作流程
- 带可比实例选择与格式化 Excel 输出的销售比较网格
- 兼容同一批演示数据集 (Appraisal_1.csv)

15.3 推荐的工作流程

- 先用 earthUI 发现重要变量、非线性关系与交互作用。earth 的自动建模提供了最佳的起点。
- 再用 mgcvUI 精化，如果你希望得到更光滑的偏效应。导入 earthUI 的节点即可无缝过渡。
- 使用 glmnetUI，当你需要一个带激进变量选择或系数符号约束的纯线性模型时。
- 在同一份数据上比较三者，以评估非线性模型所增添的复杂度是否值得。

16 演示数据集: Appraisal_1.csv

16.1 说明

该演示数据集与 earthUI 和 glmnetUI 共用。它位于 mgcvUI 源代码的 demo_mls/ 文件夹中；若已安装 earthUI，则可用：

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
```

该文件包含约 1,500 条住宅销售记录，取自一份模拟的 MLS 导出，标的物业位于第 1 行，可比销售实例位于其余各行。数据代表一个多区域市场中的独户住宅销售，物业面积、年龄与区位各不相同。

这并非真实数据，但以北加利福尼亚一个实际存在的街区为蓝本。所有标识信息均已改写或删除。

16.2 各列

列	类型	特殊类型	说明
sale_type	character	sale_type	用于区分标的物与可比销售实例的记录类型码
weight	numeric	weight	观测权重 (0 = 不参与拟合)
id	numeric	display_only	内部记录 ID
property_id	numeric	display_only	MLS 物业标识符
listing_id	character	display_only	MLS 挂牌编号
parcel_number	character	display_only	县估价官宗地编号 (APN)
street_address	character	display_only	物业地址
city_name	character	display_only	市
postal_code	character	display_only	邮政编码
county_name	character	display_only	县
contract_date	Date	contract_date	销售合同日期 (用于计算 sale_age)
sale_age	numeric	—	从合同日期至估价基准日的天数 (已预先计算)
coe_date	Date	display_only	交割日期
listing_status	character	display_only	挂牌状态 (例如“Sold”，已售)
sale_price	numeric	(目标变量)	销售价格——响应变量
rent	numeric	—	月租金 (用于多目标模型)
list_price	numeric	display_only	挂牌价格
original_list_price	numeric	display_only	初始挂牌价格
living_sqft	numeric	living_area	以平方英尺计的总居住面积
ppsf	numeric	display_only	每平方英尺单价
beds_total	integer	—	卧室数
baths_total	numeric	—	卫生间总数 (例如 2.5 = 2 个全卫 + 1 个半卫)
lot_size	numeric	lot_size	以平方英尺计的地块面积
area_id	integer	area	MLS 区域标识符
area_text	character	display_only	区域名称
age	numeric	actual_age	以年计的物业年龄
year_built	integer	display_only	建成年份
latitude	numeric	latitude	纬度 (入模前四舍五入至 3 位小数)
longitude	numeric	longitude	经度 (入模前四舍五入至 3 位小数)
latitude6	numeric	display_only	全精度纬度
longitude6	numeric	display_only	全精度经度
garage_spaces	integer	—	车库泊位数
fp_count	integer	—	壁炉数
no_of_stories	numeric	—	楼层数
style	character	—	建筑风格
view	character	—	景观类型 (例如“Neighborhood”街区景、“Hills”山景)
days_on_market	integer	dom	市场停留天数
listing_date	Date	listing_date	挂牌日期

列	类型	特殊类型	说明
sale_concessions	numeric	concessions	卖方让价

16.3 建议的快速上手步骤

1. 启动 mgcv 例程: `earthUI::launch_mgcv()`
2. 在第 1 节中新建一个用途为 **For Appraisal** 的项目
3. 把 Appraisal_1.csv 放入该项目的输入文件夹, 并在第 2 节中选中它 (必要时点击 **Refresh**)
4. 选择 `sale_price` 作为目标变量
5. 如需比例式调整, 可将 Response Transform 设为 **Log (natural)**
6. 按上表所示指派特殊类型
7. 纳入以下预测变量: `sale_age`、`living_sqft`、`baths_total`、`lot_size`、`area_id` (作为因子)、`age`、`latitude`、`longitude`、`garage_spaces`
8. 保留默认设置: REML、 $\gamma = 1.2$ 、tp 基、k = 自动、select = TRUE
9. 点击 **Fit Mgcvm GAM Model**
10. 查看 Summary、Contribution (光滑图) 与 Diagnostics 标签页
11. 下载输出 (第 7 步), 查看 CQA 排名
12. 以约 ~5.00 的 CQA 得分计算 RCA 调整额 (第 8 步)
13. 用推荐的可比实例生成销售比较网格 (第 9 步)

17 系统要求与故障排除

17.1 受支持的平台

mgcvUI 可在 **macOS**、**Windows** 与 **Linux** 上运行。本应用程序在 macOS 配合 RStudio 的环境中开发，也主要在该环境中测试。与平台相关的说明见下文。

17.2 R 版本

要求 $R \geq 4.1.0$ 。强烈推荐使用 RStudio Desktop (2023.06 或更高版本) ——它自带 **Quarto** 与 **pandoc** (渲染报告所需)，并为启动本应用程序提供了便利的环境。

17.3 必需的 R 软件包

以下软件包会作为依赖自动安装：

软件包	用途
shiny	Web 应用程序框架
mgcv	GAM 模型拟合
valengrCore	共享的估价内核（特殊角色、earthUI 设置沿用）
ggplot2, plotly	静态与交互式贡献图
gratia	光滑项提取与诊断
DT	交互式数据表
quarto, rmarkdown	Quarto 报告的生成与渲染
readxl	Excel 文件导入
shinyFiles	文件与文件夹浏览器
RSQLite, DBI	设置与项目数据库 (SQLite)
jsonlite	变量状态的序列化

17.4 可选依赖

组件	何时需要
LaTeX (TinyTeX、MiKTeX 或 MacTeX)	仅用于生成 PDF 报告。若未检测到 LaTeX 安装，PDF 选项会自动从格式下拉菜单中隐去。安装命令：tinytex::install_tinytex()
sysfonts + showtext	供 ggplot2 图形使用的 Roboto Condensed 字体。若不可用或处于离线状态，应用程序会自动退回到系统无衬线字体。
earth ($\geq 5.3.0$)	仅在使用 earthUI 导入流水线时需要。独立使用时不需要。
callr	后台模型拟合（保持应用程序的响应性）。没有它则拟合同步运行。
writexl, openxlsx	Excel 下载（模型输出、RCA、销售网格）。缺失时应用程序会提示。
bslib	Bootstrap 5 主题 (Nord Light/Dark)。
officer	早期的程序化 Word 导出 (export_gam_docx())。

17.5 与平台相关的说明

17.5.1 macOS

问题最少。系统库推荐使用 Homebrew 安装。若需要 PDF 报告：

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

配合 RStudio Desktop 运行良好。要点：

- 企业／受限机器：若临时目录或用户数据目录受到限制，设置的持久化与报告生成可能失败。应用程序会在 R 控制台中显示警告，但仍继续运行，只是不保存设置。
- **LaTeX**：从 R 中安装 TinyTeX（命令与上文 macOS 相同），或从 miktex.org 安装 MiKTeX。
- 过长的文件路径：Windows 有 260 个字符的路径长度上限。请保持工作目录路径简短。

17.5.3 Linux

各发行版之间差异最大。Ubuntu/Debian 用户在编译 R 软件包之前可能需要先装好系统库：

```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

若需 PDF 报告：使用 `tinytex::install_tinytex()`，或从系统软件包管理器安装 `texlive-full`。

对于无显示设备的无头服务器，`plotly` 交互式图形在报告中会以静态图形作为退回方案渲染。Shiny 应用程序本身仍需要一个网页浏览器连接。

17.6 优雅降级

`mgcvUI` 的设计使其即便在缺少可选组件时也能工作：

缺失的组件	行为
无 LaTeX	PDF 选项从报告格式下拉菜单中隐去。HTML 与 Word 仍然可用。
无网络 / 字体加载失败	Roboto Condensed 字体由系统无衬线字体替代。控制台记录一条提示信息。
无 RSQLite / 只读文件系统	设置不会跨会话保存。控制台显示警告。除此之外应用程序正常运行。
<code>gratia::appraise</code> 失败	诊断标签页与报告中显示占位信息。可在控制台中改用 <code>mgcv::gam.check()</code> 。
临时目录不可写	报告生成失败，并给出明确的错误信息。可通过把 <code>TMPDIR</code> 环境变量设为可写位置来解决。

17.7 故障排除

17.7.1 “PDF option is not available” (PDF 选项不可用)

未检测到 LaTeX 安装。请在 R 中运行：

```
install.packages("tinytex")
tinytex::install_tinytex()
```

然后重启应用程序，PDF 选项便会出现。

17.7.2 “Settings will not persist” (设置将不会保存)

SQLite 数据库目录无法创建或无法写入。请检查以下位置的权限：

- macOS/Linux: ~/Library/Application Support/R/mgcUI/ 或 ~/.local/share/R/mgcUI/
- Windows: %APPDATA%/R/data/mgcUI/

17.7.3 “Diagnostics plot unavailable” (诊断图不可用)

`gratia::appraise()` 函数在复杂模型（含许多张量交互项、非标准分布族）上偶尔会失败。这是一个已知的上游问题。变通办法：在 R 控制台中运行 `mgcv::gam.check(model)` 以直接查看诊断图。

17.7.4 “Report error: number of items to replace...” (报告错误：待替换的项数……)

这是早期版本中一个与 `gratia` 软件包相关的已知问题。请确保你安装的是最新版本的 `mgcvUI`。当前版本会捕获该错误，并在生成报告时为诊断部分放置一个占位内容。

17.7.5 7880 端口已被占用

若应用程序因 7880 端口被占用（来自上一次会话）而无法启动，请运行：

```
# macOS/Linux:
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 在 Linux 上安装软件包失败

缺少系统库。请安装上文 Linux 一节中列出的开发头文件，然后重新运行 `install.packages()`。

参考文献

- Wood, S. N. (2003). “Thin-plate regression splines”. In: *Journal of the Royal Statistical Society (B)* 65.1, pp. 95–114. DOI: [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). “Stable and efficient multiple smoothing parameter estimation for generalized additive models”. In: *Journal of the American Statistical Association* 99.467, pp. 673–686. DOI: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models”. In: *Journal of the Royal Statistical Society (B)* 73.1, pp. 3–36. DOI: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.
- Wood, S. N., N. Pya, and B. Säfken (2016). “Smoothing parameter and model selection for general smooth models (with discussion)”. In: *Journal of the American Statistical Association* 111, pp. 1548–1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).

基础概念

概述

简体中文版

基础概念 (*Foundations*) • *The Valuation Engineer*

Bert Craytor (主编) • 2026 年 8 月 3 日

版本 1.0.0

DOI: [10.67330/82vssv81](https://doi.org/10.67330/82vssv81)

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

《基础概念》(*Foundations*) 栏目发表对现代估价方法论基础概念的精确、可引用的阐述。各条目选取执业估价师会遇到的术语——有时以其他名称出现，有时甚至未被察觉——并将其置于形式化的语境之中：这些概念从何而来、其技术含义是什么、在实务中出现于何处，以及它们为何关乎可辩护性。

创刊号确立了特征价格 (hedonic) 框架的六个核心概念，其编排使每个条目都建立在前面条目所确立的词汇之上。这一序列从《异质商品》(*Heterogeneous Good*，即房地产之所以需要估价方法论的根本属性) 出发，经由《特征空间》(*Characteristics Space*) 与《特征价格函数》(*Hedonic Price Function*，即组织房地产所代表的属性组合的形式化工具)，再到《隐含价格》(*Implicit Price*) 与《特征价格回归》(*Hedonic Regression*，即如何从市场证据中还原每个属性的边际贡献)，最后到达《潜变量》(*Latent Variable*，即任何特征价格模型中未被还原的残差信号究竟代表什么)。

一个贯穿全部六个条目的运行示例——来自加利福尼亚州帕西菲卡 (Pacifica, California) 某社区的八个可比案例——构成本期的教学主线。

为何是这六个概念，为何按此顺序

六个条目的选择并非任意。每个概念都是现代特征价格方法论的承重构件，每个概念都确实出现在日常估价实务之中，而且每个概念被误解或未被命名的频率之高，使得将其写成可引用的形式具有实际价值。

条目的排序由概念间的依赖关系决定。异质商品是根基：它命名了房地产（以及二手车、劳动力市场和其他若干商品）的一种属性，正是这一属性使整个特征价格框架成为必要。特征空间是将异质商品表示为向量的形式化工具；没有它，任何进一步的论述都无法精确展开。特征价格函数是将特征空间中的点映射到价格的均衡对象。隐含价格是它的偏导数——销售比较网格上的每一项调整都可以被解读为对这些边际贡献的断言，无论估价师是否如此称呼它们。特征价格回归是从观测到的交易中还原隐含价格的经验估计量；正是在这一条目中，理论转变为数据分析。潜变量则从回归残差开始的地方接续：那些未被观测的属性，其效应以有结构的信号而非可丢弃的噪声的形式存留于残差之中。

按此顺序阅读的读者，应当能获得一条连续的思想线索：从“估价方法论究竟为何存在？”一直到“我们应当如何看待拟合模型中的残差？”——以作者的经验而言，大多数估价师需要多年实务的零散积累才能走完这段旅程。

运行示例

一个由加利福尼亚州帕西菲卡八个住宅可比案例构成的数据集贯穿全部六个条目：

可比案例	建筑面积 (平方英尺)	地块 (平方英尺)	景观	状况	成交价格
A	1,650	9,000	有	良好	\$1,425,000
B	1,650	7,200	无	良好	\$1,280,000
C	1,850	7,500	无	良好	\$1,407,500
D	1,450	7,000	无	良好	\$1,155,000
E	1,700	8,000	有	优秀	\$1,470,000
F	1,550	7,100	无	一般	\$1,177,500
G	1,700	10,500	无	良好	\$1,392,500
H	1,600	7,300	无	良好	\$1,312,500

这个示例被有意设计得足够小，可以徒手推理；同时又足够丰富，可以演练每一个概念。条目 001 考察可比案例 A 与 B 之间 \$145,000 的差价，将其视为两项特征（景观与地块）纠缠于单一配对销售观测中的联合贡献。条目 002 将全部八个可比案例置于一个四维特征空间中的点，并说明朴素的距离度量会对它们作出错误排序。条目 003 为该社区的特征价格函数提出一个函数形式。条目 004 以解析方式还原隐含价格，并将最初的 A 与 B 差价分解为景观与地块贡献的 \$127,039，另有 \$18,000 未获解释。条目 005 在 R 中正式拟合该模型，并报告完整的 lm() 输出，包括标准误与残差。条目 006 考察模型无法解释的那一个残差——可比案例 H 上 +\$46,455 的偏差——并将其与潜变量的解释框架联系起来，这一框架正是作者在别处发展的残差约束法（Residual Constraint Approach）方法论的基础。

该数据集可在本期的资源库中获取，路径为 `shared/data/pacifica_comps.csv`，供希望复现任何计算的读者使用。

如何阅读本期

初次接触特征价格方法论的读者应按顺序阅读六个条目。条目间的依赖是实实在在的：每个条目都假定读者已掌握前面条目所确立的词汇，跳跃式阅读会招致条目标本身无法预料的困惑。

先前接触过特征价格回归的读者——包括大多数修过高级统计课程、使用过自动估价模型 (AVM) 或读过批量估价文献的估价师——可将本期用作温习与参考。各条目在写作上力求可被独立引用；本刊其他文章若提及隐含价格，可通过 DOI 链接到条目 004，而无需重新推导该概念。

对残差约束法方法论特别感兴趣的读者，应先读条目 006 以理解潜变量的解释框架，然后回到条目 003–005 了解残差约束法 (RCA) 所扩展的特征价格工具，再到条目 001–002 了解支撑一切的基础词汇。

关心什么使估价调整具有可辩护性这一问题的读者会发现：依次阅读各条目的“为何关乎可辩护性”部分，便相当于一篇关于基于调整的估价方法之方法论地位的短文。这篇短文并未在本期任何地方被明确地组装起来，但它是每个条目的潜在主题。

编辑立场

《基础概念》不是教科书式栏目。每个条目短到可以一次读完，面向已经熟稔估价实务的读者写作，旨在给出概念的精确表述而非面面俱到的阐释。其目标是提供一个稳定的参照物，供本刊其他文章引用。

条目经编辑部审阅，并可能附有特邀领域专家的评论。条目以知识共享署名许可协议 (Creative Commons Attribution) 发布，并通过 Crossref（作为规范引用）与 Zenodo（作为版本化存档）双渠道获得 DOI。每个条目都可能随时间推移而被打磨——小幅修订更新 Zenodo 版本，重大修订则同时更新两个 DOI 并附编辑说明描述改动。Crossref DOI 始终解析到最新版本；特定历史版本可通过 Zenodo 寻址。

未来各期将把《基础概念》的词汇扩展到相邻领域：空间依赖、条件期望、内生性、遗漏变量偏差、可识别性，以及区分代理变量与潜变量的形式化工具。欢迎读者就值得阐述的概念提出建议，来函请寄编辑部。

本期六个条目：

1. 异质商品 (Heterogeneous Good) —— 估价方法论为何必须存在。
2. 特征空间 (Characteristics Space) —— 属性组合的向量表示。
3. 特征价格函数 (Hedonic Price Function) —— 从特征到价格的均衡映射。
4. 隐含价格 (Implicit Price) —— 每项特征的边际市场价格。
5. 特征价格回归 (Hedonic Regression) —— 从观测交易中估计该函数，附完整的 R 实例演算。
6. 潜变量 (Latent Variable) —— 从残差中还原有结构的信号。

参考文献。

Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Craytor, B. (2025). Residual Constraint Approach: Framework & Protocol. Zenodo. [doi:10.5281/zenodo.14787917](https://doi.org/10.5281/zenodo.14787917)

附录：Prolog 形式化

以下机器可读的配套文件 (000-overview.pl) 在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```
%% 000-overview.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Overview
%% Title: Overview
%% Author: Bert Craytor, Editor
%% Piece ID: foundations_overview
%% Kind: editorial_piece (not a concept; sits outside the dep graph)
%%
%% The Overview is an editorial piece, not a defined concept. Its .pl
%% companion records the issue's structure (which entries appear in
%% what reading order), the running example, and the reader pathways
%% the editor proposes for readers with different backgrounds.

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(foundations_overview, "Overview", foundations_editorial,
```

```

%      "Editorial overview of a Foundations issue. Frames the concepts
%      covered, situates them within appraisal practice, and is not
%      itself a defined concept in the dependency graph.",
%      published, 'dictionary-2026.1').
%
% Plus the six concept terms covered by this issue (heterogeneous_good,
% characteristics_space, hedonic_price_function, implicit_price,
% hedonic_regression, latent_variable).
% -----

:- use_terms([ foundations_overview,
                heterogeneous_good, characteristics_space,
                hedonic_price_function, implicit_price,
                hedonic_regression, latent_variable,
                residual_constraint_approach ]).
:- use_relations([depends_on/2]).

% =====
% PIECE METADATA
% =====

article_piece(foundations_overview).
article_kind(overview).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.overview').

% =====
% ISSUE STRUCTURE: the six concept entries in dependency-ordered sequence
% =====
%
% The published reading order. Each entry assumes vocabulary established
% in the previous ones. This is the same as queries:reading_order/1
% derived from the depends_on graph; recording it here lets readers and
% downstream tools query the issue's structure without invoking the
% topological-sort machinery.

issue_entry(1, heterogeneous_good).
issue_entry(2, characteristics_space).
issue_entry(3, hedonic_price_function).
issue_entry(4, implicit_price).
issue_entry(5, hedonic_regression).
issue_entry(6, latent_variable).

% Each entry's one-line theme as stated in the Overview's closing list.
entry_theme(heterogeneous_good,
    "Why valuation methodology must exist at all.").
entry_theme(characteristics_space,
    "The vector representation of bundle attributes.").
entry_theme(hedonic_price_function,
    "The equilibrium mapping from characteristics to prices.").
entry_theme(implicit_price,
    "The marginal market price of each characteristic.").
entry_theme(hedonic_regression,
    "Estimating the function from observed transactions, with a full R worked example
    ↪.").
entry_theme(latent_variable,
    "Recovering structured signal from residuals.").

% =====

```

```
% THE RUNNING EXAMPLE
% =====
%
% A single dataset of 8 Pacifica residential comparables runs through
% all six entries. The data lives in kb/pacifica_comps.pl and is
% consulted by each entry that uses it. The Overview's role here is
% only to point at it.

running_example_dataset(pacifica_comps).
running_example_n_comps(8).
running_example_location(pacifica_california).

% Which entries exercise the dataset.
:- consult('../..../kb/pacifica_comps').

% =====
% READER PATHWAYS
% =====
%
% The editor proposes four pathways through the issue, calibrated to
% the reader's background. These are advisory orderings; the canonical
% reading order is issue_entry/2 above.

reader_pathway(new_to_hedonic_methodology,
    "Read entries 1-6 in order. Dependencies are real and skipping invites confusion
    ↪ .",
    [heterogeneous_good, characteristics_space, hedonic_price_function,
     implicit_price, hedonic_regression, latent_variable]).

reader_pathway(prior_exposure_to_hedonic_regression,
    "Use as a refresher and atomic reference. Read in any order.",
    [hedonic_regression, hedonic_price_function, implicit_price,
     characteristics_space, heterogeneous_good, latent_variable]).

reader_pathway(interested_in_rca_methodology,
    "Read 006 first for the framing, then 003-005, then 001-002.",
    [latent_variable, hedonic_price_function, implicit_price,
     hedonic_regression, heterogeneous_good, characteristics_space]).

reader_pathway(interested_in_adjustment_defensibility,
    "Read the 'Why it matters for defensibility' sections in entry order. Forms an
    ↪ implicit short essay.",
    [heterogeneous_good, characteristics_space, hedonic_price_function,
     implicit_price, hedonic_regression, latent_variable]).

% =====
% EDITORIAL POSTURE
% =====

editorial_posture(
    "Foundations is not a textbook department. Each entry is short, written for an
    ↪ audience already fluent in appraisal practice, and designed to give a precise
    ↪ formulation rather than a comprehensive exposition.").

publication_model(
    [ license(cc_by),
      crossref_doi(canonical_citation),
      zenodo_doi(versioned_archive),
      revision_policy("Minor revisions bump Zenodo version; major revisions bump both
```

```
↪DOIs with editorial note.") ]).

% =====
% FUTURE DIRECTIONS
% =====
%
% Concepts the editor flags as deserving treatment in future Foundations
% issues. Recorded here so the journal's planning is machine-queryable.

future_topic(spatial_dependence).
future_topic(conditional_expectation).
future_topic(endogeneity).
future_topic(omitted_variable_bias_extended).
future_topic(identification).
future_topic(proxy_vs_latent_variable).
```

异质商品

简体中文版

基础概念 (*Foundations*) • *The Valuation Engineer*

Bert Craytor • 2026 年 8 月 3 日

版本 1.0.0

DOI: [10.67330/s1scsv65](https://doi.org/10.67330/s1scsv65)

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

形式定义。异质商品 (heterogeneous good) 是指不可相互替换的商品：每一件商品都是买方分别赋予价值的特征的独特组合，而这些特征本身并不单独交易。

直观阐释。同质商品 (homogeneous good) 是任何一件都可以的商品：一桶 WTI 原油、一蒲式耳二号黄玉米、电网上的一千瓦时电、某只股票的一股。这些商品背后的分级标准或规格的存在，正是为了使单件商品可以相互替换，从而使价格适用于类别而非特定的那一件。异质商品则是具体那一件很重要的商品——是这栋房子，而不是那栋房子；是这辆二手车，而不是旁边那一辆。

在技术意义上，两个特点使商品成为异质商品。第一，买方在意多个可区分的属性，而这些属性在市场上的各件商品之间存在差异。第二，这些属性因商品的生产技术而被捆绑在一起，无法单独购买。你不可能单独购买 200 平方英尺的建筑面积；你不可能脱离一宗土地而获得一个学区。特征成捆而来，市场为整个组合定价，而不是为特征本身定价。

房地产是异质商品的典型代表，并且异常鲜明地体现了这一属性。区位特征是完全不可复制的：任何两宗地块都不可能占据同一位置，因此每宗房地产至少拥有一项属性（其坐标）将它与市场上的所有其他房地产区分开来。物理、法律与环境属性又在此之上叠加了更多的变异维度，使得整体的特征向量对每宗房地产而言都是独一无二的。

估价师在何处遇到它。市场比较法 (sales comparison approach) 中的每一项调整，都是对估价对象与可比案例乃异质商品这一事实的隐含承认。如果房地产是同质的，成交价格本身就可以直接比较，根本不需要比较网格。网格是一种通过对组合构成差异作出调整，来调和不同组合价格的尝试。值得注意的是网格不做什么：它从不将成交价格分解为各特征的价值贡献。传统网格没有价值贡献一栏——只有逐项直接估计的调整。

成本法通过从组成部分重建组合来应对异质性——基础、框架、装修、折旧——并对每一部分分别定价。收益法则通过将组合抽象为单一的现金流轮廓来应对异质性，由资本化率与收益率把组合层面的异质性吸收进一个源自市场的乘数。三种估价方法都是对“房地产是异质商品”这一根本事实的方法论回应。它们的区别在于如何应对这一事实，而不在于是否面对它。

估价师在非正式场合同样会遇到异质性：每当他们思考可比案例的选择、配对销售分析 (paired-sales analysis) 或“同类相比”的比较时。寻找在尽可能多的属性上相似的可比案例，这一直觉正是实务工作者对异质性的回应：将估价对象与可比案例之间的组合距离最小化，组合中未被观测的成分对成交价格差异的贡献便随之减小。

为何关乎可辩护性。房地产的异质性正是估价方法论必须作为一门学科存在的原因。如果房地产可以相互替换，市场价值 (market value) 便可从挂牌价格直接观测，估价将沦为查表。正因为并非如此，价值必须从相关但不相同的房地产的价格中推断得出，而每一步推断都是一次需要说明理由的方法论选择。

正式承认异质性，可以厘清 USPAP 及各准则条款所关注的可辩护性所依赖的若干实务问题。可比案例的选择，是在特征空间中挑选与估价对象足够接近、以致从中还原的隐含价格可以推广的样本。一项调整是否合理，取决于其背后的隐含价格能否从市场证据中得到支持。各估价方法之间的最终调和，反映的是估价师对哪种方法最可信地应对了这宗房地产的异质性的判断。一旦“异质商品”被点明为深层条件，上述每一个可辩护性问题都会变得更加清晰。

估价实例演算。 加利福尼亚州帕西菲卡 (Pacifica, California) 同一社区最近的两笔成交：

属性	可比案例 A	可比案例 B
建筑面积 (平方英尺)	1,650	1,650
卧室 / 卫浴	3 / 2	3 / 2
地块 (平方英尺)	9,000	7,200
景观	有	无
成交价格	\$1,425,000	\$1,280,000

在同质商品的视角下，这是同一宗房地产：建筑面积相同、房间数相同、社区相同。在异质商品的视角下，它们是两个不同的组合，在买方分别赋予价值的两项特征（地块与景观）上存在差异。\$145,000 的价格差是这两项属性的联合贡献——两个隐含价格纠缠在同一个配对之中。

要将隐含价格逐一还原，需要更多的配对，最好是每次只改变一项属性。再有三四笔覆盖有景观/无景观、大地块/小地块各种组合的成交，两个隐含价格便开始可以分别识别。当足够多的可比案例提供了足够的变异时，特征价格回归 (hedonic regression) 便可正式完成这一解缠；而当变异有限时，估价师只能局限于区间化的配对销售推理，其可靠性明确取决于各组合在特征空间中彼此有多接近。

这个例子说明了为什么异质性是根本条件：即使是两栋“完全相同”的房子也并不相同，而估价的方法论工具之所以存在，正是为了处理它们之间的距离。

参考文献。

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

交叉引用：特征空间 (*characteristics space*)；特征价格函数 (*hedonic price function*)；隐含价格 (*implicit price*)。

附录：Prolog 形式化

以下机器可读的配套文件 (001-heterogeneous-good.pl) 在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```
%% 001-heterogeneous-good.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 001
%% Title:    Heterogeneous Good
%% Author:   Bert Craytor
%% Concept:  heterogeneous_good
%% Version:  0.1.0-draft
%%
%% This file is the Prolog galley for the article: it encodes the
%% article's main claims, worked example, and inference rules in a form
%% that other Prolog code in the journal can consult, query, and verify
%% against. Read alongside the entry.tex / entry.pdf prose.

% =====
% DICTIONARY IMPORT
% =====
```

```

%
% The journal-wide dictionary is the source of truth for term definitions
% and relation schemas. The use_module below pulls those facts in at
% runtime; the comments that follow reproduce the relevant entries so a
% reader skimming this file in isolation can see what's been imported.

:- use_module('.././../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(heterogeneous_good, "Heterogeneous Good", economic_theory,
%     "Heterogeneous goods are goods that are not interchangeable
%     because each is a distinct bundle of characteristics that buyers
%     value separately, and whose characteristics are not separately
%     traded.", published, 'dictionary-2026.1').
%
% term(characteristics_space, "Characteristics Space", economic_theory,
%     "The characteristics space of a class of heterogeneous goods is
%     the vector space whose coordinates index the attributes that
%     buyers value, and in which each good is represented by the
%     vector of its attribute levels.", published, 'dictionary-2026.1').
%
% term(implicit_price, "Implicit Price", economic_theory,
%     "The implicit price of a characteristic is the partial derivative
%     of the hedonic price function with respect to that characteristic,
%     evaluated at a point in the characteristics space; it represents
%     the marginal market price of an incremental unit of the
%     characteristic, holding all other characteristics fixed.",
%     published, 'dictionary-2026.1').
%
% term(hedonic_price_function, ...).
%
% relation(depends_on, 2, "depends on", [foundations_concept, foundations_concept],
%     foundations_methodology, "...", accepted, 'dictionary-2026.1').
% relation(related, 2, ...).
%
% These facts are NOT declared in this file; the use_module above is
% what makes them available at runtime.
% -----

:- use_terms([ % Foundations concepts (terms/foundations.pl)
    heterogeneous_good, characteristics_space, implicit_price,
    hedonic_price_function,
    % Practice terms (terms/valuation-practice.pl)
    real_estate, homogeneous_good,
    sales_comparison_approach, cost_approach, income_approach,
    comp_selection, paired_sales_analysis, adjustment,
    reconciliation, uspap, standards_rule ]).
:- use_relations([related/2, depends_on/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(heterogeneous_good).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.001').

% =====

```

```
% ARTICLE-LOCAL VOCABULARY
% =====
%
% Terms the article needs that are NOT yet in the dictionary, and that
% the editor decided are too narrow to promote. These are field names
% for the worked-example property records — not concepts in the
% journal's ontology.

local_term(gla_sqft).
local_term(bedrooms).
local_term(baths).
local_term(lot_sqft).
local_term(view).

% =====
% FORMAL DEFINITION (restated as Prolog rule)
% =====
%
% From the formal definition: a good is heterogeneous IF buyers care
% about multiple distinguishable attributes that vary across items in
% the market, AND those attributes are bundled by production rather
% than separately traded.

%! heterogeneous(?Good) is nondet.
%
% True when Good satisfies both bundling criteria. The two predicates
% buyers_value_multiple_attributes/1 and attributes_bundled_by_production/1
% must be asserted for specific goods elsewhere; here they characterize
% the formal definition itself.

heterogeneous(Good) :-
    buyers_value_multiple_attributes(Good),
    attributes_bundled_by_production(Good).

% =====
% INTUITIVE FRAMING — CONTRASTS WITH HOMOGENEOUS GOODS
% =====

% Canonical examples of homogeneous goods cited in the article.
canonical_homogeneous_example(wti_crude).
canonical_homogeneous_example(no_2_yellow_corn).
canonical_homogeneous_example(kilowatt_hour).
canonical_homogeneous_example(common_stock_share).

% Their distinguishing feature: grading or specification standards make
% individual items interchangeable.
interchangeable_via_grading(wti_crude).
interchangeable_via_grading(no_2_yellow_corn).
interchangeable_via_specification(kilowatt_hour).
interchangeable_via_specification(common_stock_share).

% Canonical heterogeneous goods.
canonical_heterogeneous_example(real_estate).
canonical_heterogeneous_example(used_car).
canonical_heterogeneous_example(unique_machinery).

% Why real estate exhibits heterogeneity unusually strongly:
% locational characteristics are perfectly non-replicable.
buyers_value_multiple_attributes(real_estate).
```

```

attributes_bundled_by_production(real_estate).
non_replicable_attribute(real_estate, location).

% =====
% APPRAISER ENCOUNTERS
% =====
%
% The three classical valuation approaches are all methodological
% responses to heterogeneity. They differ in HOW they navigate the
% bundle, not in WHETHER they confront it.

valuation_approach(sales_comparison_approach).
valuation_approach(cost_approach).
valuation_approach(income_approach).

responds_to_heterogeneity(sales_comparison_approach,
    "Reconciles the prices of distinct bundles by adjusting for composition
    ↪ differences; the grid records only adjustments, estimated directly, and never
    ↪ decomposes a price into feature value contributions.").
responds_to_heterogeneity(cost_approach,
    "Reconstructs the bundle from priced components: foundation, framing, finishes,
    ↪ depreciation.").
responds_to_heterogeneity(income_approach,
    "Abstracts the bundle into a single cash-flow profile; cap/yield rates absorb
    ↪ heterogeneity.").

% Practitioner-level encounters with heterogeneity.
practitioner_response_to_heterogeneity(comp_selection,
    "Minimize bundle distance between subject and comp in characteristics space.").
practitioner_response_to_heterogeneity(paired_sales_analysis,
    "Isolate an attribute's contribution by holding other attributes constant.").
practitioner_response_to_heterogeneity(adjustment,
    "Quantify the price impact of one differing attribute at a time.").

% =====
% DEFENSIBILITY
% =====
%
% Why naming "heterogeneous good" sharpens USPAP defensibility questions.

defensibility_consequence(comp_selection,
    "Comps must be close enough in characteristics space that recovered implicit
    ↪ prices generalize.").
defensibility_consequence(adjustment,
    "Each adjustment must be supportable from market evidence of the implicit price.")
    ↪ .
defensibility_consequence(reconciliation,
    "Choice among approaches reflects which best navigates THIS property's
    ↪ heterogeneity.").

% Root claim: valuation methodology exists as a discipline BECAUSE
% real estate is heterogeneous. If it were homogeneous, value would be
% directly observable from posted prices.
root_claim(
    "Valuation methodology exists as a discipline because real estate is heterogeneous
    ↪; in the homogeneous case, value would be directly observable.").

% =====
% WORKED EXAMPLE — PACIFICA COMPS

```

```
% =====
%
% Two recent sales in the same Pacifica, California neighborhood. The
% example demonstrates that even "identical" houses are not identical:
% the $145,000 price spread is the joint contribution of differences in
% lot size and view, two implicit prices entangled in a single pair.

% The 8-comp Pacifica dataset is shared across entries 001-006 and
% lives in kb/pacifica_comps.pl. Consulting it here brings property/1,
% in_market/2, attribute_value/3, and sale_price/2 into this article's
% namespace.
:- consult('../..kb/pacifica_comps').

% --- Derived facts from the example -----

% The attributes that differ between the two comps.
differing_attribute(comp_a, comp_b, lot_sqft).
differing_attribute(comp_a, comp_b, view).

% The price differential and what it's attributable to.
price_differential(comp_a, comp_b, 145000).
attributed_to(price_differential(comp_a, comp_b, 145000),
               [implicit_price_of(lot_sqft), implicit_price_of(view)]).

% --- Article's argument about identification -----

% A single pair cannot separately identify the two implicit prices —
% they are entangled in the single $145K observation.
joint_identification_only(price_differential(comp_a, comp_b, 145000),
                           [lot_sqft, view]).

% Separate identification requires more pairs varying one attribute at
% a time. With sufficient variation, hedonic regression performs the
% disentanglement formally; with limited variation, the appraiser uses
% bracketed paired-sales reasoning, whose reliability depends on how
% close the bundles are in characteristics space.
identification_strategy(few_pairs,
    "Bracketed paired-sales reasoning. Reliability depends on bundle proximity in
    ↪ characteristics space.").
identification_strategy(many_pairs_with_variation,
    "Hedonic regression (see hedonic_regression)").

% =====
% INFERENCE RULES
% =====

%! comparable(?A, ?B) is nondet.
%
% A and B are comparable when they are distinct properties in the same
% market. This is the operational definition of "comp" used implicitly
% in the article.

comparable(A, B) :-
    property(A),
    property(B),
    A \== B,
    in_market(A, M),
    in_market(B, M).
```

```

%! attribute_difference(?A, ?B, ?Attr, ?DA, ?DB) is nondet.
%
% A and B differ on Attr, with DA the value at A and DB the value at B.
% Use this to enumerate the dimensions on which two comps actually vary.

attribute_difference(A, B, Attr, DA, DB) :-
    comparable(A, B),
    attribute_value(A, Attr, DA),
    attribute_value(B, Attr, DB),
    DA \== DB.

%! identifiable_implicit_price(?Attr) is nondet.
%
% An attribute's implicit price is identifiable from the observed comps
% if there exists at least one pair of comparables differing ONLY in
% that attribute. With our two-comp example this fails for lot_sqft AND
% view (both differ in the same pair), which is exactly the article's
% point.

identifiable_implicit_price(Attr) :-
    comparable(A, B),
    findall(X, attribute_difference(A, B, X, _, _), Diffs),
    Diffs == [Attr].

% =====
% CROSS-REFERENCES
% =====
%
% Concepts the article cross-references (corresponding to the
% \foundationsref{} macros in entry.tex).

cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).

```

特征空间

简体中文版

基础概念 (Foundations) • The Valuation Engineer

Bert Craytor • 2026 年 8 月 3 日

版本 1.0.0

DOI: 10.67330/3yymh010

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

形式定义。一类异质商品的特征空间 (characteristics space)，是以买方赋予价值的各项属性为坐标的向量空间；在这一空间中，每件商品由其属性水平构成的向量来表示。

直观阐释。上一条目确立了异质商品是特征的独特组合。自然的下一个问题是：我们应当如何形式化地描述并比较这些组合？答案出自 Lancaster (1966)：把每件商品视为向量空间中的一个点，该空间的坐标轴正是买方在意的各项特征。

在这种表示下，一宗帕西菲卡 (Pacifica) 住宅不再是”帕克帕西菲卡的一栋房子”，而是一个向量，其坐标包括建筑面积、地块面积、景观、状况、学区、距海远近等等。若两宗房地产在所有这些属性上的取值都相同，它们将占据特征空间中的同一个点；在任何属性上存在差异的房地产则占据不同的点。空间的几何结构编码了两宗房地产之间的接近或差异程度，并涵盖了买方作出区分的全部维度。

空间的维数取决于所研究市场中哪些属性是相关的。对典型的住宅估价而言，相关空间可能有十来个维度：物理维度（建筑面积、地块、卧室、卫浴、房龄、质量、状况）、区位维度（学区、景观、距便利设施或不利设施的远近）以及法律 / 经济维度（区划、地役权）。对商业地产估价而言，相关维度转向现金流属性（租约结构、租户质量、出租率），但底层的几何思想是相同的。

特征空间之所以有用，在于它把实务工作者的定性直觉（”这些可比案例与估价对象相似”）转化为定量命题（”这些可比案例在特征空间中距估价对象很近”）。一旦定义了距离，每一个销售比较决策都可以被当作关于空间中位置的陈述来审视。

估价师在何处遇到它。估价师无时无刻不在特征空间中工作，只是通常没有叫出它的名字。销售比较网格的每一行是一条坐标轴；每个可比案例是一个由坐标值构成的行向量；每一项调整都是在保持其他轴不变的情况下沿一条轴移动的尝试。

我们将在本期通篇使用的八个帕西菲卡可比案例直接演示了这一构造。每个可比案例都是由建筑面积、地块面积、景观与状况张成的四维特征空间中的一个点：

可比案例	建筑面积（平方英尺）	地块（平方英尺）	景观	状况	成交价格
A	1,650	9,000	有	良好	\$1,425,000
B	1,650	7,200	无	良好	\$1,280,000
C	1,850	7,500	无	良好	\$1,407,500
D	1,450	7,000	无	良好	\$1,155,000
E	1,700	8,000	有	优秀	\$1,470,000
F	1,550	7,100	无	一般	\$1,177,500
G	1,700	10,500	无	良好	\$1,392,500
H	1,600	7,300	无	良好	\$1,312,500

上一条目中的可比案例 A 与 B 是前两行；其余六个案例对示例作了扩展，我们将在本期余下部分充分利用这些扩展。

每个可比案例都是 $\mathbb{R}^4$ 中的一个点：

- 可比案例 A 位于 (1650, 9000, 1, 2)。
- 可比案例 B 位于 (1650, 7200, 0, 2)。
- 可比案例 G 位于 (1700, 10500, 0, 2)。

景观与状况坐标以数值编码（景观取 {0, 1}，状况取 {1, 2, 3}）；价格不是特征空间的坐标，而是空间中位置的函数。

为何关乎可辩护性。特征空间框架使“可比案例选择究竟是什么”变得更加清晰。可比案例不只是“一宗相似的房地产”，而是特征空间相关区域中靠近估价对象的一个点。可比案例在空间中距估价对象越近，所需的调整越小，得出的指示价值也越可靠。

这一重新表述暴露了非正式网格所掩盖的若干实务问题：

- 哪些维度重要？坐标轴的选择是一个建模决策。只有当市场确实对景观单独定价时，把“景观”设为一条轴才有意义；只有当某个市场细分对“风水朝向”定价时，把它设为一条轴才有意义。可辩护的估价要求为坐标轴本身辩护，而不仅仅是为轴上的取值辩护。
- 适用什么距离度量？两个可比案例的价格可能相差 \$200,000，因为它们在特征空间中相距很远；但距离的计算需要对每条轴进行尺度化。未经变换的欧氏距离把一平方英尺建筑面积等同于一平方英尺地块，而市场并不这样看。
- 可比案例是否位于价格函数表现良好的区域？如果估价对象落在特征空间中采样稀疏的角落，就没有任何可比案例靠近它，价值指示必然是外推性的。

估价实例演算。八个帕西菲卡可比案例占据了四维特征空间中一个相对紧凑的区域。考虑这样一个问题：对于位于 (1,700, 7,500, 1, 2) 的假想估价对象——建筑面积 1,700 平方英尺、地块 7,500 平方英尺、有景观、状况良好——哪个可比案例“最近”？

如果我们不加考虑地使用坐标轴未经尺度化的欧氏距离，地块维度会占据主导地位，因为其数值 (7,000–10,500) 远大于景观 (0–1) 或状况 (1–3)。各案例到估价对象的距离如下：

可比案例	距离（未尺度化欧氏距离）	备注
A	1,501	景观相符；地块占主导
B	308	景观不符；地块接近
C	230	景观不符；地块接近
D	559	景观不符；建筑面积偏差
E	502	景观相符；状况不同
F	422	景观与状况均不同
G	3,000	景观不符；地块相距甚远
H	224	景观不符；距离最近

按原始距离，可比案例 H 是最近邻。但可比案例 H 没有景观，而估价对象有。朴素的度量误导了我们：地块维度的数值尺度挤掉了类别型的景观维度，而后者对价格实际上更重要。

一个可辩护的距离度量应当以各条轴的隐含价格 (implicit price) 来尺度化——粗略地说，就是该轴每变动一个单位值多少美元。一旦坐标轴按隐含价格尺度化，景观（约值 \$100,000）便不再被地块（约每平方英尺 \$25）所淹没。于是，在景观上与估价对象相符的可比案例 A 便浮现为恰当的首选可比案例。

这正是后续条目中发展的隐含价格工具所提供的东西：对特征空间的一种有原则的尺度化，使“接近”从几何上的猜测变为源自市场的计算。

参考文献。

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

交叉引用： 异质商品 (*heterogeneous good*)；特征价格函数 (*hedonic price function*)；隐含价格 (*implicit price*)。

附录：Prolog 形式化

以下机器可读的配套文件 (002-characteristics-space.pl) 在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```
%% 002-characteristics-space.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 002
%% Title:   Characteristics Space
%% Author:  Bert Craytor
%% Concept: characteristics_space
%% 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(characteristics_space, "Characteristics Space", economic_theory,
%      "The characteristics space of a class of heterogeneous goods is
%      the vector space whose coordinates index the attributes that
%      buyers value, and in which each good is represented by the
%      vector of its attribute levels.", published, 'dictionary-2026.1').
%
% term(heterogeneous_good, ...).
% term(hedonic_price_function, ...).
% term(implicit_price, ...).
%
% generalized_by(characteristics_space, lancaster_1966).
% depends_on(characteristics_space, heterogeneous_good).
% -----

:- use_terms([ characteristics_space, heterogeneous_good,
                hedonic_price_function, implicit_price,
                comp_selection, sales_comparison_approach,
                real_estate ]).
:- use_relations([related/2, depends_on/2, generalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(characteristics_space).
article_issue(volume(1), number(1), year(2026)).
```

```

article_id('foundations.2026.002').

% =====
% LOCAL VOCABULARY
% =====
%
% Narrow article-local field names + the dimensions of the worked
% example's characteristics space.

local_term(dimension).
local_term(coordinate).
local_term(euclidean_distance).
local_term(distance_metric).
local_term(nearest_neighbor).

% =====
% FORMAL ATTRIBUTION
% =====

% The construction is due to Lancaster (1966), as noted in citations.
attributed_construction(characteristics_space, lancaster_1966).

% =====
% THE ARTICLE'S CLAIM ABOUT REPRESENTATION
% =====
%
% A heterogeneous good becomes a point in a vector space. Each axis is
% a characteristic buyers value; each good is the vector of its
% attribute levels.

representation_of(heterogeneous_good, point_in_vector_space).

% Examples of dimensions relevant to residential appraisal.
residential_dimension(gla_sqft).
residential_dimension(lot_sqft).
residential_dimension(view).
residential_dimension(cond).
residential_dimension(bedrooms).
residential_dimension(baths).
residential_dimension(age).
residential_dimension(quality).
residential_dimension(school_district).
residential_dimension(distance_to_amenity).
residential_dimension(zoning).

% Examples of dimensions relevant to commercial appraisal — included
% to make the article's point that the dimensions shift with market.
commercial_dimension(lease_structure).
commercial_dimension(tenant_quality).
commercial_dimension(occupancy).
commercial_dimension(cash_flow_profile).

% =====
% APPRAISER ENCOUNTERS
% =====
%
% Mapping appraisal practice onto the characteristics-space framework.

practice_to_space_correspondence(

```

```

    sales_comparison_grid_line, coordinate_axis,
    "A grid line is an axis.").
practice_to_space_correspondence(
    comp_row, row_vector,
    "A comp is a vector of coordinate values.").
practice_to_space_correspondence(
    adjustment, movement_along_axis,
    "An adjustment moves along one axis holding others fixed.").

% =====
% WORKED EXAMPLE
% =====
%
% The 8 Pacifica comps are points in a 4D characteristics space spanned
% by GLA, lot_sqft, view, cond_numeric. Bring in the shared dataset.

:- consult(' ../kb/pacifica_comps ').

% Axes of the chosen space for this article's analysis.
axis(gla_sqft).
axis(lot_sqft).
axis(view).
axis(cond_numeric).

% Numeric encoding for non-numeric axes.
encode(view, no, 0).
encode(view, yes, 1).
encode(cond_numeric, 1, 1).
encode(cond_numeric, 2, 2).
encode(cond_numeric, 3, 3).

%! coord(?Property, ?Axis, ?Numeric) is nondet.
%
% Numeric coordinate of Property along Axis. Handles symbolic attribute
% values by encoding them.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V ; encode(A, V, N) ).

%! point(?Property, -Vector) is det.
%
% Vector representation of Property in the 4D characteristics space.

point(P, [GLA, Lot, View, Cond]) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond).

% =====
% THE HYPOTHETICAL SUBJECT
% =====
%
% The article tests distance metrics against this hypothetical property.

hypothetical_subject(subject).
attribute_value(subject, gla_sqft, 1700).
attribute_value(subject, lot_sqft, 7500).

```

```

attribute_value(subject, view,      yes).
attribute_value(subject, cond_numeric, 2).

% =====
% NAIVE EUCLIDEAN DISTANCE (the article's straw distance metric)
% =====
%
% Computes the unscaled Euclidean distance between Property and the
% hypothetical subject. Used to demonstrate that unscaled distances
% mislead: the lot dimension dominates because its numeric range is
% much larger than view (0..1) or cond (1..3).

naive_distance_to_subject(P, D) :-
    point(P, V_P),
    point(subject, V_S),
    sum_of_squares(V_P, V_S, S),
    D is sqrt(S).

sum_of_squares([], [], 0).
sum_of_squares([X|Xs], [Y|Ys], S) :-
    sum_of_squares(Xs, Ys, S0),
    D is X - Y,
    S is S0 + D * D.

% =====
% THE ARTICLE'S CENTRAL OBSERVATION
% =====
%
% Under naive Euclidean distance, Comp H is the nearest neighbor. But
% Comp H has no view and the subject does; view matters more to price
% than the naive metric admits. The fix (developed in entries 003-005)
% is to scale each axis by its implicit price.

claim(naive_metric_misleads,
    "Unscaled Euclidean distance crowds out small-range axes (like view) under large-
    ↪ range axes (like lot_sqft). A market-defensible distance scales each axis by its
    ↪ implicit price.").

% =====
% DEFENSIBILITY QUESTIONS RAISED
% =====
%
% Three questions the characteristics-space framing makes explicit.

defensibility_question(which_dimensions,
    "The choice of axes is a modeling decision. Defending an appraisal requires
    ↪ defending the axes, not just the values on them.").
defensibility_question(distance_metric,
    "Untransformed Euclidean distance treats one sqft of GLA as equivalent to one sqft
    ↪ of lot, which the market does not. Distance requires a scaling per axis.").
defensibility_question(extrapolation_region,
    "If the subject sits in a sparsely-sampled corner of the space, no comp is truly
    ↪ close and the indication is extrapolative.").

% =====
% INFERENCE RULES
% =====

%! same_point(?A, ?B) is nondet.

```

```

%
% Two properties occupy the same point in the characteristics space.
% (Doesn't occur in the running example, but the article notes that it
% would if all coordinates matched.)

same_point(A, B) :-
    point(A, V),
    point(B, V),
    A \== B.

%! nearest_neighbor_by_naive_distance(-P) is semidet.
%
% The comp closest to the subject under the naive Euclidean metric.
% Used to illustrate the article's point that the naive metric chooses
% Comp H, not Comp A.

nearest_neighbor_by_naive_distance(P) :-
    findall(D-Q,
        ( property(Q), Q \== subject, naive_distance_to_subject(Q, D) ),
        Pairs),
    keysort(Pairs, [_-P|_]).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).

```

特征价格函数

简体中文版

基础概念 (Foundations) • The Valuation Engineer

Bert Craytor • 2026 年 8 月 3 日

版本 1.0.0

DOI: [10.67330/9x1a0489](https://doi.org/10.67330/9x1a0489)

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

形式定义。特征价格函数 (hedonic price function) 是从异质商品特征空间中的点到成交价格的均衡映射，由竞争性市场中买方出价函数与卖方要价函数的包络构成。

直观阐释。我们现在有了两个要素：异质商品（特征的组合）与特征空间（每个组合是其中一点的向量空间）。特征价格函数将二者结合起来。它是为特征空间中每个点 $\mathbf{z}$ 赋予一个价格的函数 $p(\mathbf{z})$ ，告诉我们市场为具有该特定属性组合的商品支付多少。

Rosen (1974) 的关键洞见在于： $p(\mathbf{z})$ 不是由任何单个买方或卖方选定的，而是从市场均衡中涌现出来的。每个买方都有一个出价函数 (bid function)，表达在其偏好与收入给定的情况下，愿为任一组合支付的最高价格。每个卖方都有一个要价函数 (offer function)，表达在其成本与备选方案给定的情况下，愿意接受的最低价格。特征价格函数是这些出价与要价的包络——买卖双方实际成交所沿的边界。

对房地产而言， $p(\mathbf{z})$ 回答的是这样一个问题：如果你给定一个完整的特征组合，市场会为它给出什么价格？答案并非任意。它反映了市场中全体买方的总体偏好与全体卖方的总体成本与备选方案，并经由买方与组合的均衡匹配来调节。

特征价格函数是特征价格分析的核心理论对象。隐含价格是它的偏导数。特征价格回归是对它的经验估计。任何销售比较调整的可辩护性最终都可归结为关于这一函数的一个命题。

估价师在何处遇到它。估价师无时无刻不在对 $p(\mathbf{z})$ 进行推理——同样，通常没有叫出它的名字。若干实务活动都是对特征价格函数的隐性操作：

市场比较法中的调整。当估价师因估价对象有景观而可比案例 B 没有，而将可比案例 B 上调 \$100,000 时，估价师是在断言一个关于 $p(\mathbf{z})$ 的命题：即在其他特征保持不变的情况下，当景观特征从 0 变为 1 时，函数值增加 \$100,000。这是关于 p 在景观维度上局部斜率的陈述。

各估价方法之间的调和。成本法自下而上构建价值指示，将复制组合的各组成部分成本相加。收益法从现金流推导价值，而在均衡状态下，现金流反映产生类似现金流的组合的价格。市场比较法则直接从相似组合的观测成交中读取价格。在运转良好的市场中，三种方法应当收敛于 $p(\mathbf{z})$ 上的同一个点。当它们出现分歧时，估价师必须判断哪种方法对估价对象在特征空间中所处位置的函数测量得最可靠。

市场分析与时间调整。随着供给、需求、融资与预期的变化，特征价格函数会随时间发生移动。每年 6% 的升值率是关于 $p(\mathbf{z})$ 整体如何移动的命题；而“有景观溢价的房地产比无景观房地产升值更快”的发现，则是关于函数在特征空间中非均匀移动的命题。

为何关乎可辩护性。明确点出特征价格函数，可以厘清若干根本性的可辩护性问题：

函数是否存在？Rosen 框架假定存在一个竞争性市场，且交易量足以使均衡函数有意义。在交易稀薄的市场、快速变化的市场，或买方群体特质化的细分市场中，该函数可能定义不清或不稳定。在此类市场中执业的估价师应当承认，其推断工具是在一个不牢靠的理论基础上运行的。

函数是否处处相同？Rosen 的框架为每个市场细分给出一个函数，而不是全局一个函数。帕西菲卡 (Pacifica) 与旧金山 (San Francisco) 是不同的市场，各有不同的特征价格函数；即便在同一座城市内，独栋住宅与共管公寓也是不同的细分市场。将细分市场混为一谈的估价师，实际上是在隐含地假设同一个函数横跨两者，而这通常是错误的。

p 取什么函数形式? 该函数不必是线性、对数线性或任何其他特定形式。Rosen 只证明了 p 是出价与要价的包络; 他没有给定其形状。经验上, p 表现出非线性 (新增一间卧室的边际价值往往随建筑面积而变化)、交互作用 (景观溢价取决于地块朝向) 以及阈值效应 (低于某一质量门槛的厨房可能招致与其状况评分不成比例的折价)。可辩护的估计要求要么采用能够从市场中得到支持的函数形式, 要么使用能够从数据中还原函数形状的灵活估计量。

估价实例演算。对于所研究的帕西菲卡社区, 我们将使用一个简单的可加函数形式, 下一条目将对其作更充分的演练:

$$p(\mathbf{z}) = \beta_0 + \beta_{\text{GLA}} z_{\text{GLA}} + \beta_{\text{lot}} z_{\text{lot}} + \beta_{\text{view}} z_{\text{view}} + \beta_{\text{cond}} z_{\text{cond}}.$$

这一形式内嵌了两个强假设: 每项特征的贡献对其水平是线性的, 且各特征之间不存在交互作用。两个假设都是可检验的, 也经常被违背; 我们在此将其作为出发点采用。

对于本期运行示例中的八个可比案例, 后续条目的分析将估计出一组 β 值, 得到大致如下的函数:

参数	估计值
β_0 (截距)	\$80,777
β_{GLA} (每平方英尺)	\$575
β_{lot} (每平方英尺)	\$22.70
β_{view} (有/无)	\$86,179
β_{cond} (每级)	\$49,824

将这些值代入可加形式, 该函数便可为四维特征空间中的任何一点预测价格。代入可比案例 A 的坐标, 得到的预测价格接近其观测到的 \$1,425,000; 代入位于 (1700, 7500, 1, 2) 的假想估价对象, 则得到该房地产的价值指示。

本条目的要点不在于这些具体数字, 而在于概念上的一步: 在我们估计它之前, 该函数在原则上就已存在。函数是市场的属性。我们的估计是还原它的一种尝试。这种还原可能是不完全的、有偏的或有噪声的, 接下来的三个条目所讨论的正是如何负责任地完成这一还原。

参考文献。

Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Triplett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

交叉引用: 异质商品 (*heterogeneous good*); 特征空间 (*characteristics space*); 隐含价格 (*implicit price*); 特征价格回归 (*hedonic regression*)。

附录: Prolog 形式化

以下机器可读的配套文件 (003-hedonic-price-function.pl) 在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```

%% 003-hedonic-price-function.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 003
%% Title: Hedonic Price Function
%% Author: Bert Craytor
%% Concept: hedonic_price_function
%% 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(hedonic_price_function, "Hedonic Price Function", economic_theory,
% "The hedonic price function is the equilibrium mapping from
% points in the characteristics space of a heterogeneous good to
% transaction prices, formed as the envelope of buyer bid
% functions and seller offer functions in a competitive market.",
% published, 'dictionary-2026.1').
%
% term(bid_function, "Bid Function", economic_theory, ...).
% term(offer_function, "Offer Function", economic_theory, ...).
% term(sales_comparison_approach, ...).
% term(cost_approach, ...).
% term(income_approach, ...).
%
% formalized_by(hedonic_price_function, rosen_1974).
% depends_on(hedonic_price_function, heterogeneous_good).
% depends_on(hedonic_price_function, characteristics_space).
% -----

:- use_terms([ hedonic_price_function, heterogeneous_good,
               characteristics_space, implicit_price, hedonic_regression,
               bid_function, offer_function,
               sales_comparison_approach, cost_approach, income_approach,
               adjustment, reconciliation ]).
:- use_relations([related/2, depends_on/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_price_function).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.003').

% =====
% LOCAL VOCABULARY
% =====

local_term(equilibrium).
local_term(envelope).
local_term(functional_form).
local_term(market_segment).
local_term(linearity).
local_term(separability).

```

```

local_term(time_adjustment).
local_term(coefficient).
local_term(intercept).

% =====
% FORMAL ATTRIBUTION
% =====

attributed_formalization(hedonic_price_function, rosen_1974).

% =====
% CONSTRUCTION: ENVELOPE OF BID AND OFFER FUNCTIONS
% =====
%
% The Rosen (1974) construction:  $p(z)$  is the envelope of buyer bid
% functions and seller offer functions.

primitive_of_construction(bid_function, buyer).
primitive_of_construction(offer_function, seller).

constructed_as_envelope_of(hedonic_price_function,
                           [bid_function, offer_function]).

emerges_from(hedonic_price_function, market_equilibrium).

% Property that flows from the construction: the function is not
% chosen by any single market participant.
property(hedonic_price_function, not_set_by_individual_participant).
property(hedonic_price_function, equilibrium_object).
property(hedonic_price_function, defined_per_market_segment).

% =====
% APPRAISER ENCOUNTERS — IMPLICIT MANIPULATIONS OF  $p(z)$ 
% =====

implicit_manipulation_of_hedonic(
    adjustment,
    sales_comparison_approach,
    "Asserts a local slope of  $p$  in the relevant characteristic dimension.").
implicit_manipulation_of_hedonic(
    reconciliation,
    multi_approach_valuation,
    "Three approaches should converge on the same point of  $p$ ; divergence requires
    ↪ editorial judgment.").
implicit_manipulation_of_hedonic(
    time_adjustment,
    market_analysis,
    " $p$  shifts over time; appreciation rates are statements about that shift.").

% =====
% DEFENSIBILITY QUESTIONS
% =====
%
% The article names three questions the hedonic-price-function framing
% sharpens.

defensibility_question(function_existence,
    "In thin or rapidly-changing markets, the equilibrium function may be poorly
    ↪ defined; the inferential machinery operates on a shaky foundation.").

```

```

defensibility_question(segment_specificity,
    "One function per market segment, not one function globally. Conflating segments
    ↪ is usually wrong.").
defensibility_question(functional_form,
    "Rosen showed only that p is an envelope, not its shape. Empirically nonlinear
    ↪ with interactions and thresholds; estimation requires defending the form.").

% =====
% WORKED EXAMPLE: ADDITIVE LINEAR SPECIFICATION
% =====
%
% For the Pacifica neighborhood, the article adopts the additive linear
% specification:
% 
$$p(z) = \beta_0 + \beta_{\text{gla}} * z_{\text{gla}} + \beta_{\text{lot}} * z_{\text{lot}} + \beta_{\text{view}} * z_{\text{view}} + \beta_{\text{cond}} * z_{\text{cond}}$$

%
% The next entry (004) takes the coefficients as implicit prices; entry
% 005 estimates them by hedonic regression.

:- consult('../..kb/pacifica_comps').

specification_form(additive_linear).
specification_assumption(linearity_in_levels,
    "Contribution of each characteristic is linear in its level.").
specification_assumption(no_interactions,
    "Characteristics do not interact in the price function.").
specification_caveat(
    "Both assumptions are testable and frequently violated; adopted here as a starting
    ↪ point.").

% Coefficient values reported in the article (estimated in entry 005,
% asserted here so this entry's worked example can be queried).
coefficient(intercept,      80777.07).
coefficient(beta_gla,       574.93).
coefficient(beta_lot,       22.70).
coefficient(beta_view,      86179.03).
coefficient(beta_cond,      49824.33).

%! predicted_price(?Property, -Predicted) is det.
%
% The fitted additive-linear hedonic price function applied to a
% property's characteristics. Symbolic view/cond are mapped to {0,1}
% and {1,2,3} respectively.

predicted_price(P, Predicted) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond),
    coefficient(intercept, B0),
    coefficient(beta_gla, Bg),
    coefficient(beta_lot, Bl),
    coefficient(beta_view, Bv),
    coefficient(beta_cond, Bc),
    Predicted is B0 + Bg * GLA + Bl * Lot + Bv * View + Bc * Cond.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V
    ; A == view. V == ves -> N = 1
    ; A == cond. V == 1 -> N = 2
    ; A == cond. V == 2 -> N = 3
    ; A == cond. V == 3 -> N = 4
    ; A == cond. V == 4 -> N = 5
    ; A == cond. V == 5 -> N = 6
    ; A == cond. V == 6 -> N = 7
    ; A == cond. V == 7 -> N = 8
    ; A == cond. V == 8 -> N = 9
    ; A == cond. V == 9 -> N = 10
    ; A == cond. V == 10 -> N = 11
    ; A == cond. V == 11 -> N = 12
    ; A == cond. V == 12 -> N = 13
    ; A == cond. V == 13 -> N = 14
    ; A == cond. V == 14 -> N = 15
    ; A == cond. V == 15 -> N = 16
    ; A == cond. V == 16 -> N = 17
    ; A == cond. V == 17 -> N = 18
    ; A == cond. V == 18 -> N = 19
    ; A == cond. V == 19 -> N = 20
    ; A == cond. V == 20 -> N = 21
    ; A == cond. V == 21 -> N = 22
    ; A == cond. V == 22 -> N = 23
    ; A == cond. V == 23 -> N = 24
    ; A == cond. V == 24 -> N = 25
    ; A == cond. V == 25 -> N = 26
    ; A == cond. V == 26 -> N = 27
    ; A == cond. V == 27 -> N = 28
    ; A == cond. V == 28 -> N = 29
    ; A == cond. V == 29 -> N = 30
    ; A == cond. V == 30 -> N = 31
    ; A == cond. V == 31 -> N = 32
    ; A == cond. V == 32 -> N = 33
    ; A == cond. V == 33 -> N = 34
    ; A == cond. V == 34 -> N = 35
    ; A == cond. V == 35 -> N = 36
    ; A == cond. V == 36 -> N = 37
    ; A == cond. V == 37 -> N = 38
    ; A == cond. V == 38 -> N = 39
    ; A == cond. V == 39 -> N = 40
    ; A == cond. V == 40 -> N = 41
    ; A == cond. V == 41 -> N = 42
    ; A == cond. V == 42 -> N = 43
    ; A == cond. V == 43 -> N = 44
    ; A == cond. V == 44 -> N = 45
    ; A == cond. V == 45 -> N = 46
    ; A == cond. V == 46 -> N = 47
    ; A == cond. V == 47 -> N = 48
    ; A == cond. V == 48 -> N = 49
    ; A == cond. V == 49 -> N = 50
    ; A == cond. V == 50 -> N = 51
    ; A == cond. V == 51 -> N = 52
    ; A == cond. V == 52 -> N = 53
    ; A == cond. V == 53 -> N = 54
    ; A == cond. V == 54 -> N = 55
    ; A == cond. V == 55 -> N = 56
    ; A == cond. V == 56 -> N = 57
    ; A == cond. V == 57 -> N = 58
    ; A == cond. V == 58 -> N = 59
    ; A == cond. V == 59 -> N = 60
    ; A == cond. V == 60 -> N = 61
    ; A == cond. V == 61 -> N = 62
    ; A == cond. V == 62 -> N = 63
    ; A == cond. V == 63 -> N = 64
    ; A == cond. V == 64 -> N = 65
    ; A == cond. V == 65 -> N = 66
    ; A == cond. V == 66 -> N = 67
    ; A == cond. V == 67 -> N = 68
    ; A == cond. V == 68 -> N = 69
    ; A == cond. V == 69 -> N = 70
    ; A == cond. V == 70 -> N = 71
    ; A == cond. V == 71 -> N = 72
    ; A == cond. V == 72 -> N = 73
    ; A == cond. V == 73 -> N = 74
    ; A == cond. V == 74 -> N = 75
    ; A == cond. V == 75 -> N = 76
    ; A == cond. V == 76 -> N = 77
    ; A == cond. V == 77 -> N = 78
    ; A == cond. V == 78 -> N = 79
    ; A == cond. V == 79 -> N = 80
    ; A == cond. V == 80 -> N = 81
    ; A == cond. V == 81 -> N = 82
    ; A == cond. V == 82 -> N = 83
    ; A == cond. V == 83 -> N = 84
    ; A == cond. V == 84 -> N = 85
    ; A == cond. V == 85 -> N = 86
    ; A == cond. V == 86 -> N = 87
    ; A == cond. V == 87 -> N = 88
    ; A == cond. V == 88 -> N = 89
    ; A == cond. V == 89 -> N = 90
    ; A == cond. V == 90 -> N = 91
    ; A == cond. V == 91 -> N = 92
    ; A == cond. V == 92 -> N = 93
    ; A == cond. V == 93 -> N = 94
    ; A == cond. V == 94 -> N = 95
    ; A == cond. V == 95 -> N = 96
    ; A == cond. V == 96 -> N = 97
    ; A == cond. V == 97 -> N = 98
    ; A == cond. V == 98 -> N = 99
    ; A == cond. V == 99 -> N = 100
    ; A == cond. V == 100 -> N = 101
    ; A == cond. V == 101 -> N = 102
    ; A == cond. V == 102 -> N = 103
    ; A == cond. V == 103 -> N = 104
    ; A == cond. V == 104 -> N = 105
    ; A == cond. V == 105 -> N = 106
    ; A == cond. V == 106 -> N = 107
    ; A == cond. V == 107 -> N = 108
    ; A == cond. V == 108 -> N = 109
    ; A == cond. V == 109 -> N = 110
    ; A == cond. V == 110 -> N = 111
    ; A == cond. V == 111 -> N = 112
    ; A == cond. V == 112 -> N = 113
    ; A == cond. V == 113 -> N = 114
    ; A == cond. V == 114 -> N = 115
    ; A == cond. V == 115 -> N = 116
    ; A == cond. V == 116 -> N = 117
    ; A == cond. V == 117 -> N = 118
    ; A == cond. V == 118 -> N = 119
    ; A == cond. V == 119 -> N = 120
    ; A == cond. V == 120 -> N = 121
    ; A == cond. V == 121 -> N = 122
    ; A == cond. V == 122 -> N = 123
    ; A == cond. V == 123 -> N = 124
    ; A == cond. V == 124 -> N = 125
    ; A == cond. V == 125 -> N = 126
    ; A == cond. V == 126 -> N = 127
    ; A == cond. V == 127 -> N = 128
    ; A == cond. V == 128 -> N = 129
    ; A == cond. V == 129 -> N = 130
    ; A == cond. V == 130 -> N = 131
    ; A == cond. V == 131 -> N = 132
    ; A == cond. V == 132 -> N = 133
    ; A == cond. V == 133 -> N = 134
    ; A == cond. V == 134 -> N = 135
    ; A == cond. V == 135 -> N = 136
    ; A == cond. V == 136 -> N = 137
    ; A == cond. V == 137 -> N = 138
    ; A == cond. V == 138 -> N = 139
    ; A == cond. V == 139 -> N = 140
    ; A == cond. V == 140 -> N = 141
    ; A == cond. V == 141 -> N = 142
    ; A == cond. V == 142 -> N = 143
    ; A == cond. V == 143 -> N = 144
    ; A == cond. V == 144 -> N = 145
    ; A == cond. V == 145 -> N = 146
    ; A == cond. V == 146 -> N = 147
    ; A == cond. V == 147 -> N = 148
    ; A == cond. V == 148 -> N = 149
    ; A == cond. V == 149 -> N = 150
    ; A == cond. V == 150 -> N = 151
    ; A == cond. V == 151 -> N = 152
    ; A == cond. V == 152 -> N = 153
    ; A == cond. V == 153 -> N = 154
    ; A == cond. V == 154 -> N = 155
    ; A == cond. V == 155 -> N = 156
    ; A == cond. V == 156 -> N = 157
    ; A == cond. V == 157 -> N = 158
    ; A == cond. V == 158 -&
```

```

; A == view, V == no -> N = 0
; V = N
).

% =====
% THE CENTRAL CONCEPTUAL MOVE
% =====
%
% The function exists, in principle, before estimation. The function is
% a property of the market; estimation is recovery.

claim(function_precedes_estimation,
    "The hedonic price function is a property of the market and exists conceptually
    ↪ prior to any specific fitted model. Estimation is recovery, and recovery may be
    ↪ partial, biased, or noisy.").

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(implicit_price).
cross_reference(hedonic_regression).

```

隐含价格

简体中文版

基础概念 (Foundations) • The Valuation Engineer

Bert Craytor • 2026 年 8 月 3 日

版本 1.0.0

DOI: [10.67330/md6wke17](https://doi.org/10.67330/md6wke17)

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

形式定义。一项特征的隐含价格 (implicit price) 是特征价格函数对该特征的偏导数，在特征空间中某一点取值；它代表在其他特征保持不变的情况下，该特征增加一个单位的边际市场价格。

直观阐释。特征价格函数 $p(\mathbf{z})$ 为每个组合赋予一个价格。然而，估价实务中的大多数问题并不关乎完整组合的价格，而是关乎组合中某个具体组成部分的价格。景观值多少钱？多出一百平方英尺能增加多少？“良好”与“优秀”状况之间的差异贡献多少？

隐含价格回答的正是这些问题。这一术语源自 Rosen (1974)，他证明了差异化产品的观测价格连同其特征”定义了一组隐含的或‘特征’价格”；Lancaster (1966) 则提供了将商品视为特征组合的底层观点。对每项特征 z_i 而言，隐含价格是在其他所有特征保持不变时， $p(\mathbf{z})$ 随 z_i 变化的变化率：

$$\pi_i(\mathbf{z}) = \frac{\partial p(\mathbf{z})}{\partial z_i}.$$

隐含价格的单位是美元每单位特征——建筑面积每平方英尺多少美元、地块每平方英尺多少美元、状况每提升一级多少美元。对于景观这样的二元特征，隐含价格就是从无景观 (0) 变为有景观 (1) 所支付的全部溢价。

“隐含”一词点明了核心事实：这些价格没有被挂牌。没有人以 \$575 出售“一平方英尺建筑面积”。这一价格隐含在确实被挂牌的均衡组合价格之中，而特征价格分析工具的存在正是为了把它提取出来。

隐含价格是特征价格框架中与估价直接相关的产出。销售比较网格中的调整，原则上就是隐含价格的估计值乘以特征差异。每一项调整都是对特征空间中相关点处 π_i 取值的一次押注。

估价师在何处遇到它。销售比较调整网格就是一张隐含价格应用表，即便估价师并不这样称呼它。网格中“因建筑面积比估价对象少 200 平方英尺而 +\$10,000”的调整，是在断言：在特征空间的相关位置上，建筑面积的隐含价格为每平方英尺 \$50。”因景观较差而调整 \$100,000”的网格条目，是在断言二元景观特征的隐含价格为 \$100,000。市场比较法的全部可辩护性，都取决于这些隐含价格能否从市场中得到支持。然而在实践中，估价师是直接估计调整额的——依据配对案例、经验或同行做法——从未计算过底层的价值贡献；传统网格只记录调整本身。隐含价格的解读是对这些调整所断言内容的一种诠释，而不是对其推导方式的描述。

这一命名也阐明了为什么配对销售分析是推导调整的基础技术。恰好只在一项特征上不同的两个可比案例，可以直接估计该特征的隐含价格：价格差除以特征差。配对销售分析的可靠性，取决于这两个案例在特征空间中是否真的足够接近，以致从中还原的隐含价格能够推广到估价对象。

隐含价格有三个对实务重要的性质：

它们是局部的。在 1,650 平方英尺处的建筑面积隐含价格，未必与在 2,500 平方英尺处相同。特征价格函数一般是非线性的，其偏导数随特征空间中的位置而变化。由 1,500 平方英尺的配对销售分析推导出的调整，未必适用于 3,000 平方英尺的估价对象。

它们是有条件的。海滨细分市场中景观的隐含价格，与内陆细分市场中的并不相同。偏导数是在其他特征保持不变的情况下求取的，但该导数的取值取决于其他特征被固定在什么位置。

它们依赖于市场。隐含价格是均衡价格，因而反映交易发生时买方群体的偏好。即使特征本身没有变化，市场的变动也会改变隐含价格。时间调整实际上是在承认：昨天的隐含价格未必是今天的。

为何关乎可辩护性。隐含价格的视角揭示了非正式网格可能掩盖的调整可辩护性问题：

你处在特征空间中的什么位置？只有在估价对象于特征空间中所处位置附近估计出的隐含价格才是可辩护的。由内陆可比案例推导出的景观溢价被套用于海滨估价对象，等于在函数行为未知的特征空间区域上作外推。

局部几何结构是什么？如果两项特征存在交互作用——例如景观在更大的地块上更值钱——那么将它们分别作可加处理就是一种模型设定错误。景观的隐含价格依赖于地块面积，忽略交互作用的调整低估了函数的实际结构。

隐含价格是如何还原的？配对销售分析、回归系数、行业经验法则与“专业判断”指向的是同一个量——隐含价格——但其认识论地位截然不同。估价师的可辩护性取决于说明每项调整所使用的隐含价格出自哪个来源。

估价实例演算。对于采用上一条目可加设定的帕西菲卡 (Pacifica) 运行示例，隐含价格就是线性函数的系数：

$$\pi_{\text{GLA}} = \beta_{\text{GLA}}, \quad \pi_{\text{lot}} = \beta_{\text{lot}}, \quad \pi_{\text{view}} = \beta_{\text{view}}, \quad \pi_{\text{cond}} = \beta_{\text{cond}}.$$

使用条目 003 的估计系数：

特征	隐含价格	单位
建筑面积	\$575	每平方英尺
地块	\$22.70	每平方英尺
景观	\$86,179	每 (有/无)
状况	\$49,824	每级

这些隐含价格蕴含了一次配对销售分析。回到条目 001 中可比案例 A 与可比案例 B 的差异：

- 可比案例 A 有景观 ($\Delta z_{\text{view}} = +1$)；贡献 +\$86,179。
- 可比案例 A 的地块大 1,800 平方英尺 ($\Delta z_{\text{lot}} = +1,800$)；贡献 $1,800 \times \$22.70 = +\$40,860$ 。
- 其他所有特征相同。
- 预测价格差： $\$86,179 + \$40,860 = \$127,039$ 。

观测到的价格差为 \$145,000。隐含价格分解解释了其中约 \$127,000；其余 \$18,000 反映了小数据集中的抽样变异（更一般地说，反映了潜变量条目将要讨论的那类未观测因素）。

有两个操作层面的要点值得注意。第一，线性假设承担着实实在在的工作：在非线性的 $p(\mathbf{z})$ 下，同样的地块面积差在小地块上可能比在大地块上更值钱，上述可加分解便会有偏。第二，地块每平方英尺 \$22.70 的隐含价格是增加一平方英尺的边际价值，而不是地块的总价值。一块 9,000 平方英尺的地块对组合价格的总贡献为 $9,000 \times \$22.70 = \$204,300$ ，但再增加一平方英尺只使组合价格提高 \$22.70。

这一区分——边际与总量——是日常估价语言中大量混淆的根源，也是明确命名隐含价格所带来的最重要的概念收益之一。

参考文献。

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.

- Freeman, A. M., III (1979). Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues. *Scandinavian Journal of Economics*, 81(2), 154–173.
- Taylor, L. O. (2003). The Hedonic Method. In *A Primer on Nonmarket Valuation* (pp. 331–393). Kluwer Academic Publishers.
- Palmquist, R. B. (2005). Property Value Models. In *Handbook of Environmental Economics* (Vol. 2, pp. 763–819). North-Holland.
- Triplett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

交叉引用： 异质商品 (*heterogeneous good*)；特征空间 (*characteristics space*)；特征价格函数 (*hedonic price function*)；特征价格回归 (*hedonic regression*)。

附录：Prolog 形式化

以下机器可读的配套文件 (004-implicit-price.pl) 在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```
%% 004-implicit-price.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 004
%% Title:   Implicit Price
%% Author:  Bert Craytor
%% Concept: implicit_price
%% 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(implicit_price, "Implicit Price", economic_theory,
%      "The implicit price of a characteristic is the partial
%      derivative of the hedonic price function with respect to that
%      characteristic, evaluated at a point in the characteristics
%      space; it represents the marginal market price of an
%      incremental unit of the characteristic, holding all other
%      characteristics fixed.", published, 'dictionary-2026.1').
%
% term(hedonic_price_function, ...).
% term(paired_sales_analysis, ...).
% term(adjustment, ...).
%
% formalized_by(implicit_price, rosen_1974).
% depends_on(implicit_price, hedonic_price_function).
% -----

:- use_terms([ implicit_price, hedonic_price_function, characteristics_space,
               heterogeneous_good, hedonic_regression,
               sales_comparison_approach, paired_sales_analysis, adjustment,
               comp_selection ]).
:- use_relations([related/2, depends_on/2, formalized_by/2]).
```

```

% Provenance: the term originates with Rosen (1974).
formalized_by(implicit_price, rosen_1974).

% =====
% ARTICLE SUBJECT
% =====

article_concept(implicit_price).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.004').

% =====
% LOCAL VOCABULARY
% =====

local_term(partial_derivative).
local_term(marginal).
local_term(total).
local_term(locality).
local_term(conditionality).
local_term(market_dependence).

% =====
% IMPLICIT PRICE AS PARTIAL DERIVATIVE
% =====
%
%  $\pi_i(z) = \partial p(z) / \partial z_i$ 

implicit_price_definition(
    "Implicit price of characteristic i at point z = partial derivative of p with
    ↪ respect to z_i, evaluated at z.").

unit_of_implicit_price(continuous, "dollars per unit of the characteristic").
unit_of_implicit_price(binary, "the full premium between the two states").

% =====
% THREE PROPERTIES OF IMPLICIT PRICES (per the article)
% =====

property_of_implicit_prices(local,
    "The implicit price of GLA at 1,650 sf is not necessarily the same as at 2,500 sf.
    ↪ p is generally nonlinear; partial derivatives vary with location.").
property_of_implicit_prices(conditional,
    "Implicit price depends on where the OTHER characteristics are held fixed. View
    ↪ premium on a beachfront submarket differs from inland.").
property_of_implicit_prices(market_dependent,
    "Implicit prices are equilibrium prices reflecting the current buyer pool. Market
    ↪ shifts change implicit prices even when characteristics do not.").

% =====
% IMPLICIT-PRICE / ADJUSTMENT EQUIVALENCE
% =====
%
% Every adjustment in a sales-comparison grid can be READ as asserting an
% implicit price applied to a characteristic difference. This is an
% interpretation of what the adjustment asserts, not a description of how
% traditional adjustments are derived in practice.

adjustment_as_implicit_price_application(

```

"Interpretation: adjustment magnitude = (implicit price of attribute at relevant
 ↪ point) * (characteristic difference between comp and subject) --- what the
 ↪ adjustment asserts, not how it is derived.").

adjustment_derivation_in_practice(

"Traditional adjustments are estimated directly --- from matched pairs, experience
 ↪, or peer practice --- without computing the underlying value contributions; the
 ↪ traditional grid records only the adjustments themselves.").

```
% =====
% LINEAR-FORM IMPLICIT PRICES (from entry 003's adopted specification)
% =====
%
% Under the additive linear specification, implicit prices ARE the
% regression coefficients. The article reproduces them:

:- consult('..../kb/pacifica_comps').

% Implicit price per characteristic, as inherited from the additive
% form of p adopted in entry 003.
implicit_price_of(gla_sqft, 575.00, "per square foot").
implicit_price_of(lot_sqft, 22.70, "per square foot").
implicit_price_of(view, 86179.03, "per yes/no").
implicit_price_of(cond, 49824.33, "per step on the ordinal scale").

% =====
% PAIRED-SALES DECOMPOSITION: COMP A vs COMP B
% =====
%
% Returns to the article-001 paired comparison and decomposes it using
% implicit prices.
%
% Comp A and Comp B share GLA, bedrooms, baths, condition.
% They differ in: view (+1) and lot_sqft (+1800).
% Predicted differential = 86179 + 40860 = 127039.
% Observed differential = 145000.
% Unexplained = 17961 (~$18,000).

paired_sales_decomposition(comp_a, comp_b,
  [ contribution(view, 86179),
    contribution(lot_sqft, 40860) ],
  127039,
  145000,
  17961).

% Derived from the data + coefficients (for the linter to type-check
% against the dataset rather than the hand-computed values).
contribution_for_pair(A, B, Attr, Contribution) :-
  coord(A, Attr, VA),
  coord(B, Attr, VB),
  Diff is VA - VB,
  Diff =\= 0,
  implicit_price_of(Attr, Pi, _),
  Contribution is Diff * Pi.

% Tiny coord helper for symbolic view/cond values.
coord(P, A, N) :-
  attribute_value(P, A, V),
  ( number(V) -> N = V
```

```

; A == view, V == yes -> N = 1
; A == view, V == no  -> N = 0
; V = N
).

% =====
% MARGINAL vs TOTAL CONTRIBUTION
% =====
%
% The article emphasizes the distinction. The implicit price of lot is
% $22.70 per square foot (marginal); the TOTAL contribution of Comp A's
% 9,000-sf lot is  $9,000 * 22.70 = \$204,300$ .

distinction(marginal_vs_total,
    "Implicit price is MARGINAL ($22.70 per additional sqft of lot). Total
    ↪ contribution of a 9,000-sf lot is  $9,000 * \$22.70 = \$204,300$ . Confusing them is a
    ↪ common source of error in everyday appraisal language.").

%! total_contribution(?Property, ?Attribute, -Total) is det.
%
% Total dollar contribution of a characteristic to a property's bundle
% price under the linear specification.

total_contribution(P, A, Total) :-
    coord(P, A, V),
    implicit_price_of(A, Pi, _),
    Total is V * Pi.

% =====
% DEFENSIBILITY QUESTIONS RAISED
% =====

defensibility_question(location_in_space,
    "An implicit price is defensible only when estimated near the subject's location
    ↪ in characteristics space; a view premium from inland comps applied to a beachfront
    ↪ subject extrapolates into unknown territory.").
defensibility_question(local_geometry,
    "If view interacts with lot size, treating them as additively separable
    ↪ misspecifies the function. The implicit price of view depends on lot.").
defensibility_question(epistemic_provenance,
    "Paired-sales analysis, regression coefficient, industry rule, and professional
    ↪ judgment all point to the same quantity but have different epistemic standing.
    ↪ Defensibility requires declaring the source.").

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(hedonic_regression).

```

特征价格回归

简体中文版

基础概念 (Foundations) • The Valuation Engineer

Bert Craytor • 2026 年 8 月 3 日

版本 1.0.0

DOI: [10.67330/ams1t139](https://doi.org/10.67330/ams1t139)

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

形式定义。 特征价格回归 (hedonic regression) 是根据异质商品市场中观测到的交易对特征价格函数进行的估计，得到进入模型的各项特征的隐含价格的经验估计值。

直观阐释。 前面的条目将特征价格函数及其隐含价格描述为理论对象。该函数在原则上存在，但在实践中我们只能观测到有限的一组交易：成交的组合，以及它们成交的价格。特征价格回归就是连接二者的经验桥梁。

设定很直接。每笔交易提供一个 $(\mathbf{z}_i, p_i)$ 对：成交组合的特征向量及其成交价格。这些数据对的集合是（未被观测的）特征价格函数的一个样本。回归 (regression) 为该样本拟合一个函数形式，估计出刻画该函数的参数。参数估出之后，拟合函数的偏导数便给出隐含价格的估计值。

在回归运行之前，必须作出三项建模选择：

1. 哪些特征进入模型？这一选择决定了被估函数的维度。遗漏相关特征会使其效应渗入与之相关的已纳入特征的系数之中（遗漏变量偏差）。
2. 什么函数形式？线性、对数线性、双对数、多项式、基函数展开、半参数、机器学习。这一选择限定了被估函数可以呈现的形状，以及它能还原哪些隐含价格模式。
3. 什么估计量？普通最小二乘 (OLS)、广义最小二乘、稳健 M 估计、分位数回归、空间计量估计量、惩罚回归。这一选择决定了估计量在优化什么，以及哪些假设支撑其推断。

每项选择在方法论上都有实质意义，也都带有可辩护性上的后果。

估价师在何处遇到它。 特征价格回归是批量估价 (mass appraisal) 的基础，也是大多数自动估价模型 (AVM) 的引擎。当县税务评估员每年对 10 万宗地块重新估值时，承担工作的是对近期成交所作的特征价格回归。当 Zillow 或 Redfin 给出自动估计时，产出它的是一个特征价格模型（往往是树集成而非 OLS，但在概念上仍是特征价格估计量）。当贷款机构的二级市场 AVM 判定某宗房地产的再融资通过与否时，涉及的也是同一套机制。

单宗房地产估价也越来越多地借助特征价格回归，即便最终的价值指示来自销售比较。由社区样本回归估出的隐含价格，能够远比估价师直觉或经验法则百分比更可辩护地支持调整幅度。回归系数成为网格中所用调整的证据，而非意见。

为何关乎可辩护性。 基于回归的估计具有与配对销售分析不同的可辩护性轮廓。优势在于：回归使用全部可得成交，而非一次只用两笔；它可以在同时控制许多其他特征的情况下分离出单项特征的效应；它给出量化每个隐含价格不确定性的标准误。代价在于：回归强加了一个可能与真实价格函数不符的函数形式；它比配对销售分析需要更多数据；其结果可能对样本与模型设定的选择敏感。

基于回归的估价中反复出现若干可辩护性问题：

样本是否恰当？ 回归估计的是特征价格函数在样本所处市场细分与时间段内的运作。用内陆成交作回归再套用于海滨估价对象，无论统计结果多么漂亮，都是可辩护性上的失败。

模型设定是否恰当？ 线性设定是一个强假设。如果真实函数在某一阈值以上对建筑面积呈现收益递减，线性拟合就会在该区间内错估隐含价格。诊断、残差图与备选设定是可辩护工作流程的组成部分。

标准误意味着什么？标准误刻画的是在假定的模型设定下的抽样变异。它不涵盖模型设定错误、遗漏变量偏差或样本选择问题。错误设定下的小标准误，比正确设定下的大标准误更危险。

结果能否与其他方法互相印证？一个可辩护的、基于回归的隐含价格，应当与配对销售分析的提示一致，与成本法的构件成本所暗示的一致，也与熟悉该市场的估价师的看法一致。与三者都相左的回归系数是一份邀请函——邀请你去调查，而不是去发表。

估价实例演算。现在我们可以八个帕西菲卡（Pacifica）可比案例上演完整的特征价格回归机制。数据集位于 `shared/data/pacifica_comps.csv`。

R 代码：

```
comps <- read.csv("shared/data/pacifica_comps.csv")
fit <- lm(price ~ gla + lot + view + cond, data = comps)
summary(fit)
```

输出：

```
Call:
lm(formula = price ~ gla + lot + view + cond, data = comps)

Residuals:
    Min       1Q   Median       3Q      Max
-17995  -8243  -5433  -1409   46455

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  80777.07   164509.35   0.491   0.6571
gla           574.93     109.76   5.238   0.0135 *
lot           22.70      10.38   2.187   0.1165
view        86179.03   31342.97   2.750   0.0708 .
cond       49824.33   27761.27   1.795   0.1706
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30481 on 3 degrees of freedom
Multiple R-squared:  0.9707, Adjusted R-squared:  0.9317
F-statistic: 24.86 on 4 and 3 DF, p-value: 0.01231
```

解读。拟合模型将特征价格函数估计为：

$$\hat{p}(\mathbf{z}) = 80,777 + 575 z_{\text{GLA}} + 22.70 z_{\text{lot}} + 86,179 z_{\text{view}} + 49,824 z_{\text{cond}}$$

每个系数都是对应特征的隐含价格估计值。建筑面积每平方英尺贡献 \$575（标准误最窄、显著性最强）；景观为一宗帕西菲卡房地产贡献约 \$86,000；状况每提升一级（例如从“良好”到“优秀”）约值 \$50,000。

若干注意事项一望即知：

样本极小。八个观测、四个预测变量，只剩三个残差自由度。标准误相应地很宽，地块、景观与状况的 p 值尽管在经济上合乎情理，却未达到常规的 0.05 门槛。这是小样本的伪象，而不是这些特征无关紧要的证据。若有 80 个而非 8 个可比案例， p 值将大幅缩小。

整体拟合很强。调整后 $R^2 = 0.93$ ，模型 p 值为 0.012，表明所纳入的特征联合解释了样本中大部分价格变异。联合拟合强而单项显著性弱的组合，是多重共线性的教科书式信号；在这里它是结构性的：建筑面积、地块与状况在样本中共同变动。

一个残差格外突出。最大残差为 +\$46,455，对应可比案例 H。其余七个残差全部落在 $\pm \$18,000$ 的区间内。可比案例 H 的某种特质没有被拟合函数捕捉——模型对其价格的系统性低估，几乎是第二大残差的三倍。

这个离群残差不是噪声。它是一个信号：可比案例 H 的某项未观测特征正以纳入变量无法还原的方式对其价格作出贡献。潜变量条目讨论的正是这一问题。

与配对销售推理的相互验证。条目 004 的配对销售分析仅凭景观与地块就预测出可比案例 A 与 B 之间 \$127,039 的差异。回归给出的隐含价格与之完全相同（因为它们正是那次计算的来源）。回归将配对销售逻辑同时推广到全部八个可比案例，但在这一样本量下并未产生任何本质上新的东西；它确认了配对分析，同时报告了配对销售推理无法给出的标准误。

参考文献。

Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.

Griliches, Z. (1961). Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. In *The Price Statistics of the Federal Government*. National Bureau of Economic Research.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Triplett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

交叉引用： 异质商品 (*heterogeneous good*)；特征空间 (*characteristics space*)；特征价格函数 (*hedonic price function*)；隐含价格 (*implicit price*)；潜变量 (*latent variable*)。

附录：Prolog 形式化

以下机器可读的配套文件 (005-hedonic-regression.pl) 在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```
%% 005-hedonic-regression.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 005
%% Title: Hedonic Regression
%% Author: Bert Craytor
%% Concept: hedonic_regression
%% 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(hedonic_regression, "Hedonic Regression", econometric_method,
%      "Hedonic regression is the estimation of the hedonic price
%      function from observed transactions in a heterogeneous-goods
%      market, yielding empirical estimates of the implicit prices of
%      the characteristics that enter the model.",
%      published, 'dictionary-2026.1').
%
% term(mass_appraisal, ...).
% term(automated_valuation_model, ...). % synonym: avm
% term(omitted_variable_bias, ...).
% term(multicollinearity, ...).
% term(residual, ...).
% term(ordinary_least_squares, ...).
%
```

```

% introduced_by(hedonic_regression, court_1939).
% introduced_by(hedonic_regression, griliches_1961).
% formalized_by(hedonic_regression, rosen_1974).
% depends_on(hedonic_regression, hedonic_price_function).
% depends_on(hedonic_regression, implicit_price).
% -----

:- use_terms([ hedonic_regression, hedonic_price_function, implicit_price,
               characteristics_space, heterogeneous_good, latent_variable,
               mass_appraisal, automated_valuation_model,
               omitted_variable_bias, multicollinearity, residual,
               ordinary_least_squares,
               sales_comparison_approach, paired_sales_analysis, adjustment ]).
:- use_relations([related/2, depends_on/2,
                  introduced_by/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_regression).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.005').

% =====
% LOCAL VOCABULARY
% =====

local_term(specification).
local_term(coefficient).
local_term(standard_error).
local_term(p_value).
local_term(r_squared).
local_term(adjusted_r_squared).
local_term(residual_standard_error).
local_term(degrees_of_freedom).
local_term(f_statistic).
local_term(significance_level).
local_term(diagnostic).

% =====
% THE EMPIRICAL BRIDGE
% =====
%
% Hedonic regression is the bridge from theoretical  $p(z)$  to a sample of
% observed  $(z_i, p_i)$  pairs.

setup_of_regression(
    each_transaction_supplies_pair("(z_i, p_i): characteristics vector and sale price
    ↪"),
    sample_from("the unobserved hedonic price function"),
    fits("a functional form to that sample, estimating parameters")).

% =====
% THREE MODELING CHOICES (per the article)
% =====

modeling_choice(which_characteristics,
    "Determines the dimensions of the estimated function. Omission causes

```

```

    ↪ omitted_variable_bias.")).
modeling_choice(functional_form,
  "Constrains what shape the estimated function can take. Linear, log-linear,
  ↪ polynomial, basis-expanded, semiparametric, ML.")).
modeling_choice(estimator,
  "OLS, GLS, robust, quantile, spatial, penalized. Determines optimization criterion
  ↪ and inference assumptions.")).

% =====
% PRACTICE CONTEXTS WHERE REGRESSION IS THE ENGINE
% =====

regression_engine_context(mass_appraisal,
  "Assessor revaluation of 100,000 parcels annually.")).
regression_engine_context(automated_valuation_model,
  "Zillow/Redfin/lender AVMs; tree ensembles are conceptually hedonic estimators.")).
regression_engine_context(single_property_appraisal,
  "Regression coefficients support adjustments more defensibly than appraiser
  ↪ intuition.")).

% =====
% DEFENSIBILITY PROFILE: REGRESSION vs PAIRED-SALES
% =====

advantage_over_paired_sales(uses_all_data,
  "Regression uses all available sales rather than two at a time.")).
advantage_over_paired_sales(simultaneous_control,
  "Isolates one characteristic while controlling for many others.")).
advantage_over_paired_sales(quantified_uncertainty,
  "Produces standard errors quantifying uncertainty in each implicit price.")).

tradeoff(functional_form_constraint,
  "Regression imposes a form that may not match the true price function.")).
tradeoff(data_requirement,
  "Requires more observations than paired-sales analysis.")).
tradeoff(sample_specification_sensitivity,
  "Results can be sensitive to the choice of sample and specification.")).

% =====
% RECURRING DEFENSIBILITY QUESTIONS
% =====

defensibility_question(sample_appropriateness,
  "The regression estimates the function for the sample's market segment and time
  ↪ period; mismatch is a defensibility failure regardless of statistical fit.")).
defensibility_question(specification_appropriateness,
  "Linear is a strong assumption; diagnostics, residual plots, and alternative
  ↪ specifications belong in the defensible workflow.")).
defensibility_question(standard_error_interpretation,
  "Standard errors capture sampling variability under the assumed specification, NOT
  ↪ misspecification, omitted-variable bias, or sample selection.")).
defensibility_question(triangulation,
  "Regression coefficients should be consistent with paired-sales, cost components,
  ↪ and local appraiser opinion; isolated coefficients are an invitation to
  ↪ investigate.")).

% =====
% WORKED EXAMPLE: lm(price ~ gla + lot + view + cond)
% =====

```

```

:- consult('../..kb/pacifica_comps').

% The fitted model.
fit_call("lm(formula = price ~ gla + lot + view + cond, data = comps)").
fit_estimator(ordinary_least_squares).
fit_software(r_lm).

% Fitted coefficients, with standard errors and significance from
% the R output reproduced in the article.
fitted_coefficient(intercept, 80777.07, 164509.35, 0.491, 0.6571, ' ').
fitted_coefficient(gla, 574.93, 109.76, 5.238, 0.0135, '*').
fitted_coefficient(lot, 22.70, 10.38, 2.187, 0.1165, ' ').
fitted_coefficient(view, 86179.03, 31342.97, 2.750, 0.0708, '.').
fitted_coefficient(cond, 49824.33, 27761.27, 1.795, 0.1706, ' ').

% Diagnostics from the R output.
residual_summary(min, -17995).
residual_summary(q1, -8243).
residual_summary(median, -5433).
residual_summary(q3, -1409).
residual_summary(max, 46455).

model_summary(residual_standard_error, 30481).
model_summary(degrees_of_freedom, 3).
model_summary(r_squared, 0.9707).
model_summary(adjusted_r_squared, 0.9317).
model_summary(f_statistic, 24.86).
model_summary(model_p_value, 0.01231).

% =====
% INTERPRETIVE CAVEATS FROM THE ARTICLE
% =====

caveat(tiny_sample,
  "n=8, four predictors, three residual df. Standard errors are correspondingly wide
  ↪; lot/view/cond fail to reach the 0.05 threshold despite economic plausibility.
  ↪Small-sample artifact.").

caveat(structural_multicollinearity,
  "Adj R^2 = 0.93 with model p = 0.012 indicates strong joint fit; weak individual
  ↪significance is a textbook multicollinearity signature. Here GLA, lot, and
  ↪condition covary in the sample.").

caveat(comp_h_residual,
  "Comp H's residual is +$46,455; the other seven fall within +/- $18,000. The model
  ↪systematically under-predicts Comp H by ~3x the next-largest residual. NOT noise.
  ↪Signal of an unobserved characteristic.").

% =====
% RELATING TO PAIRED-SALES (entry 004)
% =====
%
% The regression's implicit prices ARE the values used in entry 004's
% paired-sales decomposition of A vs B. The regression generalizes that
% logic to all 8 comps simultaneously while reporting standard errors
% that paired-sales reasoning cannot produce.

confirms_paired_sales_from(entry_004,

```

```

"A vs B differential of $127,039 from view and lot alone — coefficients are the
↪ same; regression confirms the paired analysis while adding standard errors.").

% =====
% RESIDUAL OF COMP H POINTS FORWARD TO ENTRY 006
% =====

points_forward_to(latent_variable,
  "Comp H's anomalous +$46,455 residual is the signal that an unobserved
  ↪ characteristic is contributing to price. Entry 006 takes up the question.").

% =====
% PREDICTED PRICES FROM THE FITTED MODEL
% =====
%
% Computes predicted price from the fitted coefficients. The residual
% for each comp is sale_price - predicted_price.

predicted_price(P, Predicted) :-
  coord(P, gla_sqft, GLA),
  coord(P, lot_sqft, Lot),
  coord(P, view, View),
  coord(P, cond_numeric, Cond),
  fitted_coefficient(intercept, B0, _, _, _, _),
  fitted_coefficient(gla, Bg, _, _, _, _),
  fitted_coefficient(lot, Bl, _, _, _, _),
  fitted_coefficient(view, Bv, _, _, _, _),
  fitted_coefficient(cond, Bc, _, _, _, _),
  Predicted is B0 + Bg * GLA + Bl * Lot + Bv * View + Bc * Cond.

residual_of(P, R) :-
  sale_price(P, Observed),
  predicted_price(P, Predicted),
  R is Observed - Predicted.

coord(P, A, N) :-
  attribute_value(P, A, V),
  ( number(V) -> N = V
  ; A == view, V == yes -> N = 1
  ; A == view, V == no -> N = 0
  ; V = N
  ).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).
cross_reference(latent_variable).

```

潜变量

简体中文版

基础概念 (*Foundations*) • *The Valuation Engineer*

Bert Craytor • 2026 年 8 月 3 日

版本 1.0.0

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

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

形式定义。潜变量 (latent variable) 是一项未被观测的属性，它影响观测到的结果，其取值必须由它对可观测变量的作用推断得出，而无法直接测量。

直观阐释。前五个条目在处理特征向量 $\mathbf{z}$ 时，仿佛它可以被完整测量：建筑面积、地块、景观、状况等等，全都以数据列的形式可得。而在实践中，没有任何特征向量是完整的。有些属性不在数据之中，是因为它们难以测量（室内装修品质、平面布局的流畅度、微区位的朝向细节），是因为标准数据源中根本没有它们（MLS 记录不会记载采光品质、隔音效果，也不会记载近期未取得许可的改造），或者是因为它们尚未被认定为相关（某些社区层面的特征，只有在分析揭示之后才被发现能够预测价格）。

这些未被观测却相关的属性，就是潜变量。它们存在，它们影响价格，但它们不在我们回归的数据列中。它们的作用不会因为未被测量而消失；这些作用会累积到模型的残差里，从已纳入变量的视角看它们像是“噪声”，但从底层数据生成过程的视角看，它们其实是有结构的信号。

潜变量是理想化的特征价格框架与杂乱的估价数据现实之间的桥梁。明确承认它们的存在，是诚实地思考模型不确定性、残差结构以及可观测特征解释力边界的前提。

估价师在何处遇到它。估价师在三类反复出现的情境中遇到潜变量。

品质与状况调整。作为估价类别的“品质”与“状况”，本身就是把一个多维潜在构念压缩到单一序数量表上的尝试。UAD 3.6 中的 C3 状况评级，意在概括一宗房地产的递延维修、饰面磨损、机电系统年限与呈现状态——四到十项底层属性被浓缩为一个数字。这个单一数字的概括是潜在构念的代理变量；估价师的功力之一，就在于判断该代理在每个具体案例中对底层真相的捕捉程度。

网格中未获解释的残差。当三个经调整的可比案例给出价值指示，彼此的差距却超出预期时，这一分歧本身是有信息的。它很少全部是抽样噪声；更常见的情况是，它反映了估价师尚未识别、尚未调整的某项属性。追查这种残余分歧——“为什么这个可比案例的指示值比其他高出 \$50,000?”——正是估价师对潜变量推断的非正式版本。

谁也看不见的近期改造。一宗房地产可能刚做过厨房翻新、屋顶更换，或建成了经许可的附属住宅单元 (ADU)，而这些尚未反映在 MLS 或税务评估数据中。从估价师的视角看，这项改造实地可见，却不在数据之中。从回归的视角看，这项改造是潜在的。估价师必须把实地观察与数据驱动的指示相互调和，二者都应收敛于该潜在属性的真实贡献。

为何关乎可辩护性。潜变量带来了若干可辩护性方面的考量，而条目 005 中的可加线性模型并未直接处理这些问题。

遗漏变量偏差。如果某个潜变量与一个已观测变量相关，它的效应就会渗入该已观测变量的系数。如果近期成交的帕西菲卡 (Pacifica) 住宅往往拥有翻新过的厨房（一项潜在属性），而房屋成交的合同日期又在数据之中，那么成交时点（即 `sale_age`，“距成交天数”）的系数就会吸收一部分厨房翻新效应。`sale_age` 的隐含价格贡献估计值因此被扭曲——它反映的是 `sale_age` 加上未被测量的厨房翻新状态的一部分。可辩护的估价必须承认，已观测变量的系数是以模型中缺失的内容为条件的。

残差的解读。标准回归理论把残差视为均值为零的高斯噪声。潜变量视角则把残差视为噪声与信号的混合：一部分是抽样变异，另一部分是未观测属性的累积效应。分辨二者在方法论上具有实

质意义。若某残差与位置、与成交日期、与估价师身份，或与任何其他可观测量相关，就提示相应的潜变量可以被部分还原。

模型选择。纳入更多已观测变量是遗漏变量偏差的显见对策，但最要紧的那些潜变量，往往恰恰是最难加入模型的（状况的细颗粒、近期改造的品质、社区微效应）。可辩护的应对方式有时是承认模型永远无法捕捉它们，从而把未获解释的方差归入残差，而不是强行塞进额外的变量。

估价实例演算。条目 005 的回归对可比案例 H 产生了一个异常残差：模型对可比案例 H 价格的低估达 \$46,455，而其余七个残差都落在 $\pm \$18,000$ 的区间内。可比案例 H 的已观测特征——建筑面积 1,600、地块 7,300、无景观、状况“良好”——把它牢牢置于样本的中间位置；其可观测量轮廓中没有任何东西预示异常高的价格。然而它的成交价却比模型所说的高出 \$46,455。

自然的假设是：可比案例 H 具有某项潜在属性，它对价格有贡献，却没有出现在数据中。在本期数据集的构造中，情况恰恰如此：可比案例 H 经历了一次近期厨房翻新，为其组合价值增加约 \$60,000。这项翻新不在回归的变量清单之内，因此其贡献累积在残差中，而不是体现在某个系数上。估计出的残差 (\$46,455) 略小于真实的潜在贡献 (\$60,000)，因为回归会把潜在效应的一小部分分摊到其他系数上，在相关维度上使它们向上偏高。

由此可得三点操作性结论：

把离群残差当作证据，而不是误差。对可比案例 H 的残差，第一个恰当的反应不是把它作为离群值剔除，而是去调查这宗房地产的何种特质解释了这一低估。在本例中，实地查勘或建筑许可记录检索就会揭示那次翻新。模型的失败具有诊断价值。

认识到其余系数存在轻微偏差。景观、状况与地块的系数，通过与可比案例 H 已观测特征之间的相关性，吸收了翻新效应的一小部分。在 $n = 8$ 时这一偏差很小，随着可比案例与协变量的增加还会进一步缩小，但它并不为零。

残差整体承载着可还原的结构。这正是作者所提出的残差约束方法 (Residual Constraint Approach, RCA) 的实质主张 (Craytor, 2025)：在住宅估价数据中，残差在系统上并非独立同分布的高斯噪声。它们反映了未观测属性的结构化贡献——状况的细颗粒、品质的细颗粒、功能效用以及其他潜在构念——而这些贡献可以通过把残差本身当作有待建模、而非弃置的潜变量信号来部分还原。

RCA 对状况与品质的处理，使这一两层结构得以落地。估价师的序数评级——即上文讨论的粗颗粒代理变量，可标记为状况-1（状况粗颗粒）与品质-1（品质粗颗粒）——粗糙到几乎可以把它们本身也视为潜变量，但它们仍可作为回归元纳入，以得到近似的隐含价格贡献。随后，该协议把这些估计出的贡献从销售网格回归部分的已测属性中移出、送回残差，再由残差重新分配给细颗粒潜变量状况-2（状况细颗粒）与品质-2（品质细颗粒），并对照 MLS 公开备注等补充证据加以精细化。粗颗粒评级提供了初始的价值分配；细颗粒潜变量则承载可还原的结构。最终，细颗粒的价值贡献取代粗颗粒的价值贡献。

这里的单一可比案例演示，是 RCA 直觉最简单的情形。有了更多数据与更丰富的建模，残差中的结构可以被划分为若干可辨识的潜在因子，每个因子都以标准特征价格设定无法企及的方式对组合的市场价格作出贡献。潜变量不是需要被控制掉的麻烦；它们是可辩护估价建模的下一个前沿。

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

交叉引用：特征空间 (*characteristics space*)；特征价格函数 (*hedonic price function*)；隐含价格 (*implicit price*)；特征价格回归 (*hedonic regression*)。

附录：Prolog 形式化

以下机器可读的配套文件（006-latent-variable.pl）在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```
%% 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).

```

近期判例

近期法院判例——美国

近期法院判例 • *The Valuation Engineer*

William Bert Craytor • 第 1 卷, 第 1 期 • 2026 年夏季

DOI: 10.67330/fz34f711

简体中文译本, 由人工智能辅助翻译并经编辑部审校; 英文版为权威版本。

按辖区分列的案例摘要

California (加利福尼亚州)

The Retail Property Trust v. Orange County Assessment Appeals Board No. 1. California Court of Appeal (加利福尼亚州上诉法院), 2026 年 4 月判决。

上诉法院处理的问题是: 对于 COVID-19 疫情期间政府命令所施加的通行限制, 能否依据 California Revenue and Taxation Code section 170 (《加利福尼亚州税收与税务法典》第 170 条) 获得财产税估价减免。作为原告的购物中心主张, 疫情期间通行受限构成该第 170 条意义上的对财产的“损害”(damage), 因而有权获得估价减免。

对估价师的意义。该判决意见对第 170 条中“损害”一词的解释, 将影响今后基于外因造成的财产效用减损而提出的减免主张, 无论其起因是公共卫生命令、基础设施工程, 还是其他影响通行的政府行为。凡受托就通行限制对商业地产价值的影响发表意见的估价师, 都应研读法院关于“效用受限”——区别于实体损伤——是否构成可获估价减免之损害的说理。

Colorado (科罗拉多州)

Cheroutes v. Rocket Mortgage & Solidifi. U.S. District Court, D. Colorado (美国科罗拉多地区联邦地区法院); 审理中, 2026 年有进展更新。

该案起诉状主张, 2021 年对 Denver (丹佛) 一处双拼住宅所作的估价存在违反 Fair Housing Act (《公平住房法》) 的估价歧视: 该房产被估值约 \$640,000, 比上一年度的估价值低出逾 \$220,000。此前一名联邦法官已驳回被告的驳回起诉动议, 裁定 Rocket Mortgage 与估价管理公司 Solidifi 必须就相关主张进行答辩, 而不能以贷款机构和估价管理公司 (AMC) 无须为独立计费估价师的行为承担责任为由推卸责任。该案在 2026 年处于证据开示阶段。

对估价师的意义。该案是目前检验估价管理公司与贷款机构是否须依据《公平住房法》就计费估价师行为承担责任的主要活跃载体。若法院最终就实体问题判定该估价管理公司与贷款机构承担责任, 将大幅扩大因估价而面临公平住房主张的主体范围, 并很可能传导至估价管理公司与估价师之间的委托合同之中。

Florida (佛罗里达州)

针对向住宅借款人收取的 AMC 费用的佛罗里达州集体诉讼。 2026 年 1 月 2 日提起。

该起诉状对估价管理公司向住宅按揭借款人收取的费用提出质疑, 主张支付给计费估价师的费用与在交割时向借款人收取的费用之间的加价, 超出了联邦法律确立的限额, 且披露不充分。该案处于 Dodd-Frank (《多德—弗兰克法》) “惯常且合理”费用要求 (12 U.S.C. § 1639e) 与 Truth in Lending

Act (《贷款真实性法》)、Real Estate Settlement Procedures Act (《房地产交割程序法》) 项下消费者披露制度的交叉地带。

对估价师的意义。该诉讼检验的是估价管理公司报酬结构；若获准作为集体诉讼并就实体问题作出裁判，可能重塑全行业的披露实务。其结果可能通过将“估价管理公司支付额”与“向借款人收取额”之间的价差置于司法审视之下，间接影响经由估价管理公司承接业务的计费估价师的费用形成环境。

Georgia (佐治亚州)

Green Rock 保护地役权系列案。 United States Tax Court (美国联邦税务法院)，判决意见见 *Tax Notes* 2026 年 3 月 2 日的评述。

税务法院在评估佐治亚州 Piedmont (皮埃蒙特) 地区相邻地块 (Green Rock Property 与 Green Rock South) 上设定的保护地役权时，逐一检视了最高最佳使用 (highest and best use) 检验的各项要件。申请人一方的专家断言，在设定地役权之前，该等房地产的最高最佳使用是经营花岗岩采石场，并采用现金流折现分析估算出设定地役权前的公平市场价值约为 \$13.75 million，而设定地役权后的价值为 \$180,000。

法院对该检验的每一项要件都表示保留。在法律许可性方面，法院不认为将土地改划为允许采石作业的分区在合理概率上可以实现，并特别指出：Green Rock Property 的分区类别变更是在委托新的保护地役权估价后不久取得的——这一先后顺序令法院难以采信。在实体可能性方面，法院认为申请人的地质证据不足以证明在 Harman Road 地块上可以开发出可行的采石场，尽管 Piedmont 地区总体上已知存在可开采的花岗岩。法院进一步认为，即便假定法律许可性与实体可能性均已成立，在未满足该分析其余要件的情况下，这也不具有决定意义。

对估价师的意义。该判决意见是有序、逐项适用四要件最高最佳使用检验的一个范例。它强化了 *Ranch Springs, LLC v. Commissioner*, 164 T.C. No. 6 (2025) 与 *Beaverdam Creek Holdings, LLC v. Commissioner*, T.C. Memo. 2025-53 所确立的立场：即便拟议用途在周边区域的总体可行性有据可查，税务法院也不会接受带有投机性质的最高最佳使用主张。编制地役权估价的估价师应针对每一项要件建立针对具体地块的证据，而不应依赖区域性的或类推支持的可行性论证；对于时间上的先后模式——例如与委托估价同期取得的分区变更——尤应格外审慎，因为这类情形容易招致司法上的怀疑。

Maryland (马里兰州)

Baltimore (巴尔的摩市) 征收 Pimlico Race Course (皮姆利科赛马场)。 土地征收程序，2025–2026 年进行中。

巴尔的摩市依据征收权取得 Pimlico Race Course，由此引出关于专用型设施的持续经营价值与房地产价值之间关系的公正补偿问题。由于 Pimlico 的经营性业务与其设施在功能上不可分割，常规的房地产可比销售分析用处有限，而收益法估价则高度依赖于赛马经营将持续进行这一假设。

对估价师的意义。凡受托评估经营性业务与不动产在功能上不可分割的单一用途设施——例如赛马场、影视制片厂、主题公园、码头，以及某些医疗或酒店类资产——的估价师，都值得关注该程序的进展。诉讼程序中形成的方法论，很可能影响今后征收案件中公正补偿计算如何处理持续经营增值部分。

针对估价师的联邦《公平住房法》主张

Braxton v. Bank of America. 2026 年 4 月提起。

一名房主起诉 Bank of America（美国银行）及一名 Virginia（弗吉尼亚州）执业的估价师，主张其行为构成违反 Fair Housing Act（《公平住房法》）的种族歧视。原告声称，估价现场有她本人及其租户在场，且该估价师的举止表现出种族敌意。该案延续了此前估价歧视诉讼所形成的模式，包括 Maryland 联邦地区法院的 *Connolly/Mott v. loanDepot* 案，以及加利福尼亚州 Marin County（马林县）的 *Austin v. Miller* 案，即由个人原告针对估价师、估价管理公司与贷款机构提起《公平住房法》主张。

对估价师的意义。针对个体估价师的估价歧视案件持续以一定速度提起，加之联邦层面在差别性影响执法上的收缩，表明责任暴露的格局已经改变：私人原告与州一级的公平住房执法机构，如今已取代联邦机构成为估价歧视责任的主要来源。估价师应检视其文档记录做法，确保可比实例选取、调整值推导与最终价值调和在事后审视下均能独立成立；同时应意识到，勘察现场本身的言辞或举止，往往正是原告主张所依据的事实基础。

近期法院判例——欧盟

近期法院判例 • *The Valuation Engineer*

William Bert Craytor • 第 1 卷, 第 1 期 • 2026 年夏季

DOI: 10.67330/r1yn1y02

简体中文译本, 由人工智能辅助翻译并经编辑部审校; 英文版为权威版本。

范围与资料来源

本综述取材于可获取的公开一手资料——最主要包括欧盟法院 (Court of Justice of the European Union) 的官方判例数据库 (curia.europa.eu)、欧洲人权法院 (European Court of Human Rights) 的 HUDOC 数据库、德国联邦财政法院 (Bundesfinanzhof) 公布的裁判 (bundesfinanzhof.de)、德国联邦宪法法院 (bverfg.de)、用于法国判例的 Légifrance, 以及英国的 Upper Tribunal (Lands Chamber) (上级审裁处土地庭)——并参考了专业与学术出版物中的二手评述。以各国语言作出的裁判均已阅读原文并以英文加以描述; 案件名称保留其原有拼写与变音符号。

仅限订阅的法律检索数据库 (juris.de、beck-online、Westlaw UK、Lamy、Aranzadi) 未予使用。最高法院层级以下的国内法院裁判仅作选择性收录, 优先选取具有明确先例价值或方法论意义者。本文所载案件引证, 在用于实务依据或进一步出版之前, 均应对照法院的官方裁判汇编予以核实。

本基线一辑将 Germany (德国)、France (法国) 与 United Kingdom (英国) 作为主要国家辖区。Italy (意大利)、Spain (西班牙)、Netherlands (荷兰) 与北欧各国在本辑中未单独综述, 但将视材料情况在其后的季度各辑中补入。Switzerland (瑞士) 虽不属于欧盟, 但因瑞士估价准则对欧洲实务总体上具有很强的方法论影响, 将在今后各辑中予以综述。

欧洲监管背景

欧盟层面的两项监管进展对估价师的执业环境有实质影响, 尽管两者本身都不是法院判例:

EU AI Act 与自动估价模型。 Regulation (EU) 2024/1689 (EU AI Act, 欧盟《人工智能法》) 已于 2024 年生效, 适用于高风险人工智能系统的条款分阶段生效, 至 2026 年 8 月全部到位。2026 年 5 月 19 日, European Commission (欧盟委员会) 就高风险人工智能系统的分类发布了指南草案, 供利益相关方咨询。该法 Annex III (附件三) 将与自然人有关的信用状况评估列入八类推定高风险应用场景之中, 委员会的指南草案对这一类别作了详细展开。由于用于按揭发放、按揭组合监控与保险承保的自动估价模型 (AVM) 在功能上与相关贷款的信用状况评估不可分割, 欧盟境内由信贷机构或保险机构运营的 AVM, 在大多数配置下都将被归类为高风险人工智能系统。由此产生的合规后果包括: 强制性的风险管理体系、数据治理控制、日志留存不少于六个月的自动记录、使运营者能够理解并推翻系统决定的人工监督机制, 以及投放市场后的监测义务。

对估价师的意义。欧洲估价师应当预期, 其贷款机构客户所运营的、由 AVM 支持的工作流程, 将面临大幅提高的文档记录与可解释性要求。凡估价师提供模型输入、校准数据, 或对 AVM 输出进行人工监督复核之处, 这些活动本身即构成贷款机构高风险人工智能合规态势的组成部分。

Energy Performance of Buildings Directive (重订版)。 Directive (EU) 2024/1275, 即重订版《建筑能效指令》(EPBD), 于 2024 年 5 月 8 日在《欧盟官方公报》上公布, 并于 2024 年 5 月 28 日生效。成员国的转化期限为 2026 年 5 月 29 日。除其他条款外, 重订版 EPBD 要求成员

国优先改造能效表现最差的百分之十六的非住宅建筑存量与能效表现最差的百分之四十三的住宅存量；设定最低能效标准（MEPS）；要求自 2028 年起新建公共建筑、自 2030 年起所有新建建筑均须为零排放建筑；并要求自 2030 年起对新建建筑开展全生命周期碳评估。

对估价师的意义。重订版 EPBD 将通过若干渠道影响房地产估价：最低能效门槛造就了一类新的、由合规驱动的功能性贬值；能效证书（EPC）评级成为住宅与商业存量房产价格的实质决定因素；为达到合规所需的改造成本，必须在成本法以及剩余经济寿命的扣减中予以反映。估价师还应预见到，受 Mortgage Credit Directive（《住房抵押贷款指令》）约束的贷款机构，将日益要求在抵押物估价中对 EPC 评级与 MEPS 相关的资本性支出要求作出明确处理。

后续各辑说明

继本基线一辑之后的季度各辑将聚焦于变化：新的裁判、此处所列待决事项的处理结果（尤其是德国联邦宪法法院对 *1 BvR 472/26* 的处理），以及成员国为将重订版 EPBD 转化为国内法而制定的实施立法。瑞士法律框架、意大利与西班牙最高法院裁判的处理，以及法国 *Conseil constitutionnel*（宪法委员会）项下的进展，将视材料情况补入。

按辖区分列的案例摘要

Court of Justice of the European Union（欧盟法院）

***GR Real* (Case C-351/23)。** 欧盟法院 2025 年 6 月 24 日判决。

欧盟法院就一件来自斯洛伐克法院的先决裁判请求作出裁判。该案源于一项庭外按揭强制执行程序：在质疑合同条款可能不公平的诉讼尚在进行期间，消费者已因拍卖而失去其家庭住宅。法院裁定，Directive 93/13/EEC（第 93/13/EEC 号指令）项下的消费者保护保障要求，国内程序规则必须允许在设有抵押的住宅经庭外强制执行出售之前，对合同公平性进行有效的司法审查；凡允许出售在该审查完成之前即不可逆地推进的程序，均与该指令不相容。

对估价师的意义。该裁判并非直接的估价案件，但它重塑了按揭抵押物估价被委托与使用的程序环境。在若干设有庭外强制执行程序的成员国（尤其是斯洛伐克、捷克、匈牙利与波罗的海三国），其实际效果是放缓强制执行拍卖流程，并延长贷款机构抵押物估价被依赖的期间。在这些辖区为强制执行情形编制抵押物估价的估价师，应当预期估价报告的使用期限可能长于以往，并可能须面对原估价基准日与最终出售之间的市场变动的检验。

European Court of Human Rights（欧洲人权法院）

Baku (巴库) 建筑物拆除案 (Azerbaijan, 阿塞拜疆), 2025 年 7 月判决。 欧洲人权法院,《第一议定书》第 1 条。

欧洲人权法院就巴库一栋建筑物的拆除与征收认定存在侵犯和平享有财产权的行为；该拆除是在当局援引地震风险后下令实施的。斯特拉斯堡法院认定，国内当局在确定补偿时，未能就其驳回申请人所提交的估价报告给出充分理由，并且不正当地仅采纳了法院自行委托的专家意见所载金额。法院共判给 €75,000 的财产损害赔偿、€3,000 的非财产损害赔偿，以及 €1,055 的法律费用。

对估价师的意义。该裁判强化了《第一议定书》第 1 条的一条判例线索，即要求国内法院在征收程序中对相互竞争的估价证据给予有理由的斟酌，并且在采信法院委托的专家意见而否定申请人的估价报告时，这一否定本身也必须说明理由。对于从事征收与公正补偿业务的欧洲估价师而言，本案凸显了一点：当估价报告被提交以支持针对国家的主张时，报告在文档上的可辩护性——支撑可比实例选取、调整值推导与最终价值调和的明示说理——本身即是一项基本权利事项。

France (法国)

Cour de cassation (法国最高法院), Chambre commerciale (商事庭), 2024 年 6 月 19 日, n° 22-24.169。 (专业评述持续讨论至 2025 年。)

最高法院在继承税 *valeur vénale* (市场价值) 争议的语境下裁定: 继承取得的房产在应税事件 (*de cujus*, 即被继承人死亡) 之后发生的出售, 不得被继承人用于向下修正其原先申报的 *valeur vénale*, 即便该事件后的成交价格显著更低、且出售发生在死亡后数月之内亦然。该裁判适用的原则是: 就继承税而言, *valeur vénale* 以应税事件之日为准确定, 不因该房产其后的交易而予以更正。

对估价师的意义。该裁判确认了法国继承税实务中严格的估价基准日纪律, 并限制了事件后成交作为价值追溯证据的相关性。编制法国继承税估价的估价师, 应基于估价基准日当日或之前的交易建立完备的可比销售证据, 且不应将标的房产本身在事件后的成交作为佐证证据加以呈现。同样的说理很可能延伸适用于赠与税 (*droits de donation*) 与财富税 (*IFI*) 情形。

Cour de cassation, 3^e Chambre civile (第三民事庭), 2026 年 1 月 8 日, n° 23-22.803。
协商性 (非司法) 专家报告。

最高法院第三民事庭作出裁判, 重新界定了协商性 (非司法) 专家报告在法国民事诉讼中的法律价值。该裁判承接 2025 年 7 月 18 日 *décret* (法令) 所推动的更广泛程序改革, 该法令使调解与协商性争议解决成为民事程序中的默认规则。法院澄清: 由单方当事人依合同委托编制的协商性专家报告, 仍可作为证据被采纳并被赋予证明力, 但须遵循对抗原则, 并须经法院对专家独立性与方法论的评估。

对估价师的意义。该裁判强化了在法院指定司法专家 (*experts judiciaires*) 框架之外编制的 *expertises immobilières* (不动产鉴定报告) 的证据地位, 前提是报告遵循本行业的方法论标准且专家的独立性有据可查。对于承办继承、离婚或商事争议中争议性估价的法国估价师而言, 该裁判为客户委托的报告在其后诉讼中被援引与依赖打下了更坚实的基础。相应地, 格外注意记录独立性 (不存在从属关系、经济利益或与客户的亲属关系) 并适用专业标准所规定的方法, 也就变得更加重要。

Germany (德国)

Bundesfinanzhof (德国联邦财政法院), 2025 年 11 月 12 日判决 (II R 25/24, II R 31/24, II R 3/25)。 *Grundsteuer Bundesmodell* 三部曲——改革后财产税估价制度的合宪性。德国联邦财政法院 (*Bundesfinanzhof*, BFH) 第二审判庭在三件平行的上告程序中裁定: 自 2025 年 1 月 1 日起施行的改革后不动产税 (*Grundsteuer*) 中, 德国十一个联邦州依所谓 *Bundesmodell* (联邦模式) 对住宅房产估价所适用的收益资本化估价方法 (*Ertragswertverfahren*) 的相关规定, 符合宪法。原告为 North Rhine-Westphalia (北莱茵-威斯特法伦州)、Saxony (萨克森州) 与 Berlin (柏林) 的公寓所有权人, 他们以该套规整化的估价规则违反一般平等原则 (《基本法》第 3 条第 1 款) 并产生系统性失准的价值为由, 对其税额认定提出质疑。

联邦财政法院认为, 立法者所建立的估价制度在概念上以市场价值为取向 (*konzeptionell einer Verkehrswertorientierung folgt*), 其设计意图是在所有被估价对象上平均而言逼近“客观现实的房产价值”, 并处于公平市场价值 (*gemeiner Wert*) 的一个区间之内。法院认定, 立法者在权衡该估价方案的目标与此类方案必然伴随的不平等之间, 并未超出其被允许的裁量余地; 并且, 为避免通过自动化定期更新而再次形成“估价积压”而赋予的重大权重, 在宪法上是可以允许的。联邦财政法院还强调, 纳税人仍保有由该院此前 2024 年 5 月 27 日的 *Beschlüsse* (裁定) (II B 78/23 与 II B 79/23) 所确立的选择权, 即依据 § 198 BewG (《估价法》第 198 条) 通过合格估价证明较低的公平市场价值 (*gemeiner Wert*), 但受此限制: 该较低价值须由公开委任的房地产估价师运用公认估价方法出具的估价报告予以证实。

由 *Bund der Steuerzahler Deutschland* (德国纳税人协会) 与 *Haus & Grund Deutschland* (德国房地产业主协会) 支持的一项宪法诉愿 (*Verfassungsbeschwerde*), 现以案号 1 BvR 472/26 待联邦宪法法院 (*Bundesverfassungsgericht*) 审理。

对估价师的意义。这一三部曲是本综述期间最重要的德国房地产估价判例。对估价师而言, 主要的实务后果是对 § 198 BewG“较低价值”估价 (*Verkehrswertgutachten zum Nachweis des niedrigeren gemeinen Werts*) 的需求上升——即在依 *Bundesmodell* 确定的 *Grundsteuerwert* (不动产税价值) 看似超出标的房产实际市场价值的情形下。联邦财政法院对这一救济途径的明确认可, 加之其对 *Bundesmodell* 的总体尊重, 意味着估价报告如今已成为纳税人个体从系统性高估中获得救济的实际运作机制。估价师应预期此类报告的委托将持续增加, 并确保符合 ImmoWertV 2021 (《不动产估价条例》2021) 的方法论与 § 198 BewG 的要求。待决于联邦宪法法院的宪法诉愿 (1 BvR 472/26) 对该框架是否会有进一步变动具有实质意义, 值得持续关注, 预计将成为下一季度一辑的主要议题之一。

Bundesfinanzhof, 2025 年 10 月 7 日判决, IX R 26/24。土地与建筑物之间的购价分摊 (*Kaufpreisaufteilung*)。在一件涉及住宅出租房产折旧 (*Absetzung für Abnutzung, AfA*) 计提基数的所得税案件中, 联邦财政法院第九审判庭处理了如何在不可折旧的土地部分 (*Grund und Boden*) 与可折旧的建筑物部分 (*Gebäude*) 之间分摊合同购价的方法问题。原告曾以 *Bodenrichtwerte* (基准地价) 所依据的数据基础就标的房产的具体区位并无可靠记载为由, 对下级法院对其的援用提出质疑。

对估价师的意义。该裁判延续了联邦财政法院的一条判例线索 (尤其是上文所引 2025 年 11 月 12 日的 BFH II R 25/24, 以及此前 2022 年 9 月 20 日的 BFH IX R 12/21), 确认 ImmoWertV 是购价分摊所公认的估价原则来源。本案还表明了 *Bodenrichtwerte* 所依据的数据基础的重要性: 凡 *Gutachterausschüsse* (专家估价委员会) 就具体区位并无足够可比销售数据者, *Bodenrichtwert* 即可能因数据质量问题而受到质疑。对于编制购价分摊或折旧基数意见的估价师而言, 这意味着对 *Bodenrichtwert* 数据基础的记录——包括其底层的 *Kaufpreissammlung* (成交价格汇编)——成为可辩护分析的组成部分。

United Kingdom (英国)

Mehrunnisa Alikhan Ahmed Khan v Wolverhampton City Council. Upper Tribunal (Lands Chamber) (上级审裁处土地庭), 2025 年。

上级审裁处土地庭确认, 评估强制收购补偿的正确估价基准日, 是该房产依据一般归属声明 (*general vesting declaration, GVD*) 归属于收购主管机关之日。审裁处还处理了同期估价与追溯估价之间的关系: 在本案中, 地方主管机关在 GVD 之后随即作出的初次估价为 £150,000, 而 2025 年在审裁程序中编制的第二份估价——以业主发出通知后三个月的假定归属日为基准——则为 £120,000。尽管审裁处确认了法律上的估价基准日, 它仍偏向采用在时间上更接近 GVD 作出之日所计算的数额, 从而使请求人获得较高的支付额。

对估价师的意义。该裁判对从事强制收购业务的英国估价师具有方法论上的启示意义。虽然法律上的估价基准日是归属之日, 但审裁处在同一法律估价基准日之下, 偏向同期估价而非数年之后编制的追溯估价, 这表明同期市场证据即便在归属之日形式上不可采, 实际上仍具有分量。估价师应确保在 GVD 前后编制的估价报告有充分的支撑证据作为记录, 以充当主要记录, 并预见到其后的审裁程序可能优先采信同期报告而非事后的追溯重构。

近期法院判例——拉丁美洲

近期法院判例 • *The Valuation Engineer*

William Bert Craytor • 第 1 卷, 第 1 期 • 2026 年夏季

DOI: 10.67330/ywgzhg47

简体中文译本, 由人工智能辅助翻译并经编辑部审校; 英文版为权威版本。

范围与资料来源

本综述取材于可获取的公开一手资料——最主要包括美洲人权法院 (Inter-American Court of Human Rights) 判例库 (corteidh.or.cr)、巴西联邦最高法院 (STF) 与高等法院 (STJ) 门户网站 (portal.stf.jus.br、stj.jus.br)、墨西哥 *Semanario Judicial de la Federación* (联邦司法周刊, sjfsemanal.scjn.gob.mx)、阿根廷 *Corte Suprema de Justicia de la Nación* (国家最高法院, csjn.gov.ar)、智利 *Poder Judicial* (司法机关) 门户网站 (poderjudicial.cl) 与 *Diario Constitucional*, 以及哥伦比亚 *Corte Constitucional* (宪法法院, corteconstitucional.gov.co)——并参考了专业与学术出版物中的二手评述。以西班牙语与葡萄牙语作出的裁判均已阅读原文并以英文加以描述; 案件名称保留其原有拼写与变音符号。

仅限订阅的法律检索数据库 (vLex、La Ley、Thomson Reuters Proview Latin America) 未予使用。最高法院层级以下的国内法院裁判仅作选择性收录, 优先选取具有明确先例价值或方法论意义者。本文所载案件引证, 在用于实务依据或进一步出版之前, 均应对照法院的官方裁判汇编予以核实。

本基线一辑将 Argentina (阿根廷)、Brazil (巴西)、Chile (智利)、Colombia (哥伦比亚) 与 Mexico (墨西哥) 作为主要国家辖区。Peru (秘鲁)、Venezuela (委内瑞拉)、Ecuador (厄瓜多尔)、Bolivia (玻利维亚)、Paraguay (巴拉圭)、Uruguay (乌拉圭) 以及中美洲与加勒比各辖区在本辑中未单独综述, 但将视材料情况在其后的季度各辑中补入。委内瑞拉的财产返还业务与古巴制裁相关的估价业务, 因其方法论上的独特性及其与跨境实务的相关性, 将在今后一辑中予以综述。

区域背景

与欧盟综述不同, 本拉丁美洲综述没有与 EU AI Act (欧盟《人工智能法》) 和重订版 Energy Performance of Buildings Directive (《建筑能效指令》) 所代表的超国家监管层相对应的直接类比。拉丁美洲的房地产估价监管绝大部分属于国家层面; 主要的超国家影响来自美洲人权法院依据 American Convention (《美洲人权公约》) 第 21 条 (财产权) 所形成的基本权利判例法, 以及本区域多数辖区为财务报告目的采纳国际财务报告准则 (特别是 IFRS 13《公允价值计量》)。尽管如此, 仍有两条方法论主线贯穿本区域判例, 值得在此提及:

原住民共有财产的估价。拉丁美洲估价实务中最具特色的方法论挑战, 是在返还、征收与补偿情形下对原住民与非洲裔族群共有财产的估价。美洲人权法院的判例法一贯认为, 原住民共有财产与个人私有财产在宪法上受到同等保护; 但此类财产的估价方法——通常并无可比的公平交易案例, 财产由共同体共有且不可转让, 且该领土的文化与精神意义构成其价值的内在组成部分——仍不成熟。截至 2025 年底, 美洲人权法院作出的有利于原住民族的二十四项判决仍未执行或仅部分执行, 其中若干涉及尚待解决的补偿计算, 而这些计算取决于悬而未决的估价方法。

征收补偿中的通胀、货币与指数化。本区域若干辖区，尤其是阿根廷，要求估价方法与赔偿计算必须应对高通胀、以外币计价的 *tasaciones*（估价）以及复杂的指数化制度。阿根廷最高法院认为，原以外币表示的征收赔偿不得按人为的平价汇率折算为本币，并且估价必须反映房产在赔偿之时的真实经济价值。类似关切亦见于委内瑞拉（与财产返还及制裁相关的估价有关），并在巴西通胀高企时期不时出现。对于跨区域各辖区执业的估价师而言，这些通胀与货币方面的考量是估价方法的内在组成部分，其程度是欧洲与美国所没有的。

后续各辑说明

继本基线一辑之后的季度各辑将聚焦于变化：美洲人权法院的新裁判、此处所列待决事项的处理结果、哥伦比亚 Jurisdicción Agraria y Rural（农业与农村管辖法院系统）迈向 2027 年投入运作的实施进展，以及各国在 IFRS 13 公允价值计量实务方面的动向。秘鲁、委内瑞拉、厄瓜多尔与中美洲各辖区将视材料情况补入。将特别关注原住民共有财产估价的方法论发展——在该领域，美洲人权法院的实体裁判已超前于估价师与补偿主管机关所能获得的实用估价方法。

按辖区分列的案例摘要

Inter-American Court of Human Rights（美洲人权法院）

Maya Q'eqchi' Agua Caliente vs. Guatemala 与 *Comunidad Garífuna de San Juan y sus Miembros vs. Honduras*。美洲人权法院 2024 年判决，区域判例评述持续至 2025 年。

在两项同期判决中，美洲人权法院就危地马拉与洪都拉斯在划界、确权以及防止原住民与非洲裔族群祖传领土被侵夺方面的义务作出裁判。*Maya Q'eqchi' Agua Caliente* 社区历经四十五年的行政与司法行动，方才取得其领土的确定性所有权证书；洪都拉斯的 *San Juan* 加里富纳（*Garífuna*）社区则面临祖传沿海土地持续被侵夺的局面。两项判决均涉及国家依《美洲人权公约》第 21 条（财产权）承认、划界并保护原住民共有财产的义务。区域评论者注意到，美洲人权法院在这些裁判中的返还框架，较之其早先判决略趋收紧，尤其体现在其对“善意第三人”（*terceros inocentes*）——即在不知悉原住民主张的情况下取得重叠权利者——的处理上。

对估价师的意义。这些裁判重申：原住民与非洲裔族群的共有财产在整个美洲人权体系内受到宪法保护；在返还或补偿情形下对此类财产的估价，若不辅以方法论上的补充，便不能依赖常规的公平交易可比销售（此类案例通常并不存在）。在原住民返还情形下受托的估价师应当预期，估价必须整合三个不同的组成部分：本区域内物理特征可比的土地的基础市场价值；可归因于原住民使用的机会成本与改良；以及该领土的非经济性文化与精神意义——美洲人权法院判例法将其视为财产权中可予补偿的一个维度。第三部分的方法仍不成熟，是美洲人权法院说理中持续探讨的主题。

Mapuche / Consejo de Todas las Tierras vs. Chile。美洲人权法院 2024 年判决（见 2025 年判例评述）。

美洲人权法院就智利国家针对马普切（*Mapuche*）人 135 名成员的人权侵权行为作出谴责，该案涉及 *Consejo de Todas las Tierras*（1989–1992 年）的领土索还行动，以及将反恐立法适用于原住民领土主张的做法。该判决扩展了 *Norín Catrimán and others vs. Chile*（2014 年）所确立的原则，并明确承认原住民族在领土主张方面的自决权。

对估价师的意义。尽管本案的主要裁判要旨涉及刑事程序保障而非直接的估价方法，但对原住民在领土方面自决权的承认，形成了一层程序性叠加，影响智利原住民领土返还及相关情形下的估价工作。在此类情形下受托的估价师应确保其方法纳入适当的协商程序，并承认该共同体对其领土价值如何被表述所享有的自决权。

Argentina (阿根廷)

阿根廷关于 **Tribunal de Tasaciones de la Nación** (国家估价法庭) 的原则。 经 2025 年判例整合确立。

阿根廷国家最高法院 (CSJN) 与下级联邦法庭持续发展关于 **Tribunal de Tasaciones de la Nación** (TTN, 国家估价法庭) 在征收与 *expropiación inversa* (逆向征收) 事项中所作估价的原则。既定的 CSJN 原则 (Fallos 313:82, 在 2024–2025 年间反复适用) 认为, 基于该法庭成员的技术能力、其估价所依据的证据基础以及实务的统一性, TTN 的估价具有推定的分量; 欲背离 TTN 的估价, 须证明存在明显错误或遗漏。2024–2025 年间适用该原则的联邦法庭——包括 **Juzgado Federal de Oberá** (奥韦拉联邦法院) 在 *Sosa v. Entidad Binacional Yacyretá* (2024 年 6 月, 因 Yacyretá 水电站水库引发的 *expropiación inversa*) 一案中——继续以 TTN 估价作为赔偿的实际运作基准。

对估价师的意义。对阿根廷估价师而言, 实际后果是: 为对抗 TTN 估价而提交的私人估价报告面临沉重的举证负担。要推翻或变更 TTN 的估价, 私人报告必须证明 TTN 工作中存在方法上的错误或遗漏, 而不能仅仅提出另一套估价结论。编制此类报告的估价师, 应将方法论批评聚焦于 TTN 的可比实例选取、调整值推导以及对当地市场状况的处理, 而不应仅仅着眼于价值差额的大小。

以外币计价的 **tasaciones** 与赔偿平价。 CSJN 关于原以外币表示的征收赔偿不得按人为的一比一平价折算为阿根廷比索的原则, 至 2025 年仍然有效。该原则保护被征收人免受因通胀或货币政策而无法反映房产在赔偿之时真实经济价值的赔偿。这一原则在 2024–2025 年间已就长期进行的征收程序被多次重申。

对估价师的意义。阿根廷估价师在为跨越显著通胀或货币政策变动期间的征收事项编制估价时, 应明确记录估价所依据的货币基准, 并应考虑在估价基准日与预期支付日之间采用适当的指数化方法。未能处理通胀与货币风险的报告, 不大可能说服适用 CSJN 赔偿原则的法庭。

Brazil (巴西)

Supremo Tribunal Federal, RE 922.144 (Tema 865, 一般影响力主题)。 巴西联邦最高法院 2024 年判决, 2025 年持续适用。

STF 裁定, *desapropriação* (征收) 程序中的赔偿与《联邦宪法》第 100 条所确立的联邦 *precatório* (司法债务令) 制度在宪法上相容。具体而言, 公共主管机关为 *imissão provisória na posse* (临时占有) 目的最初提存的金额, 与最终由司法鉴定确定的金额之间的差额, 必须通过 *precatório* 制度支付, 而非作为即时的补充提存。该裁判在《宪法》第 5 条第 24 项 *justa, prévia e em dinheiro* (公正、事先、以货币支付) 的补偿命令, 与 *precatório* 制度的有序支付要求之间取得平衡。

对估价师的意义。该裁判对受征收主管机关委托的估价师所编制的初始提存估价具有直接的实务后果。由于初始提存中的任何低估都将在其后通过 *precatório* 制度支付 (这可能涉及大幅延迟), 原告在结构上就有动机主张初始估价被低估。编制初始提存估价的估价师应格外审慎地记录其方法, 以便在其后的司法鉴定面前为该估价辩护, 并预见到初始提存与最终司法认定之间的差额, 将依 STF 判例法与 **Decreto-Lei 3.365/1941** (第 3.365/1941 号法令) 框架计收补偿性利息 (*juros compensatórios*)。

Superior Tribunal de Justiça, REsp 2.129.162 (重复性主题)。 巴西高等法院 2025 年 5 月判决。

STJ 就征收主管机关撤回 *desapropriação* 之诉时的律师费问题, 解决了一项重复性争议。法院确立: 撤诉案件中的费用应遵循 **Decreto-Lei 3.365/1941** 第 27 条第 1 款的框架, 而非 *Código de*

Processo Civil (《民事诉讼法典》) 的一般程序规则。该案源于 Cemig (Companhia Energética de Minas Gerais, 米纳斯吉拉斯电力公司) 为一条配电线路寻求行政地役权而提起的诉讼。

对估价师的意义。该裁判属程序性而非方法论性质, 但对行政地役权与 *desapropriação* 事项中估价师受托关系的构造有所影响。在此类程序中被聘为专家证人 (*peritos judiciais*) 的估价师, 应确保其费用安排的记录援引适用的法令框架, 而非一般程序规则。

Superior Tribunal de Justiça 关于 *juros compensatórios* 的原则。 2025 年持续适用。

STJ 维持其原则: *desapropriação* 事项中的补偿性利息 (*juros compensatórios*) 仅自法院就被征收财产的权属作出裁判之后起算。该利息适用于初始提存与最终赔偿价值之间的差额, 按历史上确立的年利率百分之十二计算 (STF Súmula 618, 第 618 号判例要旨), 其作用是补偿被征收人在最终赔偿支付之前丧失占有的损失。

对估价师的意义。巴西估价师应当预期, 他们在司法鉴定中所认定的估价差额——即公共主管机关的初始提存与评估市场价值之间的差额——将成为补偿性利息计算的基数。因此, 该差额的经济意义大于其名义差额所显示的程度, 在长期程序中尤其如此。

Chile (智利)

Corte Suprema, Rol N° 31.423-2025。 智利最高法院 2025 年 9 月判决。

智利最高法院驳回了一家房地产开发商 (*inmobiliaria*) 针对国库 (*Fisco de Chile*) 提起的撤销之诉上诉; 该案属依 Decree-Law N° 2.186 (第 2.186 号法令) 第 12 条进行的特别征收程序。原告曾主张, 专家委员会 (*Comisión de Peritos*) 所定的估价未能反映该房产的最高最佳使用 (*mayor y mejor uso*) 以及既有建筑物与配套工程的价值, 并另以既有租赁合同本可产生未来租金收入为由, 请求就可得利益损失 (*lucro cesante*) 获得赔偿。法院认为, 原告的主张实质上是对下级法院证据权衡的不同意见, 而非对规范估价的法律规范的真正违反; 而下级法院的说理——其依据专家委员会的证据并结合书证与证人证言——已有充分根据。

对估价师的意义。该裁判说明了智利最高法院原则对征收程序中专家委员会估价所给予的高度尊重, 以及仅以方法论分歧为由对估价结论进行撤销审查的有限空间。在智利征收事项中编制对立估价的估价师, 应着眼于证明委员会工作中可以指明的方法错误, 而非提出另一套估价说理; 同时应意识到, 基于既有租赁安排的 *lucro cesante* 主张面临沉重的举证要求。

Corte Suprema (Coquimbo 征收案)。 智利最高法院 2025 年 7 月 17 日判决。

智利最高法院维持了上诉法院提高 Coquimbo (科金博) 征收案赔偿额的裁判: 赔偿额由国家主张的 UF 2,349.07 (依据 *Conservador de Bienes Raíces*, 不动产登记处的登记成交案例) 提高至 UF 2,969.75 (依据一名私人估价师的报告)。法院强调, 上诉法院援用该私人估价师的报告, 并结合证人证言以及另一份关于邻近房产的估价报告一并考量, 属于 *sana crítica* (合理心证) 的容许行使, 而撤销审查并不延及对此种证据权衡的再行评判。

对估价师的意义。Coquimbo 一案是 Rol N° 31.423-2025 的对照: 既然委员会的估价在受开发商质疑时能够胜出, 那么私人估价师较高的估价在有充分的方法论说理与佐证证据支撑时, 同样可以胜出。对智利估价师而言, 实务上的教训是: 私人估价的方法与文档支撑具有决定意义; 仅凭登记成交案例的证据, 不足以推翻一份运用了额外可比赛例与情境证据、论证充分的私人估价。

Tribunal Constitucional, 2025 年 10 月。 《税法典》项下的 *bienes raíces* (不动产) 估价。

智利宪法法院驳回了针对 *Código Tributario* (《税法典》) 中规范不动产 (*bienes raíces*) 为 *impuesto territorial* (土地税) 目的进行估价的条款的合宪性挑战。持异议的大法官认为, 被质疑的规范未为被执行方提供充分的机制以对估价提出异议, 损害了宪法上关于理性且公正程序的保障。

对估价师的意义。该裁判保留了智利房地产税现行的行政估价框架, 但异议意见的说理指出了现行框架中的一处脆弱之处, 今后的宪法或行政法挑战可能会重新提出。在与房地产税争议有关的情形下编制估价的智利估价师, 应关注裁判之后的政策环境, 因为为改进纳税人异议机制而进行行政改革看来是可能的。

Colombia (哥伦比亚)

Corte Constitucional, 对 Jurisdicción Agraria y Rural 的核准。 哥伦比亚宪法法院 2025 年裁判。

哥伦比亚宪法法院对创设新的 Jurisdicción Agraria y Rural (农业与农村管辖法院系统) 的立法通过了合宪性审查。该院核准了界定改革对象、其纳入司法机关的方式、专门法庭与法院的设立以及预算方面的关键条文。该新管辖体系定于 2027 年开始运作, 旨在解决土地返还争议、农业财产争议, 以及因执行 2016 年和平协议而产生的事项。截至核准之日, 涉及六万一千余公顷的农业争议仍待现行行政与司法系统处理。

对估价师的意义。该新管辖体系将成为哥伦比亚土地返还程序与农村财产事项所生估价争议的主要审理场所。在返还情形下执业的哥伦比亚估价师 (包括为 Unidad de Restitución de Tierras (土地返还局) 与 Instituto Geográfico Agustín Codazzi (阿古斯丁·科达西地理研究所) 工作者) 应当预期, 新法庭将发展出自身的估价判例, 而方法论标准可能自现行以行政为主导的框架演进。从业者应关注 2027 年的运作启动, 以及这些法庭初期案件量的发展。

现行以 IGAC 为基础的土地返还估价框架。 2025 年持续适用。

哥伦比亚的土地返还实务继续以 Instituto Geográfico Agustín Codazzi (IGAC) 编制的 *avalúo comercial* (商业估价) 作为返还与置换情形下补偿的主要依据, 并在需要时以 Lonjas de Propiedad Raíz (不动产同业公会) 编制的 *avalúos* 作为补充。该框架体现于宪法法院 2018 年第 SU-034 号裁判 (关于履行的程序要求) 以及其后的裁判, 包括 2024 年第 T-120 号裁判 (关于在以实物补偿情形下 IGAC 估价与个别受益人偏好之间的关系)。

对估价师的意义。对于从事哥伦比亚土地返还业务的估价师而言, 该制度框架把 IGAC 估价置于补偿计算的 actual 运作核心, 而 Lonja de Propiedad Raíz 的 *avalúos* 则充当补充性或争议解决机制。即将设立的 Jurisdicción Agraria y Rural 可能改变这一制度架构, 而 IGAC 估价与 Lonja 估价之间的相对分量, 正是新法庭可能着手发展的问题之一。

Mexico (墨西哥)

Suprema Corte de Justicia de la Nación, 关于 avalúo 与成交价差额的 ISR 纳税义务。 墨西哥最高法院, 2026 年初公布的裁判。

墨西哥最高法院 (SCJN) 全席宣告 Ley del Impuesto sobre la Renta (ISR, 《所得税法》) 第 125 条、第 130 条第 4 项与第 132 条合宪。这些条款规定: 当房地产取得的 *avalúo* (估价) 超出合同购买价格百分之十以上时, 其差额被视为购买方的应累计所得, 因而须缴纳 ISR。法院认为, 该规则并不违反税收的比例原则与公平原则, 理由是以实质低于市场价值的价格取得财产, 构成购买方财产的客观增加, 因而恰当地构成应税的经济利益。

对估价师的意义。该裁判对墨西哥房地产交易以及为其提供支持的估价实务具有重大的实务后果。由 *corredores públicos* (公共经纪人) 或其他获授权专业人士出具的 *avalúos*, 在其超出合同价格百

分之十以上时，如今对购买方产生直接的财税后果。估价师应当预期，交易类 *avalúos* 的方法将受到更严格的审视，且对以特别严谨方式记录可比销售依据的 *avalúos* 的需求将会上升。该裁判还隐含地提高了在各种估价方法之间作出方法论选择的分量，因为这一选择如今可能影响购买方的纳税义务。

Suprema Corte de Justicia de la Nación, Amparo en revisión 391/2024。 墨西哥最高法院第二庭 2025 年 6 月 25 日判决。

墨西哥最高法院第二庭审结了一家矿业公司提出的合宪性挑战，其针对的是一项影响采矿区块及相关附属权利（临时占用、地役权、用水特许优先权与租约续期条款）征收法律框架的法令。法院认为，该法令并不违反不溯及既往原则，因为受影响的条款所涉及的只是对未来权利的期待，而非既得权利。

对估价师的意义。该裁判对墨西哥矿业权与矿业地产估价具有重要意义。法院将采矿区块延期权及相关条款定性为“单纯期待”而非既得权利，影响采矿特许权的估价以及对其估价中监管风险的评估。在墨西哥编制矿业地产或相关权利估价的估价师，应在估价中明确处理监管风险这一维度，并应避免把或有的、依赖于延期的价值当作已完全既得的价值来对待。