

Valuation Engineer Journal

가치의 공학—측정하고, 절제하며, 재현한다.

제 1 권 · 제 1 호 · 2026 년 여름



이번 호의 내용

부동산 감정평가에서 이론과 실무의 간극
Residual Constraint Approach(RCA) —체계와 절차
earthUI · glmnetUI · mgcvUI —열린 모형화 인터페이스

기초 개념 (**F0-F6**) 이질재 · 특성 공간 ·
헤도닉 가격함수 · 잠재가격 · 헤도닉 회귀 · 잠재변수

최근 판례 미국 · 유럽연합 · 라틴아메리카

Valuation Engineer Journal

제 1 권 · 제 1 호
2026 년 여름
ISSN 3143-5033

편집장 William Bert Craytor, SRA

차례

사명	2
편집인의 글	3
창간호에 부쳐	4
독자 서한	6
서신 교환에 대한 초대	7
논문	8
부동산 감정평가에서 이론과 실무의 간극	9
Residual Constraint Approach(RCA) —체계와 절차	39
earthUI —earth(MARS) 패키지를 위한 대화형 인터페이스	90
glmnetUI —glmnet(라소·엘라스틱 넷) 패키지를 위한 대화형 인터 페이스	133
mgcvUI —mgcv(일반화 가법 모형) 패키지를 위한 대화형 인터페이스	168
기초 개념	204
개관 (F0)	205
이질재 (F1)	211
특성 공간 (F2)	218
헤도닉 가격함수 (F3)	225
잠재가격 (F4)	231
헤도닉 회귀 (F5)	238
잠재변수 (F6)	246
최근 판례	252
최근 판례—미국 (R1)	253
최근 판례—유럽연합 (R2)	256
최근 판례—라틴아메리카 (R3)	261

사명

Valuation Engineer Journal 은 가치의 공학에 관한 엄밀하고 실무에 뿌리를 둔 연구와 사례 연구를 게재합니다. 곧 기술적 판단을 경제적 결과와 잇는 방법과 모형, 측정입니다. 본 학술지는 재현성과 전제의 명시, 서술의 명료함을 중시합니다.

편집인의 글

섹션: 편집자의 말

창간호에 부쳐

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 감수하였습니다.

영문판이 정보입니다.

William Bert Craytor, SRA — 편집장

제 1 권 제 1 호 · 2026 년 여름

발행일: 2026 년 7 월 1 일

DOI: 10.67330/0tymfa32

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

본 오픈액세스 논문은 크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스 (CC BY 4.0)의 조건에 따라 배포됩니다.

초록. 제 1 권 제 1 호 편집자의 말: 본 저널이 존재하는 이유, 다섯 편의 논문과 세 개의 섹션이 각각 기여하는 바, 그리고 함께해 달라는 청.

주제어: 편집자의 말, 감정평가 엔지니어링, 창간호

The Valuation Engineer Journal(감정평가 엔지니어 저널) 창간호에 오신 것을 환영합니다.

본 저널은 하나의 간극 때문에 존재합니다. 감정평가 이론은 오래전부터 감정평가를 시장 행태에 대한 엄밀한 측정으로 서술해 왔지만, 일상의 실무는 여전히 대체로 수작업에 의한 비교사례 선정과 오직 판단만으로 추정된 보정치에 의존하고 있습니다. 그 사이, 시장이 부동산의 특성에 실제로 얼마를 지불하는지 측정할 수 있는 통계적 도구들 — 헤도닉 회귀 (hedonic regression), 현대적 스플라인 모형, 교차검증된 기계학습 — 은 수십 년에 걸쳐 성숙해 왔음에도 실제 감정평가 업무에는 거의 수용되지 못했습니다. 감정평가 엔지니어 (valuation engineer) 는 그 간극을 메우는 실무자를 가리키는 우리의 명칭입니다. 자격을 갖춘 감정평가사의 역량에 데이터 엔지니어링, 통계학, 그리고 재현 가능한 프로토콜을 결합한 사람입니다.

이번 호의 다섯 편의 논문은 그 작업의 토대를 놓습니다. 첫 번째 논문은 이론과 실무 사이의 간극 그 자체를 살피며, 이 직역 (職域) 의 기존 자체가 이미 일상적 방법이 제공하는 것보다 더 많은 측정을 요구하고 있음을 논증합니다. 두 번째는 — 그리고 본 저널이 의도하는 모든 것의 중심에 있는 것은 — 잔차제약접근법 (Residual Constraint Approach) 의 온전한 제시입니다. 이는 회귀 채널이 측정 가능한 것을 측정하도록 하고, 측정할 수 없는 것은 명시적인 잔차 분해로 규율하며, 총액을 제약함으로써 판단이 가치를 정련할 수는 있으나 결코 창작해 낼 수는 없도록 하는 프로토콜입니다. 이어서 세 편의 소프트웨어 논문이 earthUI 와, 별점회귀 및 일반화가법모형을 위한 그 자매 인터페이스들 — 본 저널이 딛고 선 공개 도구들 — 을 기록합니다. 이들은 인쇄 직전에 하나의 응용프로그램으로 통합되었으며, 각 논문에 실린 발행인 주기 (註記) 가 이를 설명합니다.

세 개의 섹션이 논문들과 함께합니다. *Foundations*(기초)는 일곱 편의 짧은 항목 (F0–F6)으로 문을 열어, 이질재 (heterogeneous goods)에서 출발해 특성공간, 헤도닉 가격함수, 잠재가격 (implicit prices), 헤도닉 회귀를 거쳐 잔차 속에 깃든 잠재변수에 이르는 개념의 사슬을 짚어 갑니다. 각 항목은 그 학문적 계보를 인용하고, 고유의 DOI를 지니며, 기계가 읽을 수 있는 Prolog 형식화로 끝맺습니다 — 우리가 호를 거듭하며 이 분야의 정밀한 공유 어휘 체계로 키워 가고자 하는 지식베이스의 시작입니다. *Recent Court Decisions*(최근 법원 판결)는 감정평가 실무와 관련된 미국, 유럽연합, 라틴아메리카의 판결들을 개관합니다. 법적 환경 또한 이 공학적 문제의 일부이기 때문입니다. *Letters to the Editor*(독자 서신)는 서신 지면을 엽니다. 쓰이도록 만들어진 방법은 그 저자가 결코 예상하지 못한 방식으로 검증받게 마련이며, 그러한 발견들 — 그리고 그것이 불러일으키는 이견들 — 이 자리할 곳이 바로 여기입니다.

이번 호의 모든 항목은 CC BY 4.0에 따라 PDF와 HTML 두 가지 형식으로 오픈액세스 출판되고, Crossref DOI를 부여받으며, Zenodo에 독립적으로 보존됩니다. 우리는 감정평가사, 경제학자, 엔지니어 모두로부터의 기고 — 논문, Foundations 항목, 서신, 그리고 이견 — 를 청합니다.

William Bert Craytor, SRA

편집장

독자 서한

섹션: 독자 서신

독자 여러분께 드리는 서신 초대

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 감수하였습니다.

영문판이 정본입니다.

William Bert Craytor, SRA — 편집장

제 1 권 제 1 호 · 2026 년 여름

발행일: 2026 년 7 월 1 일

DOI: 10.67330/r55mw896

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

본 오픈엑세스 논문은 크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스 (CC BY 4.0)의 조건에 따라 배포됩니다.

초록. 편집장이 독자 서신 지면을 엽니다. 본 저널이 독자로부터 받기를 바라는 것은 무엇이며, 어떻게 보내면 되는가.

주제어: 독자 서신, 서신 교환, 동료 논평

말하기만 하는 저널은 반쪽짜리 저널입니다. 이 지면은 나머지 반쪽을 위해 존재합니다.

The Valuation Engineer Journal(감정평가 엔지니어 저널)은 쓰이도록 만들어진 방법들을 신습니다 — 그리고 쓰이는 방법은 그 저자가 결코 예상하지 못한 방식으로 검증받게 마련입니다. 잔차제약접근법 (Residual Constraint Approach)을 적용해 보시고 귀하의 시장에서 제대로 작동하지 않는 단계를 발견하셨다면, 우리는 그 서신을 원합니다. 어떤 Foundations(기초) 항목이 어떤 용어를 귀하의 실무가 요구하는 것보다 좁게 정의하고 있다면, 우리는 그 서신을 원합니다. 우리가 요약한 어떤 법원 판결이 해당 관할의 실무자에게는 달리 읽힌다면, 그 서신 또한 원합니다. 이견, 재현, 정련, 그리고 단순한 정정 모두가 이 지면의 범위 안에 있습니다. 실무로부터의 짧은 보고 — 효과가 있었던 기법, 해결한 데이터 문제, 다른 이들에게 일러 줄 만한 실패 — 도 마찬가지입니다.

지켜야 할 원칙은 몇 되지 않습니다. 서신은 기명이어야 하고, 간결해야 하며 (한 쪽이 가장 좋습니다), 구체적이어야 합니다 — 다루고자 하는 논문, 절, 항목을 밝히시고, 작업파일 (workfile)에서 하시듯 근거를 제시해 주십시오. 분량과 명료성을 위해 편집할 수 있으나, 언제나 저자의 동의를 얻습니다. 서신은 이 지면의 다른 모든 것과 마찬가지로 CC BY 4.0에 따라 오픈엑세스로 출판되며, 그 저작이 논의된 저자에게는 같은 호에서 답할 기회가 주어집니다.

서신은 저널 웹사이트 (journal.valuation-engineer.com)를 통하거나 편집장에게 직접 보내 주십시오. 본 저널이 받게 될 첫 실제 서신들은, 이 저널이 출판물을 넘어 하나의 대화가 되는 순간을 알릴 것입니다 — 우리는 그것을 고대하고 있습니다.

William Bert Craytor, SRA

편집장

논문

부동산 감정평가에서 이론과 실무의 괴리: 기능, 경제, 긴장에 관한 구조적 고찰 (한국어판)

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 검토했습니다. 영문판이正本입니다.

William Bert Craytor

제 1 권 제 1 호 · 2026 년 5 월 21 일

2026 년 7 월 1 일 발행

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

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

이 오픈 액세스 논문은 [크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스 \(CC BY 4.0\)](#)의 조건에 따라 배포됩니다.

초록. 부동산 감정평가 (real estate appraisal) 는 상당한 공익적 중요성을 지니는 위치에 서 있다. 그럼에도 과거와 현재의 조건 아래에서의 시장가치 의견 (opinion of market value) 을 중심에 두는 그 공식 이론 (doctrine) 은, 감정평가 산출물이 현대 금융 시스템에서 실제로 사용되는 방식과 점점 더 큰 긴장 관계에 놓이고 있다. 이 논문은 그 논거를 다섯 단계에 걸쳐 구축하며, 각 단계마다 명시적이고 질의 가능한 Prolog 지식베이스 (knowledge base) 를 함께 제시한다. 앞의 두 단계는 감정평가가 수행하는 사회적·경제적 기능을 확립하고, 세 번째 단계는 그 주된 최종 산출물인 시장가치 의견을 지목하며, 네 번째 단계는 계량적·금융적 예측 (forecasting) 에 견주어 감정평가가 지니는 시간적·방법론적 지향과 양자를 가르는 규제상의 경계를 살핀다. 이 논의들은 다섯 번째 단계의 중심 주장으로 수렴한다. 즉 주택담보대출에서 담보인정비율 (loan-to-value ratio, LTV) 로 조정된 시장가치는 장래 경매가치 (foreclosure value) 의 암묵적 대리 지표 (proxy) 로 기능하는데, 이는 현재가치 이론이 인정하지 않는 미래지향적 용법이다. 많은 감정평가사 (appraiser) 는 이러한 담보평가상의 역할을 알지 못한 채, 그 역할이 요구하는 만큼 장래 가치 추세에 대한 주의를 기울여 시장가치를 다루지 않는다. 이 논문은 바로 그 문제를 드러내고 후속 연구를 촉구하고자 한다.

주제어: 부동산 감정평가, 시장가치, 감정평가 이론, 지식표현, 감정평가 기준, 경매가치

목차

1 사회적 중요성	1
2 경제적 중요성	6
3 감정평가의 최종 산출물로서 시장가치 추정치가 갖는 중요성	10
4 감정평가 활동의 시간적 지향	13
5 경매가치의 대리 지표로서의 시장가치 × 담보인정비율 (LTV)	20
6 결론	26

1 사회적 중요성

부동산을 비롯한 자산의 감정평가 (appraisal) 는 매수인, 매도인, 대출기관, 보험자, 투자자, 법원, 과세관청, 그리고 일반 공중이 재산과 그 밖의 부 (富) 의 이전, 금융, 과세, 관리에 관하여 충분한 정보에 근거한 공정한 결정을 내리도록 도움으로써 필수적인 사회적·경제적 기능을 수행한다. 고품질의 감정평가는 증거, 분석, 전문가적 판단에 기초한 독립적이고 충분히 뒷받침된 가치 의견

을 제공함으로써 시장의 투명성, 공중의 신뢰, 효율적 자본배분, 그리고 금융기관의 안정성에 기여한다. (LaCour-Little and Malpezzi 2003; K'Akumu and Larsen 2024; Hendershott and Kane 1995; Lee, Peng, and Liao 2020)

이와 달리 저품질의 감정평가는 여러 시간 지평에 걸쳐 심각한 결과를 낳을 수 있다. 단기적으로 부정확한 감정평가는 개별 매수인이나 매도인에게 과다지급, 저가매도, 과중한 과세, 부적절한 대출 결정을 통하여 즉각적인 재산상 손해를 입힐 수 있다. 중기적으로 체계적인 감정평가의 결함은 시장 신호를 왜곡하고, 여신심사를 저해하며, 투기적 행태를 조장하고, 불공평한 과세나 보험 보장에 기여하며, 소송과 거래상의 불신을 증가시킬 수 있다. 장기적으로 감정평가 품질의 지속적 실패는 시장과 제도에 대한 신뢰를 훼손하고, 금융 불안정, 자원의 오배분, 지역의 쇠퇴 또는 과잉투자, 그리고 가계·기업·지역사회에 영향을 미치는 광범위한 사회적 불평등에 기여할 수 있다. (LaCour-Little and Malpezzi 2003; Mayer and Nothaft 2022)

부동산과 그 밖의 자산은 흔히 개인·기업·공공의 부에서 상당한 부분을 차지하므로, 감정평가의 품질은 단지 사적인 기술적 관심사에 그치지 않고 경제적 회복력, 공정성, 장기적 공익과 결부된 중대한 공익의 문제가 된다.(Mayer and Nothaft 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 경제적 중요성

부동산을 비롯한 자산의 감정평가는 신용시장의 효율적 작동, 투자 결정, 조세 제도, 재무보고, 보험 인수, 상속재산 관리, 기업 거래, 공공정책 기획을 뒷받침함으로써 근본적인 경제적 역할을 수행한다. 신뢰할 수 있는 감정평가는 담보가치와 시장가치에 대한 믿을 만한 척도를 확립하도록 도와, 차주와 대출기관, 투자자, 정부, 기업 사이에서 자본이 더 효율적으로 흐를 수 있게 한다.(LaCour-Little and Malpezzi 2003; Ebrahim and Hussain 2010)

단기적으로 정확한 감정평가는 거래상의 불확실성을 줄이고, 대출 결정을 개선하며, 공정한 협상을 촉진하고, 과대평가 또는 과소평가에서 비롯되는 즉각적인 금전적 손실을 예방하는 데 기여한다. 중기적으로 감정평가의 품질은 주택담보대출 위험, 투자 배분, 조세 형평, 보험의 적정성, 시장참여자의 가격 설정 행태에 영향을 미침으로써 더 넓은 시장의 안정성과 효율성을 좌우한다. 이 기간에 부실하거나 체계적으로 편향의 (bias) 된 감정평가는 자산 거품의 팽창, 대출 포트폴리오의 부실화, 개발 유인의 왜곡, 채무불이행 위험과 소송 위험의 증가에 기여할 수 있다. (LaCour-Little and Malpezzi 2003; Hendershott and Kane 1992; Griffin and Maturana 2016)

장기적으로는 감정평가 품질의 누적적인 경제적 결과가 한층 더 중대해져, 자본 형성의 양상, 사회기반시설 투자, 세대 간 부의 이전, 공공 수입, 주거 부담능력, 그리고 금융시스템 전반의 회복력에 영향을 미친다. 평가실무의 결함이 지속되면 자본의 만성적 오배분, 은행의 불안정, 투자자 신뢰의 저하, 토지와 생산적 자산의 비효율적 이용을 초래할 수 있는 반면, 견실한 감정평가 기준은 안정적인 시장, 건전한 경제 기획, 장기적 경제성장을 뒷받침하는 데 도움이 된다.(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          % Function, Horizon
    , economic_effect/4            % Quality, Horizon, Function,
    ↪ Effect
    , economic_stakeholder/1
    , economic_benefits/2          % Stakeholder, Function
    , economic_positive_effect/3
    , economic_negative_effect/3
    , economic_horizon_summary/2   % Horizon, Effects
    , economic_consequences_of_quality/2 % Quality, ConsequenceList
    , economic_functions_at_horizon/2
    , economic_stakeholders_affected_by/2
    , economic_reliable_appraisals_improve/1
    , economic_appraisal_quality_matters_at/1
  ]).

:- discontiguous economic_effect/4.
:- discontiguous 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 감정평가의 최종 산출물로서 시장가치 추정치가 갖는 중요성

시장가치 감정평가는 부동산이나 자산이 통상적인 시장 조건에서 실현할 수 있는 개연적 가격에 대한 전문적 추정치를 제공하며, 그럼으로써 신용 익스포저와 잠재적 경매 결과를 평가하는 대출기관, 투자자, 담보분석가에게 기초적인 기준점 역할을 한다. 이 점에서 시장가치 감정평가는 담보위험 분석에서 중요한 역할을 수행한다.(Agarwal, Ben-David, and Yao 2015) 경매가치 (foreclosure value) 또는 청산가치는 매각 여건의 악화, 매각기간의 단축, 법적 비용, 물리적 훼손, 부정적인 시장 심리 등으로 인해 시장가치와 상당히 달라질 수 있으나, 담보분석가는 스트레스 시나리오 아래에서의 회수 가능성을 추정할 때 그 출발점으로 시장가치 감정평가에 자주 의존한다.(Campbell, Giglio, and Pathak 2011; Pennington-Cross 2006; Clauretie and Daneshvary 2009)

담보분석가는 시장가치, 대출 잔액, 담보인정비율 (LTV), 시장 추세, 부동산의 상태, 환금성, 지역의 수급 여건 사이의 관계를 분석함으로써, 차주가 채무를 불이행할 경우 손실이 발생할 가능성과 그 심도를 평가한다.(Qi and Yang 2009)

단기적으로 정확한 시장가치 감정평가는 대출기관이 신중한 여신심사 결정을 내리고 적절한 담보 여유분을 설정하도록 돕는다. 중기적으로는 금융기관이 경기순환에 따라 변화하는 담보의 적정성을 평가하도록 도움으로써 포트폴리오 위험관리, 대출채권 유동화, 충당금 배분, 규제 준수를 뒷받침한다. 장기적으로는 일관되게 신뢰할 수 있는 감정평가 실무가 과도한 차입, 담보가 부실한 대출, 그리고 시장 스트레스 국면에서 경매 손실과 광범위한 경제적 교란을 증폭시킬 수 있는 담보자산의 체계적 가격 왜곡의 위험을 줄임으로써 금융시스템의 안정에 기여한다.(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
]).

:- discontinuous collateral_effect/4.
:- discontinuous 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 감정평가 활동의 시간적 지향

감정평가사는 일반적으로 장기간에 걸친 확률적 예측이 아니라, 특정한 기준시점 —소급 (과거), 현재, 또는 정해진 제약 아래에서의 장래 (미래)—을 기준으로 한 시점 (point-in-time) 시장가치

의견을 도출한다. 이러한 시간적·방법론적 지향은 규제되는 감정평가 실무와 계량분석가, 재무분석가, 투자분석가, 경제학자의 업무를 가르는 핵심적 구별 중 하나이다.(Quan and Quigley 1991)

대부분의 감정평가 업무, 특히 전문감정평가실무통일기준 (Uniform Standards of Professional Appraisal Practice, USPAP) 이나 국제평가기준 (International Valuation Standards, IVS) 과 같은 기준에 따른 시장가치 평가에서, 감정평가사는 하나의 기준시점을 기준으로 관측 가능한 시장 증거, 당시의 여건, 공인된 방법론에 근거를 둔 조정된 점추정치 (또는 허용되는 경우 좁은 범위) 를 산출한다. USPAP 은 가치 결론을 특정 금액, 수치의 범위, 또는 기준값과의 관계로 제시하는 것을 명시적으로 허용하지만, 조정을 거친 의견은 여전히 그 시점에 결부된 신뢰할 수 있는 점 (點) 지시치로 남는다.(The Appraisal Foundation 2024; International Valuation Standards Council 2025)

감정평가사는 추세, 수요와 공급, 감가, 환원을 또는 할인율, 흡수율에 관한 시장참여자의 합리적 기대를 필연적으로 반영하지만, 이는 투기적 변동을 예측하기 위해서가 아니라 기준시점의 시장에 의해 뒷받침되는 가치를 논증하기 위해서이다. 장래 가치 의견 (예: 사업 완공 시점의 안정화 가치) 은 명시적인 특별가정 (extraordinary assumption) 이나 가상조건 (hypothetical condition) 을 요구하며, 시장에서 도출된 근거와 함께 정해진 장래 기준시점에 한정된다.

이와 달리 계량분석가, 재무분석가, 투자분석가는 확률분포, 몬테카를로 모의실험, 시나리오 분석, 민감도 검토, 또는 위험가치 (value-at-risk) 지표를 사용하여 여러 기간에 걸친 가능한 결과의 범위를 포괄하는 추정치를 통해 불확실성을 다루는 경우가 많다. 이들의 초점은 하나의 시점에서의 단일한 시장가치보다는 서로 다른 시간 지평에 걸친 기대 장래 성과, 위험조정 수익률, 또는 특정 당사자에게 갖는 투자가치에 놓이는 것이 보통이다.(Hull 2018)

이러한 업무는 또한 감정평가사 자격이 요구하지도 증명하지도 않는 기술적 역량을 전제로 한다. 예컨대 다변량 적응 회귀 스플라인 (multivariate adaptive regression spline, MARS) 모델을 적합시키려면, 표본 내의 높은 적합도 (높은 R^2) 를 과적합의 대가로 얻는 것이 아니라 검증용 자료에까지 이어지도록 (높은 교차검증 R^2) 모형의 복잡도를 통제하는 판단이 필요하며, 이는 다시 기저 알고리즘에 대한 실무적 이해를 전제한다.(Friedman 1991; Steurer, Hill, and Pfeifer 2021) 이러한 계량적 숙련과 더불어, 이 업무는 고도로 구조화된 개념을 다루고, 의미의 미세한 차이를 식별하며, 낯선 방법을 소화해 내는 개념적 역량에 보답한다. 이는 자격증이 아니라 능력의 문제이다. 자격을 갖춘 감정평가사가 그것을 지닐 수도 있고 무자격 분석가가 그것을 결여할 수도 있으며, 바로 그렇기 때문에 이러한 능력만으로는 특정 업무가 규제선의 어느 쪽에 속하는지를 결정하지 못한다.

감정평가사 역시 제한적으로 미래지향적 요소를 다룰 수 있다. 현금흐름할인법 분석에서 수익 흐름, 공실률, 비용, 복귀가치를 추정하거나, 개발 또는 소송 맥락에서 장래 가치를 추정하는 것이 그 예이다. 이러한 요소들은 현재의 시장 증거와, 논증 가능성과 근거 없는 추측의 회피를 강조하는 전문직 기준에 뿌리를 둘 때 감정평가의 범위 안에 머문다.(Diaz and Wolverton 1998; Gallimore 1996; Crosby, Lizieri, and McAllister 2010) 현대의 감정평가 실무는 특히 대량 평가, 자동평가모형, 포트폴리오 업무에서 통계적 도구, 시나리오 고려, 위험분석을 점점 더 많이 통합하고 있으나, 핵심 산출물은 대체로 여전히 시점 기준의 시장가치 의견이다.(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)

감정평가 실무는 주로 증거 중심적이며 과거에서 현재에 이르는 (또는 제한적으로 장래의) 점 추정치를 지향하는 반면, 재무·계량 분석은 더 확률적이고 분포적이며 여러 시간 지평에 걸쳐 명시적으로 미래를 지향한다고 특징지을 수 있다.(Geltner 1991) 감정평가사는 “현재의 증거와 정해진 가정 아래에서, 지정된 기준시점을 기준으로 가장 잘 논증되는 시장가치 지시치는 무엇인가?”라고 묻는다. 재무·계량 분석가는 그보다 “장래의 어느 기간에 걸쳐 어떤 결과의 범위가 어떤 확률로 나타날 가능성이 높으며, 자원은 어떻게 배분되어야 하는가?”라고 묻는 경우가 더 많다.

이 경계는 절대적이지 않으며 시장의 복잡성과 함께 계속 변화하고 있다. 그럼에도 목적 (시장 가치 대 (對) 투자가치), 증거 요건, 규제 체계에서의 근본적인 차이는 여전히 존속한다.

그러나 실제로 작동하는 선은 능력이 아니라 법에 의해 그어진다. 결정적인 물음은 단지 특정 업무가 감정평가사 자격을 법적으로 요구하는 업무인지 여부이다. 계량분석가, 재무분석가, 경제

학자, 투자분석가, 포트폴리오 관리자 및 관련 전문직은 그 업무가 규제 대상 감정평가에 해당하지 않는 한, 부동산 감정평가사 자격 없이도 장래 자산가치를 적법하게 추정하거나 예측하거나 모형화하거나 분석할 수 있으며, 여기에는 분포 기반 또는 범위 기반의 추계도 포함된다.(Murphy 2012) 핵심적 구별은 다음에 달려 있다.

1. 가치의 종류 (예: 시장가치 점추정치 대 투자가치 또는 확률적 예측치),
2. 시간적 초점 (특정 기준시점 대 일정 기간 또는 시간 지평),
3. 산출물의 형태 (점 또는 좁은 범위 대 완전한 확률분포),
4. 의도된 목적, 그리고 그 업무가 감정평가로 표시되는지 여부,
5. 그리고 USPAP 과 같은 기준의 준수를 요구하는 규제 대상 대출 또는 법적 맥락의 관여 여부.

이 체계는 주로 미국의 금융기관개혁구제집행법 (FIRREA, 1989) 과 같은 은행 규제에서 비롯되었으며, 이 법은 연방 관련 거래에 대하여 신뢰할 수 있고 독립적인 시장가치 의견을 확보하고자 하였다.(United States Congress 1989; K'Akumu and Larsen 2024) 주식 리서치의 목표주가, 타당성 조사, 포트폴리오 위험 모형화, 옵션 가격결정, 예측 분석 등 정교한 가치평가 활동의 상당수는 규제 대상 감정평가 기능이 아니라 투자, 예측, 자문의 목적에 봉사하므로 자격 규제의 밖에 놓인다.

```
%% 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_appraiser).

%% --- 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 R2 against CV-R2
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 경매가치의 대리지표로서의 시장가치 × 담보인정비율 (LTV)

앞의 여러 절은 감정평가가 무엇을 위한 것인지, 무엇을 산출하는지, 그리고 그 이론적·규제적 경계가 어디에 놓이는지를 서술하였다. 이 절은 그 논의가 쌓아 온 긴장에 도달한다.

주택담보대출과 담보분석의 밑바탕에 놓인 더 깊은 실무적 현실 가운데 하나는, 형식적으로는 감정평가가 특정 기준시점의 현재 시장가치 의견으로 구성됨에도 불구하고, 실제 경제적 기능에서는 여신심사에 사용되는 시장가치 감정평가가 장래의 일정한 보유기간에 걸친 담보 회수 가능성의 암묵적 대리지표 역할을 하는 경우가 많다는 점이다.(Foster and Van Order 1984; Kau et al. 1995; Deng, Quigley, and Van Order 2000)

80% 라는 담보인정비율은 임의로 정해진 것이 아니다. 역사적으로 이 비율은 부분적으로 다 음의 위험으로부터 대출기관을 보호하려는 위험 완충장치로 발전해 왔다.

- 완만한 시장 하락,
- 경매 비용,
- 청산 할인,
- 보유 비용,
- 법률 비용,
- 미수이자,
- 그리고 채무불이행 이후 담보를 처분하는 데 걸리는 기간 동안의 불확실성.(Foote, Gerardi, and Willen 2008; Qi and Yang 2009)

이 완충장치는 또한 대출금액 자체—시장가치와 허용 담보인정비율의 곱—가 회수 가능한 경매가치에 대한 대출기관의 실무적 추정치로 읽힐 수 있는 이유이기도 하다. 대출금을 시장가치의 일부로 제한하는 것은 사실상 강제매각 대금이 대출 잔액을 밀돌지 않으리라는 데 거는 내기이기 때문이다. 그러나 그 할인폭이 어느 정도여야 하는지는 80% 라는 어렵수가 시사하는 것보다 훨씬 확정하기 어렵다. 경매 할인율의 추정치는 시장, 방법, 시기에 따라 크게 달라진다. 자주 인용되는 강제매각 연구 하나는 그 값을 비교 가능한 정상거래 가격보다 약 27% 낮은 수준으로 제시하는 반면 (Campbell, Giglio, and Pathak 2011), 부동산의 상태, 위치, 선택편의를 보정한 다른 분석들은 실질적으로 다른 크기를 보고한다.(Pennington-Cross 2006; Clauretie and Daneshvary 2009)

이와 달리 관행적인 80% 기준선은 본질적으로 전국 단위의 단일 모수로서, 극도로 이질적인 지역시장과 신용순환의 연속적 국면 전반에 걸쳐 일률적으로 적용된다. 이 기준선은 어떤 특정 시장의 할인율도 추적할 수 없으므로, 시장가치에서 경매 회수액으로의 정밀한 사상(寫像) 이 아니라 매우 거친 일률적 근사치로 기능한다. 오히려 그 완충은 얇은 편이다. 20% 의 할인폭은 해당 연구가 보고하는 27% 의 할인율보다 낮으므로, 위에 열거한 처분 비용을 차감하기도 전에 관행적 최대 대출금이 강제매각 총대금을 넘어설 수 있다.

차주의 자기자본이 너무 얇아 완충장치를 이루지 못하는 경우에는 그 자리를 모기지보험이 대신한다. 이는 정책적 조절장치라기보다 위험배분의 질충으로 이해하는 것이 가장 적절하다. 자산은 적으나 소득은 안정적인 차주층이 존재하고, 이들에게 LTV 80% 를 넘겨 대출할 때 추가되는 채무불이행 위험과 손실 위험은 보험자가 예측하고 부담한다. 고 LTV 관행 대출에 붙는 민간 모기지보험을 통해 사적으로 부담되거나, 미국 연방주택청 (Federal Housing Administration, FHA) 을 통해 공적으로 부담되는 것이다. 그러나 그러한 질충이 일상화되어 있는 시장 자체는 연방 주택정책의 산물이다. 관행 대출을 지배하는 정부지원기업 (GSE) 은 그러한 신용보강 없이는 LTV 80% 를 넘는 대출채권을 매입하는 것이 정관상 금지되어 있다 (United States Code 1970). 보험이 종료되는 조건 또한 위험 계산이라기보다 제도적 선택이며, 제도에 따라 크게 다르다.¹

그러므로 감정평가사는 공식적으로는 “현재의 시장가치”를 추정하고 있지만, 대출기관은 그 가치를 경제적으로 미래지향적 위험의 렌즈를 통해 해석하고 있다. 사실상 대출기관이 묻고 있는 것은 다음에 가깝다.

“이 차주가 예견 가능한 장래의 어느 시점에 채무를 불이행한다면, 담보는 잔존 채무와 경매 관련 손실을 충당하기에 충분한 상태로 남아 있을 가능성이 높은가?”

이는 본질적으로 미래지향적인 물음이다.

이러한 상황은 다음 두 가지 사이의 의도적인 개념적 분리에 기초한다.

- 감정평가사의 “현재 시장가치” 의견, 그리고
- 장래 담보가치 훼손에 대비한 예측적 완충장치로서 그 의견을 활용하는 대출기관의 실무.

역사적으로 그리고 법적으로, 규제당국은 감정평가사가 투기적 예측자가 되는 것을 막고자 하였다. 그 이유는 다음과 같다.

- 투기적 예측이 과거의 대출 위기에 기여하였고,
- 법원은 감정평가가 관측 가능한 증거에 결부되기를 원하였으며,
- 은행 규제당국은 표준화되고 방어 가능한 평가 실무를 추구하였기 때문이다.

이에 따라 감정평가 이론은 다음을 향하여 발전하였다.

- 현재가치,
- 현재의 시장 증거,
- 정해진 기준시점,
- 그리고 제약된 가정.

그러는 동안 대출기관, 투자자, 담보분석가는 바로 그 동일한 감정평가를 확률적 장래위험 체계 안에서 일상적으로 사용한다.

실제로 주택담보대출의 여신심사는 다음을 아우르는 통합적 체계이다.

1. 현재의 시장가치 평가,
2. 장래 위험 추정,
3. 채무불이행 확률,
4. 경매 시점,
5. 청산 손실의 심도,
6. 거시경제적 기대,
7. 그리고 포트폴리오 스트레스 분석.

감정평가사는 형식적으로 주로 (1) 에 기여하며, 대출기관은 그 결과를 내부적으로 나머지 항목들과 결합한다.

¹관행 대출에서 차주가 보험료를 부담하는 민간 모기지보험은 LTV 78% 에서 자동으로 종료되며, 1998 년 주택소유자보호법 (Homeowners Protection Act) 에 따라 80% 에서 해지를 청구할 수 있다.(United States Congress 1998) 연방주택청 (FHA) 의 보험료는 더 엄격하다. 2013 년 6 월 3 일 이후 실행된 대출로서 최초 담보인정비율이 90% 를 초과하는 경우 (예: 계약금 3.5%) 에는 연간 보험료가 대출 기간 전체에 걸쳐 부과되는 반면, 90% 이하의 대출은 11 년 이 지나면 종료된다.(U.S. Department of Housing and Urban Development 2013) 따라서 다수의 FHA 차주는 LTV 80% 부근의 관행 대출로 대환하는 방법으로만 모기지보험에서 벗어난다.

이 구별은 실무에서 흐려진다. 담보가 장래에도 충분한 가치를 유지할 것으로 믿지 않는 한 대출기관이 대출을 실행하지 않으리라 하면, 감정평가 기준이 형식적으로는 그렇게 서술하기를 피한다 하더라도 감정평가는 간접적으로 장래가치 보호 모형의 일부로 기능하는 셈이다.

이 긴장은 다음과 함께 한층 더 뚜렷해졌다.

- 자동평가모형 (AVM),
- 기관투자자의 포트폴리오 분석,
- 유동화 모형,
- 기계학습 기반 가치평가 시스템,
- 바젤 자기자본 모형화,
- 그리고 대규모 주택담보대출 스트레스 테스트.

현대의 담보분석은 부동산 가치를 단일 시점의 정태적 추정치가 아니라 시간에 걸친 확률적 대상으로 다루는 경향이 점점 강해지고 있다.(Basel Committee on Banking Supervision 2017; Steurer, Hill, and Pfeifer 2021; Board of Governors of the Federal Reserve System 2021)

전통적인 감정평가 규제는 역사적 타협의 산물이라고 볼 수도 있다.

- 감정평가사는 규율 있고 증거에 기반한 현재가치의 기준점을 제공하고,
- 금융기관은 장래 위험의 해석에 대한 책임을 진다는 것이다.

그러한 분리가 현대 금융시스템에서 여전히 온전히 일관성을 유지하는지는 정당하게 제기될 수 있고 또 점점 더 논쟁이 되고 있는 물음이다.

여기에서 감정평가 실무 자체에 대한 실천적 함의가 따라 나온다. 대다수의 감정평가사는 이러한 담보평가상의 역할을 알지 못한다. 시장가치를 미래지향적 완충장치로 사용하는 대출기관의 용법은 감정평가 업무의 내부에서는 좀처럼 보이지 않기 때문이다. 그리하여 그들은 그 역할이 요구하는 만큼 장래 가치 추세에 대한 주의를 기울여 자신의 시장가치 결론을 다루지 않는다. 현재 가치 의견이 실제로 어떻게 소비되는지를 인식하는 것은 그 간극을 좁히는 첫걸음이며, 이 논문이 후속 연구를 위해 열어 두고자 하는 영역이다.

```
%% 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 결론

앞의 지식베이스들에 부호화된 다섯 개의 서술은 하나의 전문직을 다섯 개의 상호보완적인 관점에서 그려 낸다. 각각을 따로 떼어 놓고 보면 저마다 자족적인 서술로 읽힌다. 감정평가가 깊어지는 사회적·공익적 이해관계, 감정평가가 뒷받침하는 경제적 기능, 담보위험 분석에서의 실무적 역할, 인접한 계량적 업무와 감정평가를 가르는 규제상의 경계, 그리고 형식적 구성과 실제적 사용 사이의 긴장이 그것이다. 그러나 이들을 함께 놓고 그 구조를 Prolog 로 명시할 때, 그것들은 한층 더 흥미로운 무엇인가를 그려 낸다. 즉 공식 이론과 경제적 기능이 의도적으로 떼어 놓인 채 유지되어 온 전문직, 그리고 그 산출물에 오늘날 적용되는 분석 도구들에 의해 내적 일관성이 점점 더 시험받고 있는 전문직의 모습이다.

앞의 세 지식베이스는 공통의 스키마를 공유한다. effect(Quality, Horizon, Domain, Effect) 가 그것이다. 각각은 감정평가의 품질을 인과적 투입으로 설정하고 그 결과가 단기·중기·장기의 시간 지평에 걸쳐 전파되는 것으로 파악하며, 다만 그 결과를 읽어 내는 영역에서만 차이를 보인다. 경제적 렌즈는 신용시장, 투자 배분, 과세, 자본 형성을 통해 그것을 읽고, 사회적 렌즈는 형평, 공중의 신뢰, 지역사회와 결과를 통해 읽으며, 담보의 렌즈는 여신심사, 포트폴리오 관리, 금융시스템의 안정을 통해 읽는다. 이 공유된 스키마는 우연이 아니다. 그것은 세 서술 전체를 관통하는 실질적 주장을 반영한다. 곧 가치평가 업무의 품질에는 결과가 따르고, 그 결과는 시간에 따라 커지며, 동일한 근원적 실패 양태가 누구의 이익을 전면에 놓느냐에 따라 다르게 나타난다는 것이다. 술어의 인수 쪽에 놓인 공유 원자들—특히 품질과 시간 지평의 어휘—은 세 어휘 체계를 결합 가능하게 만들며, 그리하여 체계적 과대평가와 같은 하나의 실패를 경제적 가격 왜곡, 사회적 불형평, 담보의 오측정을 통해 동시에 추적할 수 있게 한다.

네 번째 지식베이스는 어조를 바꾼다. 그것은 감정평가 품질의 결과를 서술하지 않고, 방법론적으로 그리고 법적으로 감정평가가 무엇인지를 서술한다. 그 술어들은 인과적이라기보다 분류적

이며, 그 추론 기제는 영향이 아니라 관할에 관한 물음에 답한다. 이러한 전환은 필연적이다. 앞의 세 서술이 취하는 결과 중심의 구성은 평가 대상이 되는 활동의 정의를 전제하는데, 그 정의 자체가 다투어지고 있기 때문이다. 범위 (scope) 술어들은 감정평가사가 하는 일—증거에 기반한 현재가치 의견과 좁게 한정된 미래지향적 요소—을 기록하고, 규제 (regulation) 술어들은 그 활동이 자격 있는 실무의 범위 안에 놓이는 조건을 기록한다. 전향 추론 규칙들 (*requires_license/1*, *exempt_from_appraisal_license/1*, *regulated_appraisal/1*) 은 주어진 가치평가 업무가 규제의 울타리 안에 있는지 밖에 있는지를 물을 수 있는 장치를 제공하는데, 이는 앞의 세 지식베이스가 이미 답이 나와 있다고 전제하는 물음이다.

다섯 번째 지식베이스는 하나의 테제를 제시한다. 다른 것들이 서술하고, 분류하고, 결과를 추적하는 데 그치는 반면, 대리지표 (proxy) 지식베이스는 현재가치 감정평가와 장래가치 예측 사이의 세심한 이론적 분리—규제 지식베이스가 기록하고 있는 그 분리—가 주택담보대출에서 감정평가가 실제로 수행하는 경제적 기능과 긴장 관계에 있다고 주장한다. 80% 의 담보인정비율 관행, 그것이 완충하는 일곱 가지 구체적 위험, 축자적으로 기록된 대출기관의 근원적 물음 (“이 차주가 예견 가능한 장래의 어느 시점에 채무를 불이행한다면, 담보는 충분한 상태로 남아 있을 가능성이 높은가...”), 그리고 일곱 가지 구성요소로 이루어진 통합적 여신심사 체계는 모두 동일한 주장을 뒷받침한다. 곧 감정평가는 형식적으로는 현재가치 의견이지만 경제적으로는 장래가치 완충장치라는 것이다. 이러한 구도를 낳은 역사적 타협—감정평가사는 규율 있는 현재가치의 기준점, 대출기관은 장래위험 해석의 자리—은 감정평가 자격제도가 등장한 여건 아래에서는 방어 가능한 대응이었으나, 해당 서술이 열거한 현대적 증폭 요인들 (AVM, 기계학습 기반 가치평가 시스템, 바젤 자기자본 모형화, 유동화 분석, 대규모 스트레스 테스트) 은 부동산 가치를 시간에 걸쳐 확률적으로 다룬다. 이 도구들은 이론적으로 분리되어 있던 층위들을 하나의 계산 파이프라인 안으로 다시 끌어들이며, 그 결과 층위 사이의 개념적 분리는 유지하기가 한층 더 어려워진다.

그러한 화해가 달성 가능한지, 그리고 어떤 조건에서 가능한지가 대리지표에 관한 서술이 닫히며 남기는 물음이다. 지식베이스들은 그 물음에 답하지 않는다. 다만 그것들은 답을 찾는 데 관련된 고려사항들에 대한 명시적이고 질의 가능한 서술을 제공한다. 여러 시간 지평과 영역에 걸친 가치평가 품질의 결과, 어떤 결과가 규제 대상 실무의 범위 안에 놓이는지를 결정하는 관할의 경계, 그리고 그 경계가 지금 압박을 받고 있는 구조적 이유가 그것이다. 그 서술은 방법론에 관하여, 기준의 개혁에 관하여, 감정평가 실무와 계량적 가치평가라는 더 넓은 체계 사이의 관계에 관하여 더 구체적인 논변을 세울 수 있는 토대가 된다.

지식표현에 관한 소고

명시적으로 기록해 둘 만한 세 가지 구조적 양상이 있다. 이들은 어느 하나의 서술에서가 아니라 부호화 과정 자체에서 드러난 것들이다.

첫째, 이 지식베이스 계열은 평평하지 않다. 다섯 중 셋은 하나의 스키마를 공유하고, 하나는 그 스키마가 언제 적용되는지를 결정하는 관할상의 한정을 제공하며, 하나는 형식적 구성과 경제적 기능 사이의 간극에 관한 테제를 제시한다. 다섯 가지를 모두 결합하는 연결 층위는 그것들을 동일한 행렬의 다섯 열로 다루지 않는다. 그것은 앞의 셋을 결과에 대한 서술로, 넷째를 그 서술들이 작동하는 범위를 제약하는 것으로, 다섯째를 그 제약 자체가 압박을 받고 있는 지점을 식별하는 것으로 다룬다. 이러한 비대칭적 구조는 결합이 아니라 하나의 특징이다. 그것은 서술적 주장, 관할적 주장, 비판적 주장이 진정으로 서로 다른 인식론적 역할을 차지하는 전문직 실무의 실제 조직을 반영하기 때문이다.

둘째, 모듈들을 구별하는 접두사 규율 (*economic_*, *social_*, *collateral_*, *scope_/regulation_*, *proxy_/underwriting_*) 은 충돌 위험이 있는 곳에서만 실질적인 일을 한다. 모듈들 사이에 직접적인 대응물이 있는 술어는 접두사를 가질 만하지만, 하나의 개념 영역에만 고유한 술어는 그렇지 않다. 인수 쪽의 공유된 값 어휘—품질, 시간 지평, 영역의 원자들—야말로 모듈들이 소통할 수 있게 하는 것이며, 그 값들에까지 접두사를 붙이면 바로 그 소통이 무너진다. 그리하여 드러난 양상은 일반적이다. 술어에는 접두사를, 인수에는 공유 원자를 두되, 술어가 하나의 렌즈에만 고유한 경우에는 일률적 접두사가 아니라 선택적 접두사를 쓰는 것이다.

셋째, 대리지표 지식베이스가 식별한 중심적 긴장은 범위·규제 지식베이스가 식별한 규제상의 비대칭과 밀접히 관련되지만 그것과는 구별된다. 양자는 병렬적인 2 항 술어—*regulatory_asymmetry/2* 와 *proxy_doctrine_practice_gap/2*—로 기록되어 있으며, 동일한

근원적 구도의 서로 다른 축을 서술하기에 의도적으로 이름을 달리하였다. 규제상의 비대칭은 누가 규제를 받으며 얼마나 무겁게 받는가에 관한 것이고, 이론과 실무의 괴리는 감정평가사의 산출물에 관하여 형식적으로 무엇이 이야기되는가와 그것으로 경제적으로 무엇이 행해지는가에 관한 것이다. 이 두 긴장은 서로를 강화한다. 규제는 감정평가사를 현재가치 이론 쪽으로 밀어붙이고, 경제적 기능은 그 산출물을 장래가치의 용법 쪽으로 끌어당긴다. 이를 합쳐 보면, 공식적 자기서술을 자신의 산출물이 흘러 들어가는 분석 파이프라인과 조화시키기가 점점 더 어려워지는 전문직의 모습이 드러난다.

참고문헌

- 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%.

잔차 제약 접근법 (RCA): 방법론적 틀과 프로토콜 (한국어 판)

2025 년 3 월 5 일 Zenodo 에 최초 발표된 논문의 개정판에 대한 한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 검토했습니다. 영문판이正本입니다.

William Bert Craytor, SRA

제 1 권 제 1 호 · 2026 년 여름

2026 년 7 월 1 일 발행

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

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

이 오픈 액세스 논문은 [크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스 \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)의 조건에 따라 배포됩니다.

초록. 전통적 감정평가 (appraisal) 방법의 중대한 한계는 짝지은 사례 비교분석 (matched pair analysis) 과 같은 낡은 수작업 기법에 의존한다는 점에 있다. 이러한 기법은 부동산 특성의 가치 기여 (value contribution) 를 마치 존재하지 않는 것처럼 건너뛴다. 전통적 거래사례비교표 (sales grid) 에는 가치기여를 적을 난이 없고 보정 (adjustment) 난만 있을 뿐이다. 그러나 실제로 가치기여야말로 전통적 감정평가사 (appraiser) 의 과대평가와 과소평가를 막아 주는 수학적 제약 (mathematical constraint) 의 필수적 토대를 제공한다.

잔차 제약 접근법 (Residual Constraint Approach, RCA) 은 다변량 적응 회귀 스플라인 (Multivariate Adaptive Regression Splines, MARS) 과 특성 가치기여에 대한 엄밀한 수학적 제약을 함께 사용하는 다단계 감정평가 절차를 통합함으로써 전통적 거래사례비교법 (Sales Comparison Approach, SCA) 을 발전시킨다. 이때 특히 문제가 되는 것은 측정되지 않은 특성, 즉 미관, 디자인, 상태, 품질과 같은 잠재변수 (latent variable) 이며, 그 가치는 전통적 감정평가사의 주관적 판단과 편향 (bias) 에 좌우되는 것이 보통이다.

RCA 의 수학적 제약은 측정 가능한 부동산 특성이 거래가격에 기여하는 가치를 추정하기 위해 MARS 회귀분석 (regression) 을 전문적으로 적용하는 데 의존한다. 그러나 이 추정치는 통상 실제 거래가격의 약 80% 만을 설명하며, 디자인, 기능적 효용, 상태, 품질과 같이 주관적으로 평가되는 특성은 여기에서 제외된다. 샌프란시스코 베이 에어리어 (San Francisco Bay Area) 에서는 이러한 주관적 구성요소를 나타내는 잔차 (residual) 가 부동산 총가치의 약 20% 를 차지한다. 이 80/20 이라는 수치는 전형적인 데이터 여건 아래 베이 에어리어에서 저자가 겪은 경험을 반영한 것이며, 시장을 불문하고 변하지 않는 상수로 읽어서는 안 된다. 달성 가능한 R^2 가 더 낮은 시장에서 이 틀이 제시하는 대응은, 그러한 한계를 그대로 받아들이고 더 넓은 오차범위를 보고하는 것이 아니라 측정되지 않은 가치결정요인을 찾아내어 측정하는 것이다. 해안 지역에서의 바다까지의 거리, 구릉지 주거지역에서의 표고, 그리고 MLS 데이터가 제공하지 않을 때 외부 자료원에서 수집하는 유사한 변수들이 그 예이다. 다른 분야에서는 MARS 의 잔차를 흔히 추정오차로 취급하지만, 부동산 감정평가에서 잔차는 잠재변수 가치의 간접적 척도 역할을 한다. 처음에는 실용적이지 않아 보일 수 있으나, 잔차는 각 구성요소의 기여액 합계가 잔차와 같기만 하다면 최종 감정평가 결론에 영향을 주지 않으면서 서술적 구성요소들로 의미 있게 분해될 수 있음이 증명으로 밝혀진다.

이 방법론은 저자가 20 년에 걸쳐 실증적으로 적용해 온 경험에서 비롯되었으며, R 및 Python 환경에서 MARS 를 반복적으로 구현하는 과정을 통해 발전해 왔다.

주제어: 부동산 감정평가, 주거용 감정평가, 잔차 제약 접근법, MARS, 다변량 적응 회귀 스플라인, 거래사례비교법, 잔차분석, 잠재변수, 자동 감정평가

목차

1 서론	4
2 배경	6
3 감정평가공학자	8
4 가정	9
5 RCA 작업 흐름	11
5.1 MARS 분석의 유형	11
5.2 처리 작업 흐름	11
5.3 데이터 작업 흐름	12
5.3.1 MARS 를 위한 데이터 요건	12
5.3.2 프로토콜: 종결 거래 리스팅	13
5.4 프로젝트 폴더 구조	15
5.4.1 프로토콜: 다중 MARS(earth) 실행	16
5.5 핵심 Excel 워크북	17
5.6 백업	17
6 워크북 표기법	18
6.1 서술적 차원	18
6.1.1 프로젝트 차원 (“a”)	19
6.1.2 실행 버전 차원 (“r”)	20
6.1.3 작업 흐름 단계 (“w”)	20
6.1.4 Excel 워크북 시트 (“s”)	20
6.1.5 부동산 리스팅 (“p”) — MLSData 행	21
6.1.6 MLS 변수 (“v”) — MLSData 열	21
6.1.7 최종 표기 예시	21
6.1.8 정당화	22
6.1.9 행렬은?	22
6.1.10 첫 번째 단순화	22
6.1.11 두 번째 단순화	23
7 RCA 거래사례비교표	25

8 변수	28
8.1 변수 기원 온톨로지	28
8.2 변수 용도 온톨로지	29
8.3 2 차 및 3 차 항	30
8.3.1 보정 변수	31
8.3.2 집계 변수	31
8.3.3 특수 변수 및 표기법	32
8.3.4 잔차의 분해	32
9 RCA 의 핵심 증명	33
10 MARS 모형	35
10.1 모형 변수 및 상호작용 “ g ” 함수	35
10.2 g 함수의 색인	35
10.3 그래프의 차원	36
10.4 최종 모형 식	36
10.5 모형을 그래프로 나타내는 데 필요한 차트의 수	37
11 데이터의 특성	39
11.1 잔차의 목적	39
11.2 근사 회귀변수로서의 상태 및 품질 등급	39
11.3 데이터 오류	40
11.4 정확성이란 무엇인가	41
11.5 이상치로서의 대상 부동산	43
11.6 예상 밖 데이터를 다루는 데서 CVR2 의 중요성	43
12 잔차	44
12.1 순위와 점수	44
12.2 CQA-잔차 곡선	46
12.3 CQA 곡선의 특성	46
12.3.1 CQA 곡선 구성요소의 분석	46
12.3.2 시장 동학과 사회적 함의	47
12.4 대상 부동산의 잔차	48
12.5 RCA 가치 결론	48
12.5.1 일반적 고려사항	48
12.5.2 검토, 감사, 이익	48
13 결론	50

1 서론

이 논문은 주거용 부동산을 비롯한 각종 부동산의 시장가치 (market value) 를 정확하게 추정하기 위한 현대적 접근법의 논리적, 수학적 토대를 제시한다. 이 방법은 다변량 적응 회귀 스플라인 (Multivariate Adaptive Regression Splines, MARS) 소프트웨어를 북부 캘리포니아의 주택 — 특히 샌프란시스코 베이 에어리어 일대 지역사회의 주택 — 에 적용해 온 20 년의 경험에서 개발된 것으로, 감정평가 정밀도에서 상당한 진전을 제공한다. MARS 는 부동산 특성과 가치 사이의 비선형 관계를 능란하게 포착하는 한편, 영향력이 가장 큰 변수를 효율적으로 식별하고 그 우선순위를 정함으로써 감정평가의 질을 높인다. 이러한 능력 덕분에 감정평가사 (appraiser) 는 새로운 시장 지역의 가치 동학을 전통적 방법이 허용하는 것보다 훨씬 빠르게 밝혀낼 수 있다. MARS 는 여기서 제안하는 잔차 제약 접근법 (Residual Constraint Approach, RCA) 의 중추를 이룬다. RCA 라는 이름은 다음과 같은 그 엄밀한 틀에서 비롯된다.

- RCA 는 측정된 특성과 측정되지 않은 특성 모두의 가치기여 (value contribution) 에 엄격한 제약을 부과한다.
 1. 잔차 (residual) 를 포함하여 MARS 에서 도출된 모든 가치기여의 합은 각 비교사례 (comparable) 부동산의 순 거래가격과 같아야 한다.
 2. 잔차는 하위 구성요소로 분해하여 각각에 값을 배정할 수 있으나, 그 합계는 해당 비교사례의 잔차와 같아야 한다.
- RCA 는 미적 매력, 상태, 품질과 같은 주관적 특성의 가치 총액을 잔차에 묶어 둠으로써 그 추정에서 발생하는 편의 (bias) 를 최소화한다. 유능한 감정평가공학자 (valuation engineer) 가 MARS 모형을 구축한 경우 이 잔차는 정확하게 계량된다.

주관적 특성이 시장가치에 큰 영향을 미치는 주거용 부동산을 일차 대상으로 설계되기는 했으나, RCA 는 더 넓은 범위의 자산 감정평가에도 적용될 잠재력을 지닌다. 미국 주거용 부동산 시장의 규모가 약 45 조 달러로 추정된다는 점을 고려하면 정확한 감정평가는 매우 중요하다. 패니메이 (Fannie Mae) 는 통상적인 감정평가가 실제 거래가격에서 약 5% 벗어난다고 밝히고 있으나, 감정평가사가 거래가격에 맞추라는 압력을 받는 경우가 많으므로 이 수치는 편의를 가린다. 더욱이 감정평가의 60%–70% 가 — 비교 가능한 거래가격이 존재하지 않는 — 재융자 (refinancing) 를 뒷받침하는 것이므로, RCA 가 제공하는 정밀성의 필요는 자명해진다. 연방준비제도 (Federal Reserve) 의 2012 년 분석은 2008 년 위기 이전에 이루어진 감정평가의 15%–20% 가 거래가격에서 $\pm 10\%$ 넘게 벗어났다고 추정했는데, 이 범위는 부동산 유형과 시장에 따라 다르기는 하나 오늘날의 재융자 감정평가에도 여전히 유효할 가능성이 높다. 정확성은 왜 중요한가? 담보인정비율 (loan-to-value, LTV) 이 낮은 거래에서는 — 예컨대 신용도가 우수한 매수인이 100 만 달러 주택에 50 만 달러를 계약금으로 넣는 경우 — 대출기관이 경매 (foreclosure) 위험이 미미하다고 보아 감정평가를 면제해 주기도 한다. 그러나 감정평가를 건너뛴 매수인은 과다 지급의 위험을 지며, 매수 후에야 비로소 하자를 발견하게 된다. 이는 RCA 와 같은 견고한 감정평가 방법의 가치를 부각한다. 감정평가의 정확성을 계량하는 일은 여전히 어렵다. 연구자들에게는 “시장가치” 로부터의 편차를 직접 채는 척도가 없어 간접적인 대리지표에 의존한다. 주관적 판단과 들쭉날쭉한 감정평가사 역량에 뿌리를 둔 전통적 감정평가 방법은 분석을 한층 복잡하게 만든다. 사용되는 전략에는 다음과 같은 것들이 있다.

1. Tzioumis (2016) 는 11 개 주의 감정평가사 4,782 명과 감정평가 259,436 건을 연구하여, 3 분의 1 은 계약가격의 1% 이내, 절반은 계약가격보다 1%–3% 높은 구간, 그리고 소수는 평균 5% 높은 수준이라는 사실을 발견했다. 이는 상향 편의를 시사한다. 이 연구는 감정평가사를 두 유형 — 계약가격에 근접하는 “시장가치형” 과 일관되게 더 높게 평가하는 “시장가치 초과형” — 으로 구분했다.
2. Krivorotov and LaCour-Little (2020) 는 매매 목적 감정평가의 90% 가 계약가격 이상이 있음을 발견하고, 특히 재융자 맥락에서 희망 가치에 맞추려는 경향인 “정박 편의 (anchor bias)” 를 부각했다.

이러한 연구들은 감정평가 실무에 편익이 존재함을, 그리고 그것이 전통적 도구의 한계로 증폭됨을 드러낸다. 만약 모든 감정평가사가 한결같이 유능하고 편향되지 않았다면 그들의 평가액은 수렴했을 것이다. 일부 집단에서 과대평가가 지속된다는 사실은 구조적 문제를 알리는 신호다.

위에서 인용한 편의 연구들과 그것이 가리키는 구조적 양상은 전통적 거래사례비교 실무의 더 깊은 특징을 반영하며, RCA 는 바로 이 특징을 다루기 위해 설계되었다. 전통적 방법론은 인식론적 지위가 크게 다른 두 개의 경로로 작동한다. 측정 가능한 특성 경로 — GLA(주거면적), 방 수, 대지면적, 연식, 위치 — 에서 감정평가사는 짝지은 사례 비교분석 (matched-pair analysis) 이나 비교거래사례에 대한 국지적 선형 회귀분석을 통해 보정치를 객관적으로 도출할 것이 요구된다. 유능한 감정평가사는 이 작업을 수행하며, 그 결과로 나온 보정치는 적어도 데이터에 비추어 대체로 방어 가능하다. 그러나 여기서조차 전통적 비교표는 보정치만 기록할 뿐, 가치기여를 적을 난은 아예 갖고 있지 않다. 또한 전통적 보정치는 가치기여 사이의 차액으로 계산되지도 않는다. 그것은 밑에 깔린 기여액을 한 번도 계산하지 않은 채 짝지은 사례나 유사한 절차로부터 직접 추정된다. 이와 달리 RCA 에서는 모든 보정치가 해당 특성에 대한 대상 부동산과 비교사례의 가치기여 사이의 차액이며, 증명 II 가 그 귀결을 정밀하게 밝히는 구조가 바로 이것이다. 주관적 특성 경로 — 상태, 시공 품질, 디자인, 기능적 효용, 미관 — 에서는 동원할 수 있는 뒷받침 절차가 빈약한 것에서 실낱같은 것에 이른다. 주관적 특성의 보정치는 직접적인 비교사례 집합 밖에 있는 부동산에 대한 짝지은 사례 비교분석으로, 인정된 동료 실무 관행에 대한 원용으로, 또는 감정평가사가 해당 시장에서 문서로 남긴 경험에 대한 참조로 뒷받침될 수 있다. 이 세 가지 모두 전문직 기준에는 형식상 부합한다. 그러나 그 어느 것도 측정 가능한 경로가 일상적으로 달성하는 통계적 엄밀성에는 미치지 못한다.

그 결과는 기준이 요구하는 바와 방법론이 제공할 수 있는 바 사이의 절차적 비대칭이다. 기준은 측정 가능한 주관적이든 모든 보정치에 대한 뒷받침을 요구한다. 방법론은 한 경로에서는 견고한 뒷받침을 제공하지만, 다른 경로에서는 표본 크기가 하나뿐인 일화적 짝지은 사례에서부터 다른 감정평가사들이 관례적으로 쓰는 값의 인용에 이르기까지 여러 층위의 더 약한 절차들을 내놓는다. 조망 보정치에 대해 짝지은 사례 작업을 성실히 수행하는 정직하고 유능한 감정평가사는 근거 없이 동료 관행을 인용하는 감정평가사보다 의미 있게 나온 뒷받침을 산출하지만, 표준 양식으로 두 감정평가서를 비교하는 심사자는 이 둘을 쉽게 구별할 수 없다. 양식이 기록하는 것은 보정치의 값이지 그 도출의 엄밀성이 아니기 때문이다. 주관적 특성에 대해 더 강력한 절차가 실무 감정평가사들에게 일반적으로 제공된 적이 없었기에, 방법론은 이 비대칭을 용인해 왔다.

RCA 는 두 경로 모두를 개선하지만, 그 방식은 범주적으로 다르다. 측정 가능한 경로에서 MARS 회귀분석은 짝지은 사례 비교분석이나 통상적인 선형 회귀분석이 표현할 수 없는 비선형 관계, 변수 간 상호작용, 기저함수 (basis function) 구조를 포착한다. 전통적 비교표의 39 개 비교사례가 아니라 해당 시장 지역의 적격 거래사례 전체 — 통상 수백 건 — 에 대해 적합시킨 결과, MARS 에서 도출된 가치기여는 객관적이면서 동시에 그것이 대체하는 방법론보다 실질적으로 더 정확하다. 주관적 경로에서 RCA 는 전통적 뒷받침 절차 어느 것과도 종류가 다른 일을 한다. RCA 는 개별 주관적 특성의 보정치를 추정하려는 시도 자체를 하지 않는다. 모든 주관적 특성 — 잠재변수 (latent variable) — 의 기여는 집합적으로 잔차로 나타난다. 즉 각 비교사례의 실제 순 거래가격과 MARS 가 예측한 가격의 차액이며, 이는 감정평가사의 개입 없이 데이터만으로 산출된다. 그 잔차를 디자인, 기능적 효용, 조망, 상태, 품질 등 명명된 구성요소로 분해할 때는 그 합계가 잔차 총액과 같아야 한다는 제약이 걸리며, 증명 II(9절) 에 따라 그 제약을 만족하는 어떤 분해도 동일한 가치 결론을 낳는다. 작업 흐름에 남는 판단은 비교사례들의 순위화된 잔차 분포 안에서 대상 부동산의 위치를 정하는 일이며, 이는 검토 가능한 순위표에 비추어 이루어지고 감정평가서에 명시적으로 문서화된다. 부실한 RCA 감정평가도 여전히 가능하다 — 경험이 부족한 VE 는 과적합된 MARS 모형을 만들거나, 대상 부동산을 개연성 없는 자리에 놓거나, 데이터의 이상치를 놓칠 수 있다 — 그러나 그 부실함은 표준 양식 심사로는 드러낼 수 없는 방식의 부실함이 아니라, 심사자가 모형과 순위표를 검토함으로써 짚어낼 수 있는 방식의 부실함이다.

재용자 문제를 넘어, 감정평가사는 소유자에게 정확한 평가액을 제공할 의무를 지며, 오차가 수만 내지 수십만 달러의 손실로 이어질 수 있는 고액 매수에서는 더욱 그러하다. 그럼에도 많은 이들이 감정평가에 600 1,000 달러를 쓰는 것을 주저하며 그 부가가치를 의심한다. RCA 와 같은 진보된 방법은 초기에는 자원 집약적이지만 더 높은 정밀도를 약속한다. 개선과 경험이 쌓이면 그 효

울성 — 그리고 채택률 — 은 높아질 것이다.

2 배경

저자의 학문적 배경은 수학, 컴퓨터과학, 통계학이다. 직업상으로는 소프트웨어 엔지니어이며 감정평가와 회계 분야의 부수적 경력을 갖고 있다. 2002 년부터 간헐적으로 감정평가를 수행해 왔고, 2004 년경부터는 주거용 부동산의 보정치와 가격 추세를 산정하는 데 MARS 를 적용해 왔다.

MARS 로 감정평가를 한다는 것은 감정평가에서 “깊은 경험” 을 쌓는 일이다. MARS 는 시장가치에 기여하는 변수와 그 값의 변화가 거래가격에 미치는 영향을 다른 어떤 통계적 방법보다도 상세하게 반영하는 모형을 생성한다. 따라서 MARS 로 감정평가를 하는 감정평가사는 MARS 모형이 산출하는 변수값과 거래가격 사이의 인과관계를 끊임없이 탐구하게 된다.

투명성을 위해 밝히자면, RCA 방법론은 다음과 같은 특징을 가진 주거지역에 가장 잘 적용된다.

1. 부동산들이 서로 상당히 다른 복합적 근린지역
 - 노후 주택
 - 개보수 품질의 편차
 - 다양한 건축양식과 디자인
 - GLA(주거면적), 대지면적, 표고, 조망 등 부동산 속성의 편차.
2. 복합적 시장 여건
 - 빠르게 변하는 시장 여건
 - 다양한 유형의 매수인
3. 양질의 데이터 원천: RCA 프로토콜은 데이터 마이닝에 기초하므로 풍부한 양질의 데이터를 필요로 한다. 데이터가 희박한 지역에도 쓸 수는 있으나, 소요되는 시간과 비용에 비해 그만큼 가치가 없을 수 있다.
4. 시장가치의 정확한 추정이 요구되는 경우. RCA 는 대부분의 경우 $\pm 1-2\%$ 의 정확도를 제공할 수 있어야 한다. 감정평가공학자의 보고서는 최종 가치 결론의 정확성에 관한 조건을 언급해야 한다. 데이터나 정보가 부족하거나, 데이터 마이닝을 통해 시장의 일관된 양상이 드러나지 않았다면 정확도는 떨어질 것이며, 이 사정을 의뢰인에게 알려야 한다. 구체적으로, 높은 R^2 와 CVR^2 (교차검증 R^2 . MLS 데이터 집합을 무작위 훈련 분할과 검정 분할로 반복 분할하여 계산한다), 의미 있는 MARS 모형, 그리고 순위화된 잔차에서 나타나는 부동산 CQA(상태-품질-매력) 특성의 뚜렷한 양상이 가치 결론의 정확성과 신뢰성을 판정하는 주된 수단이다.

지난 20 여 년에 걸친 이들 RCA 프로토콜의 개발은 MLSListings Inc.(Sunnyvale, CA) 및 현재 MLSListings 를 통해 접근 가능한 다른 MLS 의 이용 가능성에 기반한 것으로, 저자의 데이터 분석 경험의 95% 이상이 여기서 나왔다. 저자의 전문성은 북부 캘리포니아의 주요 15 개 카운티에 걸친 방대한 감정평가 업무와 부차적인 10 개 카운티에서의 추가 경험에서 비롯되었다.

RCA 방법론은 다음 두 가지 핵심 요건을 갖춘 시장에서 가장 효과적이다.

1. 고품질 감정평가 데이터의 풍부한 이용 가능성
2. 다음 특징을 갖는 복합적, 이질적 주거지역:
 - 다양한 건축양식
 - 개보수 정도의 편차
 - 수십 년에 걸친 개발 양상

캘리포니아 카운티 경험

이 논문에서 제시하는 서술은 다음 북부 캘리포니아 카운티들에서 부동산을 평가한 경험에 기초한다.

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

위 지역들은 실질적으로 실리콘밸리와 그 주변 카운티를 포괄하며, 센트럴 밸리와 시에라 산기슭을 포함한 보다 농촌적인 지역까지 뻗어 있다. 저자는 미국의 덜 발달된 다른 지역의 많은 감정평가사가 쓸 수 있는 것에 비해 언제나 풍부하고 상대적으로 정확한 데이터에 접근할 수 있었다는 점을 밝혀 둔다.

사용된 MLS 데이터는 RESO 인증을 받았으며 RCF(RESO Common Format) 를 준수한다. RCF 인증에도 불구하고 일부 부동산에서는 통상 부차적 중요성을 갖는 데이터가 누락되거나 부정확한 경우가 있는데, 이는 최초 세무 평가관의 오류이거나 이후 중개인의 MLS 입력 오류에서 비롯된다. 이러한 오류에는 특성이 존재하지 않을 때 아예 입력을 하지 않는 것과 같은 일정한 양상이 있는 경우가 많다. 예컨대 카포트가 없을 때 “carport spaces”(카포트 주차대수) 에 0 을 입력하는 대신 아무것도 입력하지 않는 식이다. 카포트가 드문 지역에서는 분석자가 카포트에 대한 쓸모 있는 보정치를 얻기 위해 결측 데이터를 0 으로 대체할 수 있다. R/earth 를 비롯한 MARS 구현체들이 결측 데이터를 처리할 수 있기는 하지만, 이를 어떻게 다룰지는 감정평가공학자 (Valuation Engineer, VE) 가 결정할 몫이다.

MLSListings 데이터는 플로리다에서 하와이에 이르는 다른 몇몇 주의 MLS 에서도 얻을 수 있다. 이는 프로토콜과 코드가 여러 주에서 얼마나 잘 작동하는지 시험하는 데 유용하다.

3 감정평가공학자

정의 1: 감정평가공학자

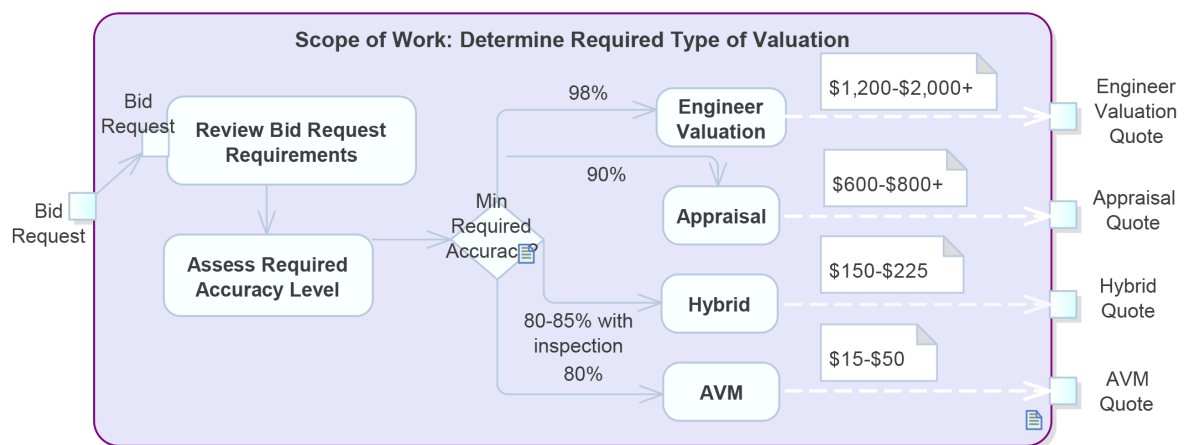
모호함을 피하기 위해, “감정평가공학자”(Valuation Engineer, 줄여서 “VE”) 라는 용어는 면허를 가진 감정평가사의 전문성에 더해 데이터 마이닝, 통계학, 프로그래밍, 상세한 프로토콜, 인공지능을 활용하여 복합적 부동산의 시장가치를 판정하는 데 관한 고급 지식과 기술, 경험을 겸비한 사람을 가리킨다. 아울러 감정평가는 사기 가능성이 이미 확인된 영역이므로 어떤 형태로든 면허를 요하는 업무라는 점을 인식하는 것도 중요하다.

대출기관은 통상 다음과 같은 몇 가지 감정평가 선택지를 제시한다.

1. 자동가치산정모형 (Automated Valuation Models, AVM) — 비용이 가장 낮은 선택지
2. 하이브리드 감정평가 — 무면허 조사자를 활용하되 면허 감정평가사가 감독
3. 표준 감정평가 — 면허 감정평가사가 수행

감정평가 업무에서 통상 고려 대상에 들지 않는 것이 공학자 수준의 감정평가 (Engineer-level Appraisal) 다. 이를 수행할 기술 역량을 갖춘 감정평가사가 애초에 많지 않다. 그러나 그 수요는 마땅히 있어야 한다. 이러한 감정평가는 RCA 방법론 (MARS 와 제약) 을 활용함으로써 현저히 높은 정확도를 제공하고 상당히 양호한 객관성을 거의 보장할 수 있다.

신규 택지개발지구의 규격화된 부동산에는 표준 감정평가로 대체로 충분하다. 그러나 30 년 이상에 걸친 상당한 개보수 이력이 있는 지역의 노후 주택에는 공학자 감정평가가 권장된다. 두 방법 사이의 선택은 부동산의 특성과 시장 여건에 기초한 체계적 의사결정 절차를 따라야 한다.



4 가정

가정 1: 지리적 범위 및 부동산 유형

이 논문의 범위는 주로 캘리포니아의 대도시권 및 도시 지역에 있는 단독주택과 콘도미니엄에 초점을 맞춘다. 거래사례비교법은 다른 부동산 유형과 지역에도 적용되지만, 농지, 다세대 부동산, 상업용 부동산을 평가할 때는 통상 수익방식과 원가방식이 더 지배적인 역할을 한다. 그러한 경우 RCA 접근법은 수정이 필요하거나 적용 범위가 제한될 수 있다.

가정 2: 데이터 저장

CSV와 Excel은 MLS 시스템에서 널리 사용되며 RCA 프로그램에 데이터를 공급하는 주된 방식으로 상정한다. 다만 브로커 피드를 이용해 SQL 데이터베이스로 데이터를 지속적으로 공급할 수도 있다. 백업과 중간 저장에는 SQL 데이터베이스가 자주 쓰이는데, 그중 큰 역할을 하는 것이 무료 오픈소스인 SQLite와 PostgreSQL이다. 보고서 조각에는 텍스트나 RTF 파일이, 그래픽에는 PNG, JPG, TIFF 파일이 사용된다. Microsoft의 SQL Server를 쓰는 이들도 많지만, 무료판인 “SQL Server Express”는 제약이 매우 크고 (예컨대 CPU 코어를 하나만 쓸 수 있다) 다른 판본은 상업적 이용 시 상당한 라이선스 비용이 든다. 캘리포니아의 부동산 데이터베이스는 매우 클 수 있고 조인(join)에 시간이 오래 걸릴 수 있으므로, 대규모 조인에는 8 코어 이상에 128G RAM을 갖춘 컴퓨터가 유용하다. SQL 데이터베이스는 임의로 경계 지어진 구역 안의 모든 부동산을 찾아야 할 때도 중요하다. 따라서 예컨대 군집분석(Cluster Analysis)으로 근린지역을 정의했다면 SQL 데이터베이스를 쓰게 될 것이다. 그러나 데이터베이스마다 지리정보를 다루는 내장 함수가 달라 서로 호환되지 않는다. RCA에 관한 이 연재에서는 복잡한 SQL 작업에는 PostgreSQL을, 단순한 작업에는 SQLite를 전제한다.

가정 3: MARS 처리에는 “R/earth”를 사용한다.

2017년까지 Salford Systems의 MARS 소프트웨어는 연 270 달러로 감당할 만한 해결책이었다. 그러나 Minitab이 Salford Systems를 인수한 뒤 MARS는 연 16,000 달러의 SPM 패키지에 묶였고, 이는 대부분의 소규모 사업자에게는 감당하기 어려운 가격이다. 다행히 Stephen Milborrow가 관리하는 R 프로그래밍 언어의 “earth” 패키지 (Milborrow 2024a)가 Minitab SPM MARS의 기능에 필적하거나 그것을 능가하는 수준으로 발전했다. earth 패키지에는 추가적인 이점도 있다. 무료로 이용할 수 있고, R의 포괄적인 통계 생태계와 매끄럽게 통합되며, Pandas를 통해 Python에서도 접근할 수 있어 다양한 작업 흐름에 뛰어난 유연성을 제공한다. MARS에 관한 논의는 R/earth의 사용을 전제한다.

가정 4: 표기법

CSV 파일이나 Excel 스프레드시트와 같은 외부 원천에서 온 원자료는 표준 라틴 문자로 표기한다. 여기에는 MLS(Multiple Listing Service) 부동산 데이터, R/earth 패키지를 이용한 MARS(다변량 적응 회귀 스플라인) 설정값, 그리고 프로젝트별 데이터가 포함된다. 이 표기법은 R이나 Python에서 그에 대응하는 데이터 프레임 표현에도 적용된다. 부동산 특성과 그 여러 유형 및 부분집합을 논할 때는 정밀성과 분석적 명료성을 높이기 위해 그리스 문자로 표기를 바꾼다. 이러한 수학적 표기가 더 엄격해 보일 수 있으나, 여기에는 추상적인 부동산 특성을 그 원자료 표현과 구별한다는 중요한 목적이 있다. 관행상 그리스 문자는 일관되게 부동산 특성을, 라틴 문자는 스프레드시트 데이터 구조를 가리킨다. 예컨대 기호 M 은 행에 부동산을, 열에 그에 대응하는 특성을 배치한 MLS 데이터 행렬을 나타낸다.

가정 5: 프로그래밍 언어

이 분야에서는 R 과 Python(w/Pandas) 이 지배적인 언어이며, 속도가 중요할 때는 C 나 C++ 가 쓰인다. R/earth 자체도 R 과 C 로 작성되어 있다.

가정 6: 색인

색인 방식은 프로그래밍 언어와 수학에서 서로 달라 혼동을 일으킬 수 있다.

- 수학과 R 은 색인이 1 에서 시작한다.
- C, C#, C++, Python 은 0 에서 시작한다.

이 논문에서는 일관성과 단순성을 위해 모든 색인을 (Python 이나 C 처럼) 0 에서 시작한다. 구체적인 방식은 다음과 같다.

- 리스트, 행렬, 데이터 프레임의 경우:
 - 색인 0 은 “대상 부동산”(연구 대상이 되는 주된 부동산) 에 예약된다.
 - 예: 행이 부동산인 표에서 행 0 은 대상 부동산이고, 열 0 은 모든 부동산의 고유 ID 를 담는다.
 - 실무 요령: 대상 부동산을 맨 앞 (행 0) 에 두고, 나머지는 거래일이 최근인 것부터 오래된 것 순으로 나열하며 순차적으로 (0, 1, 2 등) 번호를 매긴다.
- 대상 부동산이 없는 경우:
 - 데이터 구조에 부동산이 포함되지 않는다면 (예컨대 변수나 함수만 있는 경우) 색인 0 은 건너뛰고 1 에서 시작한다.
 - R 사용자를 위한 참고: R 은 1 기반 색인을 쓰므로, 이 논문의 수식을 R 로 코딩할 때는 0 기반에서 1 기반으로 맞추기 위해 모든 색인에 1 을 더해야 한다.

5 RCA 작업 흐름

RCA 작업 흐름을 이해하는 것은 효과적인 분석에 필수적이다. 이 절차는 분석 과업과 의사결정 지점의 구조화된 순서를 따른다

5.1 MARS 분석의 유형

MARS 분석의 작업 흐름은 시장분석만 필요한지, 아니면 특정 부동산의 평가나 감정평가가 요구되는지에 따라 달라진다. 특정 부동산의 감정평가에는 통상 그 시장 지역에 대한 분석이 필요하다는 점에 유의하라.

- 시장분석 (Market Analysis): 카운티, 시, 또는 획정된 권역과 같은 넓은 지역에 초점을 맞춘다. 이 접근에는 특정 대상 부동산이 없으므로 가치 결론도 필요하지 않다. 작업 흐름은 모형 개발, 검증, 그리고 뒷받침 분석자료의 생성을 중심으로 한다. 최적 모형이 식별되고 문서화되면 절차는 종료된다.
- 단일 부동산 분석 (Single Property Analysis) 은 시장분석의 모든 요소에 더해 특정 부동산의 가치 판정을 포함한다. 이는 전통적 감정평가 방법과 근본적으로 다르다. 전통적 접근에서 가치는 통상 39 개의 비교 거래사례를 보정하고 가중평균을 계산하여 도출한다. 이와 달리 RCA 는 주로 다음 두 요소를 통해 가치를 판정한다.

1. MARS 모형의 예측가치
2. 추정된 부동산별 잔차 (그리고 그에 결부된 CQA 점수)

RCA 에서 거래사례비교표 (Sales Grid) 는 주로 변수별 기여를 보여줌으로써 대상 부동산과 비교사례 사이의 가치 차이를 설명하는 역할을 한다. RCA 의 핵심 장점은 안정성이다. 의뢰인이 비교사례의 선정을 두고 이견을 제기할 수는 있으나, 제시를 위해 어떤 비교사례를 고르든 가치 결론은 일관되게 유지된다. 모든 거래사례가 (그것이 15 건이든 600 건이든) 동일한 모형으로 보정되기 때문이다. 오래된 거래도 현재의 시장 여건에 맞추어 자동으로 보정되지만, 일반적으로는 최근 거래가 바람직하다. 기간이 길어질수록 측정되지 않은 시장 특성의 비중이 커지므로, 모형이 가격 요인을 포착하는 효력은 시간적 거리가 멀어질수록 약해진다.

5.2 처리 작업 흐름

전체 처리 작업 흐름은 다음 단계들로 서술할 수 있다.

1. 업무 범위 (Scope of work).
2. 초기 프로젝트 폴더와 파일 설정.
3. MLS 데이터 다운로드. 필요한 부동산 레코드 수와 기간 범위 제한 설정에 관해서는 프로토콜 1.1-1.3 을 참조하라.
4. 설정 시트 완성: 회귀 변수, 상호작용, 집계, 계산.
5. MARS(R/earth) 회귀분석을 실행하여 가격 모형 생성.
6. MARS 모형으로부터 부동산 가치기여 및 잔차 (순 잔차와 잔차/SF) 생성.
7. 잔차/SF 로 부동산 순위 매기기.
8. 각 비교 거래사례의 CQA 점수 산출.

9. 부동산의 전반적 매력을 참되게 대표하는 품질 순위표가 산출되었다는 전제 아래, 순위표의 전반적 양상과 추세를 검토하여 대상 부동산의 최적 위치를 찾고 CQA 점수를 배정한다.
10. 그에 대응하는 잔차/SF 를 잔차/SF · (대상 부동산) GLA 의 형태로 대상 부동산에 배정한다.
11. **MARS** 가격 추정치와 배정된 잔차를 더하여 대상 부동산의 가치 결론에 이른다.
12. 대상 부동산의 잔차를 서술적 구성요소와 가치기여로 분해한다.
13. 비교사례에 대해 모형이 산출한 가치기여와 보정치, 그리고 모든 소계와 보정 후 거래가격을 계산한다.
14. 보정 후 거래가격은 대상 부동산의 최종 가치 결론과 같아진다.
15. 거래사례비교표에 넣을 가장 유사한 부동산 6 12 건을 선정한다.
16. 비교사례의 잔차 총액을 잔차 구성요소별 가치기여로 분해하고, 분해가 끝나면 잔차 총액 항목의 값을 0 으로 설정한다.
17. 모든 소계와 거래사례비교표의 비교사례 보정 후 거래가격을 재계산한다. 값을 확인한다.
18. 거래사례비교표를 보고서에 올린다.

RCA 에서 최종 가치 결론은 거래사례비교표의 비교사례 값들로부터 직접 도출되는 것이 아니라는 점에 유의하라. RCA 방법에서 거래사례비교표는 대상 부동산과 비슷해 보이는 부동산들이 왜 최종 가치 결론 또는 대상 부동산의 평가액보다 비싸게 또는 싸게 팔리는지를 설명하는 도구일 뿐이다. 이는 달리 보아야 한다. 대상 부동산의 가치 결론은 MARS 모형의 가격 추정치와 감정평가공학자가 추정한 잔차에 달려 있다. 그런데 이 두 양은 R/earth 회귀분석에 투입된 모든 부동산에 크게 의존한다. 그 회귀분석이 부동산 잔차의 순위표를 만들어 냈고, 그 순위표가 대상 부동산의 CQA 점수를 결정했으며, 감정평가공학자가 순위표 안에서 위치를 정함으로써 대상 부동산의 잔차가 확정되었기 때문이다. 거래사례비교표의 보정치는 실은 RCA 가 부과한 수학적 제약에 의해 결정된다. 잔차의 분해는 다소간 VE 의 재량에 맡겨져 있지만, 분해된 가치기여의 합계는 언제나 그에 결부된 잔차와 같아야 한다. 증명 2 로부터 우리는 잔차 구성요소의 실제 값이 VE 의 판단에 따라 달라질 수 있더라도 특정 부동산에 대한 그 합계는 완전히 제약되어 있으며, 따라서 VE 가 잔차 분해를 어떻게 바꾸든 해당 부동산의 보정 후 거래가격은 달라지지 않는다는 것을 안다.

5.3 데이터 작업 흐름

이 논문에서는 RCA 방법의 상위 수준 작업 흐름만을 제시한다는 점에 유의하라. RCA 분석의 여러 단계와 하위 단계는 다른 논문에서 다룬다.

5.3.1 MARS 를 위한 데이터 요건

거래 개시일과, 감정평가 또는 가치산정의 기준일에 근거한 종료일을 지정해야 한다. 이때 보통 필요한 회귀 변수의 추정 개수와 1 을 넘는 각 상호작용 차수별로 필요한 상호작용의 개수를 고려한다.

5.3.2 프로토콜: 종결 거래 리스팅

프로토콜 1.1: 필요한 MLS 종결 거래 리스팅 건수의 결정

이 프로토콜은 MARS 회귀분석에 투입할 종결 거래 리스팅의 권장 최소 건수를 계산하는 방법을 설명한다. 우선 대략적인 추정을 한 다음, 1 차, 2 차, 경우에 따라 3 차 차수로 MARS 를 실행하여 초기 모형을 생성하고, 아래 규칙에 따라 행 수를 검토하여 더 많은 행이 필요한지 판단해야 한다. 이 프로토콜은 독립변수가 100% 독립임을 전제하는데, 부동산에서는 그런 경우가 드물다. 예컨대 GLA 와 방 수는 대체로 상관관계가 높고, GLA 는 대지면적과도 상관을 보일 가능성이 크다. 여기서 “자유도 (Degrees of Freedom, DOF)” 의 계산 문제가 나온다. 참으로 독립적인 변수 하나가 자유도 하나를 갖는다. 그러나 이른바 독립변수들 사이에 종속관계가 있으면 정확한 DOF 계산에는 이 논문의 범위를 넘어서는 분석이 필요하다. 변수 간 공선성이 있을 가능성이 크므로 아래 추정치는 아마도 다소 높은 편이라는 점만 유념하라.

1. 1 차 항 (상호작용 없음) 에 대해서는 변수당 ~ 15 행 규칙이 잘 통한다
2. 허용되는 2 원 상호작용 하나마다 부동산 약 15 30 건을 추가해야 한다
3. 고차 상호작용 (변수 3 개 이상) 에는 각각 30 건 이상의 부동산이 필요하다.

다만 이것이 절대적인 규칙은 아니다. 실제로 필요한 건수는 다음에 좌우될 수 있다.

- 데이터의 신호 대 잡음 비율
- 모형화하려는 관계의 복잡성
- 반응변수의 변동성
- 데이터의 품질

예:

다음과 같은 경우라면:

- 주요 변수 10 개
- 허용되는 2 원 상호작용 5 개
- 허용되는 3 원 상호작용 2 개

최소 거래 건수는 다음과 같이 계산할 수 있다.

$$\begin{aligned} N &= (\text{변수 10 개} * \text{변수당 15 행}) + \\ &\quad (2 \text{ 원 상호작용 5 개} * 2 \text{ 원 상호작용당 22 행}) + \\ &\quad (3 \text{ 원 상호작용 2 개} * 3 \text{ 원 상호작용당 30 행}) \\ &= 320 \text{ 행} \end{aligned}$$

또는 다음과 같이 할 수도 있다. 어떤 데이터 집합에 R/earth 를 실행한 뒤, 모형의 기저함수 개수에 1 을 더한 값을 자유도의 추정치로 삼고, 여기에 20 30 을 곱하여 필요한 입력 레코드 수를 구할 수 있다.

프로토콜 1.2: 거래 기간은 매수인 선호의 변화가 최소인 구간을 포괄해야 한다

과거 부동산 데이터를 수집할 때, 시장 선호의 변화를 적절한 변수로 포착할 수 있다면 엄격한 기간 제한은 없다. 해당 시장의 건축양식 변천을 다룰 때는 다음의 핵심 사항을 고려하라. 지배적 건축양식의 변화 (예컨대 California Ranch 에서 Mediterranean 으로 이행) 는 다음을 이용해 모형화할 수 있다.

- 택지개발지구 식별자
- MLS 지역 코드
- 건축연도 구간

그러나 이러한 양식 변화를 특정 변수로 충분히 포착할 수 없다면 중요한 제약에 부딪힌다. 거래일 (Date of Sale) 은 통상 시장 여건을 보정하는 데 쓰이는데, 이를 동시에 건축 선호를 통제하는 데까지 쓰면 이 두 개의 별개 효과가 뒤섞이게 된다. 그러한 경우 바람직한 접근은 다음 중 하나다.

1. 데이터 수집을 새로운 건축양식이 도입된 이후 기간으로 한정한다
2. 비교사례 선정을 동일한 건축양식을 공유하는 택지개발지구로 제한한다

이렇게 하면 시간에 따른 시장 여건의 적절한 보정은 그대로 허용하면서도 부동산 특성의 일관성이 유지된다. 아니면 적어도 이 경우에는, 데이터에서 건축양식을 추출할 수 있다는 전제 아래 이 두 변수 사이의 상호작용을 고려해 볼 수도 있다. 그러나 그러한 추출이 언제나 가능한 것은 아니다.

프로토콜 1.3: 거래가 적은 지역이라면 할 수 있는 최선을 다하라

거래 데이터로 MARS 모형화를 할 때 의미 있는 분석을 위해서는 대략 15 건의 거래가 필요하다. 거래가 그보다 적다면 (예컨대 부동산 5 건) 각 레코드를 두 번씩 복제하여 15 건 기준선에 이르게 할 수 있다. 이 접근에서 중요하게 고려할 사항은 다음과 같다.

1. 과적합을 줄이기 위해 1 차 회귀만 사용한다
2. 데이터 집합이 지나치게 제한적이므로 교차검증은 무시하거나 건너뛰다
3. 이러한 상황에서도 MARS 가 짝지은 사례 비교나 표준 선형 회귀분석보다 나은 성능을 보일 가능성이 크지만, 데이터 복제는 불확실성 측정치와 통계적 유의성의 타당성에 영향을 준다는 점에 유의하라

장기적으로 가장 좋은 해법은 실제 거래를 추가하여 데이터 집합을 늘리는 것이다. 그러나 제한된 데이터로 진행할 수밖에 없다면, 통계적 한계를 인정하는 전제 아래 데이터 복제가 쓸 만한 임시 해법이 된다.

5.4 프로젝트 폴더 구조

프로젝트 폴더 구조를 논해 두는 것이 좋겠다. 시장분석과 단일 부동산 감정평가는 별도의 폴더를 두는 것이 좋다. 두 폴더는 유사한 하위 폴더 구조를 갖는다.

```

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
│   │       │   │       └── ...
│   │       └── ...
│   └── ...
└── ...

```

1 단계 처리는 버전 폴더와 그 하위 폴더들을 생성한다. 이때 MLS_Original.xlsx 를 복사하는데, 이 스프레드시트는 특정 일시에 MLS 에서 내려받은 내용을 그대로 반영하며 결코 변경되어서는 안 되는 것이지만, 시간이 긴 프로젝트에서는 새로운 MLS 다운로드가 필요해질 가능성이 있기 때문이다. 다음으로 이 단계는 설정 파라미터가 포함된 확장 MLS.xlsx 워크북의 사본을 버전 폴더에 만든다. MARS 를 완전히 한 번 실행할 때마다 하나씩, 버전 폴더는 여러 개가 있을 수 있다.

이어지는 단계들은 워크북에 열을 추가하고 정렬하는 등의 변경을 가하며, 감정평가가 완료되면 갱신된 판본을 MLS_Stage(N) 으로 프로젝트 데이터 폴더에 기록한다. 다시 말하면, 루트의 MLS.xlsx 폴더는 끊임없이 변화하며, 버전 사본은 변수의 삭제와 추가를 비롯한 여러 변경을 거치는, 원본 MLS 워크북의 진행 중인 변형이다. 처리용 R 또는 Python 코드는 대개 Data 의

MLS.xlsx 워크북을 입력으로 실행된다. 따라서 RCA 실행의 각 버전 뒤에 있는 모든 데이터의 사본이 언제나 보존된다.

더는 개선이 불가능해질 때까지 필요한 만큼 여러 버전을 만들고, 그 시점에서 최선의 버전을 최종본으로 선정한다. 디렉터리를 훑어볼 때 바로 알아볼 수 있도록 폴더 이름 끝에 “_Final” 을 붙일 수도 있다. 흥미롭거나 경합하는 모형을 담은 옛 버전 일부는 대체로 보존해 두는 것이 좋다. 각 버전 폴더에는 보고서에 삽입할 수 있는 여러 텍스트, 도표, 그래프 파일이 들어 있다. 최종 MLS_Stage5.xlsx 워크북에는 보고서에 필요한 만큼의 시트가 들어 있다. 특히 Fannie Mae 거래 사례비교표로 옮길 스프레드시트가 있을 수 있다. 또한 집계의 계산 과정이나 그 분해도 반드시 포함된다. 어려운 시장 지역에서 완벽한 모형을 찾아내려는 이 씨름은 RCA 접근법에서 가장 시간이 많이 드는 부분일 수 있다.

거래사례비교표를 재현하는 데 필요한 모든 것을 버전 폴더에 복사해 두어야 하며, 코드가 자주 바뀌는 경우에는 사용된 R 또는 Python 코드도 포함된다.

5.4.1 프로토콜: 다중 MARS(earth) 실행

프로토콜 2.1: 시장 지역의 가격 요인을 정밀하게 파악하기 위한 과적합 실행

초기 earth 실행 단계에서는 밑에 깔린 가격 요인을 발견하기 위해 과적합을 전략적으로 활용해야 한다. 예컨대 거래 800 건으로 대표되는 시장 지역에 작업장 (workshop) 이 딸린 부동산이 단 3 건뿐일 수 있는데, 그 소수 사례에서는 작업장이 최종 거래가격에 큰 영향을 줄 수 있다. 전진 단계만 실행하면 이러한 작업장의 대략적인 가치기여를 보여주는 과적합 모형이 산출된다. 그러나 800 건 중 그러한 거래가 3 건뿐이므로 작업장 항은 후진 단계에서, 그리고 교차검증에서는 확실히 제거될 것이다. 전진 단계만의 실행이 작업장에 부여한 값은 기록해 두었다가 잠재변수 분석으로 옮겨, 거래사례비교표에 등장하는 작업장 보유 부동산에 사용할 수 있다. 이러한 과적합 모형은 후진 단계의 가지치기나 교차검증 실행에서 제거될 다른 여러 희소 데이터 특성에 대해서도 상세한 추정치를 제공한다.

이러한 탐색적 실행은 각각 별도의 버전 폴더로 저장하라. 기록된 가치기여는 나중에 잔차를 분해할 때 VE 가 인용하고자 할 증거이다. 물론 VE 는 회귀분석이 한 번도 보지 못한 부동산 — 이후에 등장하는 대상 부동산 같은 — 에 대해서도 모형이 견고하도록 궁극적으로는 가지치기와 교차검증에 의존해야 한다는 점을 이해해야 한다. 과적합 실행은 연구를 위한 도구일 뿐 결코 최종 모형이 아니다.

프로토콜 2.2: 서로 다른 earth() 파라미터 아래에서의 다중 교차검증 실행

시장 지역은 구조와 규모가 서로 다르므로, VE 는 대개 일련의 실행을 통해 실험해야 한다. 어떤 변수가 가격에 가장 큰 영향을 주는지 이해하기 위해서, 기존의 측정 변수 집합이 가치를 추정하는 데 충분한지 아니면 새로운 데이터 (수변까지의 거리, 표고, 상업시설까지의 거리 등) 를 추가해야 하는지 판단하기 위해서, 그리고 비효율적이거나 오도하는 상호작용은 배제하면서 새로운 종류의 상호작용을 발견하기 위해서다. 최종적으로 VE 가 원하는 것은 훈련 데이터에 대한 양호한 R2 와 검정 데이터에 대한 최대한 좋은 CVR2 이다. 그에 이르려면 새로운 데이터를 수집하거나, 기존 데이터의 정확성을 높이거나, 시장 지역의 규모와 성격에 맞게 (예컨대 가용 자유도에 맞추어 — 프로토콜 1.1 참조) earth() 파라미터를 조정해야 할 수 있다.

위의 프로젝트 구조가 다수의 실행 버전 폴더를 지원하는 이유가 바로 이것이다. 각 후보 모형이 그 파라미터 및 R2/CVR2 와 함께 보존되므로 여러 실행에 걸친 궤적을 비교할 수 있다. R2 는 오르는데 CVR2 는 정제하거나 떨어지는 버전은 시장 신호가 아니라 잡음에 적합한 것이며, 그 사실을 드러내 주는 것이 바로 버전 이력이다.

프로토콜 2.3: Remarks 및 서술적 특성 열의 Prolog 확장

아직 다소 실험적인 단계에 있기는 하지만, earthUI 에는 이제 내장 Prolog 엔진 (vProlog) 이 탑재되어 있으며 확정절 문법 (Definite Clause Grammars, DCG) 을 이용해 MLS 의 자유 서술 열 — 무엇보다 Public Remarks, 그 밖에 중개인 remarks 와 방 특성이나 기존 구조물 같은 심표 구분 특성 목록 — 을 파싱한다. 이 확장은 “owned solar”, “3 car garage”, “40x50 workshop” 과 같은 표현을 회귀분석에 투입할 수 있는 통상의 특성 열로 바꾸어 주며, 이렇게 복원된 희소 데이터 특성은 잠재변수 분석에 곧바로 반영된다 (뒤에 나오는 근사 회귀변수로서의 상태 및 품질 등급에 관한 소절을 보라. 이 기능 자체는 이번 호의 earthUI 관련 논문에서 서술되어 있다).

텍스트 확장은 후보 예측변수 집합을 크게 늘리며, 그에 따라 검증을 위한 다수의 earth/RCA 실행이 필요해진다. 확장된 열은 모형에 넣고 빼기를 시험해 보아야 하고, 공선적인 경우에는 파생 계수나 불리언으로 통합해야 하며, 잡음이 아니라 시장 신호를 담고 있는지 CVR2 로 확인해야 한다. 그러한 실험 각각을 프로토콜 2.1 및 2.2 에서와 똑같이 별도의 실행 버전 폴더로 다루어라. 또한 파싱 문법 자체가 개발 중이므로, VE 는 추출된 열을 모형에서 신뢰하기에 앞서 원문과 대조하여 검토해야 한다.

5.5 핵심 Excel 워크북

대부분의 MLS 서비스가 부동산 데이터를 Excel CSV 워크시트로 내려받게 하므로, 데이터와 각종 설정 파라미터를 저장하는 데는 Excel 이 아마도 가장 편리한 수단일 것이다. 프로젝트마다 MLS.xlsx 라는 핵심 Excel 워크북이 하나씩 있으며, 여기에는 MLS 에서 내려받은 종결 부동산 거래 리스팅을 저장하는 “MLSDData” 시트 하나와 여러 개의 설정 및 파라미터 시트가 들어 있다.

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

원자료는 MLS_Original.xlsx 같은 이름의 워크북에 저장한다. 여기서 다른 이름, 이를테면 “MLS.xlsx” 로 사본을 만들어 R 또는 Python RCA 프로그램에 적재한다. 이 두 워크북은 모두 Data 폴더 아래에 둔다. Data/MLS.xlsx 워크북은 변수의 여러 조합, MARS 파라미터, 그 밖의 설정 변경을 시험하며 아마도 여러 차례의 실행을 거치게 된다. 각 회차는 VER_YYYYMMDDHHMMSS 형식으로 날짜와 시각을 이름에 담은 새 입출력 폴더를 만든다. 각 폴더에는 입력 MLS.xlsx 워크북의 사본과 네 개의 처리 단계 (2-5 단계) 각각에 대해 생성된 Excel 워크북, 그리고 출력 텍스트 문서, 도표, 그래프가 들어간다. 최종 단계인 5 단계에는 의뢰인의 유형과 SOW(업무 범위) 에 따라 보고서용 시트가 여럿 더 포함될 수 있다.

5.6 백업

새로 작성했거나 수정한 Excel 스프레드시트를 프로젝트 버전 폴더에 기록할 때, 프로젝트 폴더안의 SQLite 나 PostgreSQL 데이터베이스에도 백업을 만들 수 있다. 데이터베이스 파일은 이후 검토나 참조 과정에서 의도치 않게 손상될 가능성이 더 낮다.

6 워크북 표기법

6.1 서술적 차원

워크북 “M” 에는 대략 여섯 개의 서술적 차원이 있다.

1. 감정평가 프로젝트 “a”
2. 실행 버전 “r”
3. 작업 흐름 단계 “w”
4. 시트 “s”
5. 부동산 리스팅 “p”
6. 변수 집합 “v”

따라서 특정 워크북 M 안의 개별 셀은 다음과 같이 서술할 수 있다.

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

(참고: “*” 는 유효한 임의의 색인을 뜻한다)

또는:

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

여기서:

- n_p = 행의 수 (시트에 따라 다름)
- n_v = 열의 수 (시트에 따라 다름)
- n_r = 실행 버전의 수
- n_m = 작업 흐름 단계의 수
- n_g = 분석 프로젝트의 수
- n_h = 워크북 시트의 수

“W” 는 각자 자기 표를 갖는 여섯 개의 핵심 데이터 시트로 이루어진 Excel 워크북 전체를 가리키며, 연관된 R 또는 Python 프로그램을 실행하면 이들이 각각 별개의 내부 데이터 프레임으로 적재된다는 점을 이해하라. R 과 Python 모두 이름을 받아 그것이 데이터 프레임의 어느 행 또는 열에 해당하는지 찾아 주는 내부 함수를 갖고 있다. 예컨대 아래 첨자 v_j 에서 $v = (“GLA”, “LotSize”, “BathRms”, “BedRms”, ...)$ 이고 $j=0$ 이면 $v_j = “GLA”$ 이다. R 이든 Python 이든 내부적으로 “GLA” 를 그것이 저장된 열 — 예컨대 해당 데이터 프레임의 15 번째 열 — 로 해석해 낼 수 있다. 따라서 이 논문에서는 이름 배열에 색인을 걸고, 그 이름이 실행 중인 프로그램에 전달되면 프로그램이 그 이름을 내부 데이터 프레임의 실제 열 또는 행 색인으로 다시 변환한다. 다소 복잡하기는 하다. 그러나 정밀을 기하자면 이것이 필요하고 또 가장 편리하다. 이는 분석자 역할과 개발자 역할 사이의 “책임 분리” 인데, 실제로는 감정평가공학자 (VE) 가 두 역할을 모두 수행할 수도 있다. R 의 “data frame” 은 Python 에서 “dataframe” 이라 불린다는 점에 유의하라.

또한 MARS 실행 파라미터 시트는 W 의 차원이 아니라는 점에 유의하라. W 는 MARS 에 투입되는 데이터의 버전에 관한 것이고, MARS/earth 가 그 데이터를 어떻게 다룰지를 정하는 파라미터는 그와 별개이기 때문이다.

명료성을 위해, 데이터 워크북 시트가 적재되는 R 또는 Python 내부 데이터 프레임과 보다 직접적으로 대응하는 복합적 다차원 표기법을 워크북에 대해 제시한다. “무거운” 표기법이다. 그러나

증명 등에 들어가면 우리가 다루는 대상에 관한 가정을 둬으로써 이 짐의 상당 부분을 덜어낸다. 그럼에도 이 표기법은 처음부터 작업 흐름과 데이터의 전체 그림을 파악하는 데 도움이 된다.

우리가 처리할 워크북 집합에 대한 표기법에서 시작하자.

6.1.1 프로젝트 차원 (“a”)

업무 범위 (SOW) 가 승인되고 나면, 설정 파라미터, 출력 그래프, 도표, 스프레드시트, 보고서 조각을 포함한 입출력 데이터를 저장할 프로젝트 폴더를 마련해야 한다. VE 는 보고서가 분실되었을 때 그것이 곧바로 드러나도록 모든 프로젝트를 정수와 같은 단순한 순차 ID 로 식별해야 한다. 프로젝트가 취소되는 경우에는 해당 프로젝트 폴더 안에서 취소되었음을 일관되게 표시할 수단이 있어야 하되, 폴더를 삭제해서는 결코 안 된다. 다른 방식도 가능하지만, 여기서는 프로젝트 ID 가 0 에서 시작하여 1 씩 증가한다고 가정한다. 또한 유효한 모든 프로젝트 ID 가 처리 프로그램으로 생성된 순서집합 “a” 에 저장되어 있다고 가정한다.

감정평가 프로젝트 데이터는 워크북의 프로젝트 워크시트에 있으나, 중앙 SQL 데이터베이스에 저장할 수도 있다.

프로젝트 필드로는 다음을 제안한다.

1. Local ID: 회사 단위로 순차 생성되는 짧은 정수 ID. 저장용 기본키로 삼기에 좋으며, 프로젝트가 하나씩 추가될 때마다 데이터베이스가 자동으로 생성할 수 있다.
2. Global GUID: 원한다면 다음을 이용해 GUID(전역 고유 ID) 를 생성할 수 있다.

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

3. 부동산 소재지 (시장분석) 또는 주소
4. 감정평가 기준일
5. 의뢰일
6. 임장활동일
7. 최초 보고서 작성일
8. 1 차 수정 보고서 작성일
9. 2 차 수정 보고서 작성일
10. 3 차 수정 보고서 작성일
11. 의뢰인명
12. 의뢰인 주소
13. 대출기관명
14. 대출기관 주소
15. 소유자명
16. 소유자 주소
17.

프로젝트 ID 목록의 예:

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

6.1.2 실행 버전 차원 (“r”)

작업 흐름의 두 번째로 입력 데이터에 MARS 회귀분석을 수행한다. 모형을 개선하기 위해 여러 파라미터를 바꾸어 가며 몇 차례에서 여러 차례에 이르는 실행을 하게 될 것이다. 각 실행은 실행 일시에 기반한 이름으로 출력용 새 버전 폴더를 생성한다. 예를 들면,

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

```
latest_version : r1 = "ver20241210_083124"
```

경쟁력 있는 모형과 출력을 가진 유망한 버전은 최종본이 결정될 때까지, 또는 쓸모가 있다면 그보다 더 오래 보존해야 한다. 최종 버전임을 표시하기 위해 폴더 이름 끝에 “Final” 을 붙여 수정할 수도 있다.

6.1.3 작업 흐름 단계 (“w”)

RCA 작업 흐름은 아래에 열거한 것처럼 통상 다섯 개의 개별 단계로 나뉜다. 각 단계는 다음 단계를 시작하기 전에 성공적으로 완료되어야 한다. VE 는 한 단계가 끝난 지점에서 처리를 멈추고 결과를 검토한 뒤 진행할 수 있어야 한다. 시장 지역은 상당히 복잡하고 종종 매우 특이하기 때문에 이것이 필요하다. 새로운 접근을 요하는 이례적 문제에 부딪힐 수 있다. VE 는 R 이나 Python 코드에 뛰어들어 필요한 수정을 가하고 나아가 새로운 접근을 시도할 수 있어야 한다. 그러한 경우 작업 중인 동일 단계를 여러 번 실행하면서 다양한 접근을 실험해 보고 싶을 수 있다. 다만 VE 는 — 몇 가지 단서 (예컨대 3 차 집계는 2 단계의 이전 실행에서 이미 결정되어 있어야 한다) 와 함께 — 명령 하나로 모든 단계를 실행할 자유도 가져야 한다. 이는 각 단계가 완료될 때 그 모든 내부 데이터를 파일이나 데이터베이스에 저장하여, 다음 단계가 이전 단계의 데이터를 메모리로 다시 적재함으로써 마치 이전 단계가 여전히 메모리에 있는 것처럼 시작될 수 있어야 함을 뜻한다. 단계는 다음과 같다.

```
w = ("stage1", -설정
      "stage2", -MARS 처리
      "stage3", -최종 거래사례비교표
      "stage4", -보고서 조각
      "stage5") -검토/정리
```

```
stage2 : w2 = "stage2"
```

6.1.4 Excel 워크북 시트 (“s”)

RCA 핵심 Excel 워크북에는 각각 내부 데이터 프레임으로 대응되는 다음 시트들이 있다.

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

2 차 집계를 별도의 시트에 두는 이유는, 가능한 모든 상호작용 쌍을 보여주는 대칭 표로 정의하는 것이 가장 손쉬워서 각 쌍에 결부된 가치기여와 보정치를 어느 집계 변수로 모을 (더할) 것인지 지정할 수 있기 때문이다. 3 차 집계는 이렇게 할 수 없는데, 3 차원 대칭 표가 필요하지만 Excel 에서는 불가능하기 때문이다. 게다가 20x20x20 표라면 다루어야 할 셀이 8,000 개가 된다. 3 차 항의 이름은 그냥 1 차 집계 표에 함께 나열하는 편이 더 쉽다. 예를 들면 다음과 같다.

변수	집계 대상
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 부동산 리스팅 (“p”) — MLSData 행

부동산 리스팅은 MLSData 시트의 행을 이룬다. 다른 시트에는 나타나지 않다가, 최종 비교사례 선정으로 보고서 조각을 만드는 4 단계에 이르러 등장한다. 부동산 ID 목록의 예는 다음과 같다.

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

6.1.6 MLS 변수 (“v”) — MLSData 열

변수의 처리에 관해서는 뒤에서 더 다룬다. 여기서는 MLSData 시트에 통상 다양한 유형의 변수 30 40 개가 있으나 MARS 회귀분석의 독립변수로는 대체로 10 20 개만 선정된다는 점만 밝혀 두면 충분하다. 이 변수들은 다른 시트에도 나타나는데, 통상 MARS 에 제출할 수 있는 모든 가능한 변수의 목록 형태이다. 변수 목록의 예는 다음과 같다.

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

6.1.7 최종 표기 예시

위 표기법을 하나의 최종 예시로 모아 보자. 위 목록들을 사용하여 다음과 같이 두자.

$$\begin{matrix} r_1 & & a_{141} \\ w_1 & W & p_2 \\ & s_0 & v_0 \end{matrix} \quad (3)$$

여기서는 0 기반 색인을 쓰고 있으므로 목록의 첫 원소는 색인 0 을 갖는 것으로 본다는 점에 유의하라. Python 도 0 기반 색인을 쓰지만 R 은 1 기반 색인을 쓴다.

따라서 위 표기는 다음에 해당하는 셀 값을 가리킨다.

- 변수 v_0 , 즉 열 “GLA”

- 부동산 p_2 , 즉 ID “MLS823542”
- 시트 s_0 , 즉 “MLSData”
- 작업 흐름 단계 w_1 , 즉 “stage2”
- 버전 r_1 , 즉
“ver20241209_151130”
- 프로젝트 a_{141} , 즉 “142”

위 표기는 다음과 같이 쓸 수도 있다.

$$\begin{array}{ccc} \text{“ver20241209_151130”} & \begin{array}{c} \text{“142”} \\ W \\ \text{“MLSData”} \end{array} & \begin{array}{c} \text{“MLS823542”} \\ \text{“GLA”} \end{array} \\ \text{“stage2”} & & \end{array} \quad (4)$$

6.1.8 정당화

“표기법에 그냥 색인이나 숫자를 쓰면 안 되는가?” 우리의 표기법에는 두 가지 요건이 있다.

1. 부분집합. 예컨대 분석자는 변수의 부분집합처럼 목록의 부분집합을 자주 선택한다. MLSData 시트에는 변수가 35 40 개 있을 수 있지만 특정 실행에는 그중 10 20 개만 사용된다. VE 는 실행할 때마다 MARS/earth 에 제출하는 변수를 자주 바꾼다.
2. 반복. 예컨대 $i = 1, 2, \dots, n_1$
우리의 알고리즘, 프로토콜, 증명은 대개 순서 목록의 모든 항목을 처음부터 끝까지 순회하거나 특정 항목을 골라낼 수 있어야 한다.

따라서 항목의 부분집합이나 목록을 나타내는 문자가 필요하고, 그 문자에는 목록의 항목들을 순회하여 데이터 프레임의 행이나 열에 색인을 거는 데 쓰이는 문자열을 얻어낼 또 하나의 아래 첨자가 있어야 한다.

6.1.9 행렬은?

선형대수의 행렬은 어떤가? 행렬은 정수든 실수든 숫자만 다룬다. 자료구조로서의 행렬은 R 과 Python 모두에서 지원된다. 그러나 MLS 데이터에서 우리가 다루는 변수 중 상당수는 “문자열” 또는 문자 데이터라 불린다. 어떤 변수가 숫자로만 되어 있더라도 그것은 실수가 아니라 그저 이름인 경우가 많다. 예컨대 어느 도시의 MLS 지역 코드는 (“660”, “661”, “662”, “663”, ..., “672”) 일 수 있다. earth 회귀분석은 변수를 “factor” 변수로 지정할 수 있게 하며, 이 경우 숫자는 순서 없이 단순한 이름으로 취급된다. 그러한 경우 산출되는 모형은 순서를 따라 흐르는 함수가 아니라 각 숫자에 대한 가치기여만을 제공한다. 다른 예로는 상태에 대한 “C1”, ..., “C6” 과 품질에 대한 “Q1”, ..., “Q6” 이 있다.

6.1.10 첫 번째 단순화

핵심 Excel 프로젝트 워크북에 대한 위 표기법은 여러 요소를 한데 묶기에는 좋지만, 알고리즘이나 수학적 증명을 설명하며 늘 지고 다니기에는 너무 무겁다. 지루한 작업을 할 때는 주어진 프로젝트, 버전, 단계, 시트 안에 있다고 가정할 수 있다. 이미 함축되어 있지 않다면 무엇을 다루고 있는지만 밝히면 된다. 무엇을 다루는지 일단 지시하고 나면 위에서 실제로 필요한 것은 다음뿐이다.

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

여기서: n_p = 부동산 수
 n_v = 입력 변수 수
 n_s = 입력 시트 수

6.1.11 두 번째 단순화

색인을 쓰기보다 시트를 그것이 뒷받침하는 데이터 프레임의 종류에 따라 이름 붙이는 편이 더 편리하다. 시트마다 행과 열의 기능이 다르다. 그리하여 W 라는 데이터 프레임 집합을 다음과 같이 M, V, A, I, C 데이터 프레임으로 나눈다.

◇ MLS 데이터

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

여기서: n_p = 부동산 수
 n_v = 입력 변수 수

참고: 대상 부동산은 기본적으로 언제나 부동산 목록의 첫머리에 온다

“MLSDData” Excel 스프레드시트와 그에 대응하는 데이터 프레임에서 첫 행은 언제나 대상 부동산을 위해 예약된다. 그 아래의 다른 MLS 부동산들과 함께 정렬되는 일이 결코 없다. MARS 회귀분석에 독립 데이터로 투입되지 않으며, 목표 데이터 즉 종속변수로 투입된다. 부동산 목록에서도 마찬가지다. 모든 부동산 목록에서 색인 0 의 첫 원소는 달리 밝히지 않는 한 언제나 대상 부동산 ID 로 본다.

대상 부동산 없이 시장 지역의 모형을 만드는 과업이더라도 첫 행에는 어떤 가상의 부동산 데이터를 적재한다. 이렇게 하면 프로그램을 작성하거나 알고리즘을 규정하거나 증명을 하기가 훨씬 편해진다. 따라서 c 로 부동산 목록을 나타낸다면, c_0 는 대상 부동산이고 c_i , $i = 1, \dots, n_p$ 는 MARS/earth 회귀분석에 입력될 비교사례 리스팅이다.

◇ 회귀 변수 지정

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

여기서: n_v = 사용 가능한 변수 수
 n_s = 지정의 수

통상 회귀분석 입력으로 사용 가능한 변수의 부분집합을 선택하므로 $n_s \leq n_v$ 이다.

◇ 계산

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

여기서: n_v = 변수 수
 n_c = 계산 지정 수

◇ 상호작용

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

여기서: n_f = from 변수 수
 n_t = to 변수 수

데이터 프레임 I 는 대칭 데이터 프레임으로, “from” 행과 “to” 열에 변수가 알파벳순으로 나열된다. 셀에 “1” 을 표시하면 해당 변수 쌍 사이의 상호작용이 허용됨을 뜻한다. 그렇지 않고 “0” 이면 상호작용이 허용되지 않음을 뜻한다. 2 차 상호작용에서는 가능한 상호작용이 하나뿐이므로 비교적 단순하다. 3 차 회귀에서는 세 변수 사이에 세 개의 쌍 상호작용이 있으며, 세 변수 상호작용 항 (또는 변수) 을 생성하려면 그 모든 쌍의 상호작용이 허용되어야 한다.

◇ 집계

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

여기서: n_v = 변수 수

7 RCA 거래사례비교표

RCA 프로토콜에는 R/earth 에 접근하여 모든 세부 단계를 수행할 수 있는 R 또는 Python 프로그램이 필요하다는 점을 유념하라. 데이터 집약적이어서 그 작업을 수작업으로 수행하기란 도저히 불가능하다. 동시에 프로그램에 버그가 없다는 전제 아래, 필요한 작업이 사람의 실수 없이 정확하게 수행되도록 보장하려면 소프트웨어 프로그램이 있어야 한다.

또한 이 일을 하는 VE 는 특히 예상하지 못한 새로운 시장 지역, 근린지역, 부동산이나 그 밖의 새로운 요구사항, 문제에 부딪혔을 때 그러한 프로그램을 필요에 따라 수정할 수 있어야 한다.

RCA 처리에는 세 가지 유형의 거래사례비교표가 있다.

1. RCA 처리용 거래사례비교표. 이 논문에서 “MLSDData” 로 표기하는 것으로, 매우 큰 데이터 프레임과 그에 대응하는 스프레드시트이며, 집계 과정을 거쳐 아래에서 말하는 RCA 거래사례비교표에 이르게 될 여러 중간 세부 항목을 담고 있을 수 있다. 예시를 여기에 실기에는 너무 크다. 감정평가공학자들은 대개 이 스프레드시트를 자신의 취향에 맞게 다듬을 것이다. 그러한 표의 개발에 관해서는 앞으로 별도의 논문을 쓸 생각이다.
2. RCA 최종 거래사례비교표. 비 GSE 보고서에 쓰이며 표 2와 비슷한 형태다. 이는 독자에게 어느 정도 익숙함을 주기 위해 예컨대 Fannie Mae 양식을 어느 정도 본뜬 것이다. 다만 비교사례 하나당 열의 수가 Fannie Mae 표보다 많다는 점에 유의하라. 각 비교사례에는 “Measure”(측정값) 와 “Adjustment”(보정) 열 옆에 “Value Contribution”(가치기여) 열이 있고 대상 부동산에도 하나가 있는데, 이는 Fannie Mae 양식에 없는 열이다. 더 단순한 형태는 표 1에 있다.
3. GSE 업로드용 거래사례비교표. Alamode 같은 소프트웨어에 거래사례비교표를 올릴 때 쓴다. 다만 모든 감정평가 양식 소프트웨어가 스프레드시트에서 거래사례비교표 데이터를 올리는 기능을 제공하는 것은 아니다. 부득이하게 그러한 표에는 가치기여를 위한 별도의 열을 둘 수 없는데, Fannie Mae 같은 GSE 는 가치기여라는 것을 전혀 이해하지 못하기 때문이다. 전략은 그들이 요구하는 양식을 주되, RCA 거래사례비교표와 변수 집계의 세부와 같은 그 밖의 뒷받침 문서로 이를 보강하는 것이다.

RCA 처리용 거래사례비교표는 Fannie Mae 거래사례비교표처럼 생기지 않았고, 대략 그것의 전치 (transpose) 에 해당하여 행과 열이 뒤바뀌어 있다. 각 행은 어느 부동산의 데이터를, 각 열은 어느 부동산 변수의 데이터를 담는다. 부동산 행은 수백 내지 수천 개, 열은 200 개 이상이 되기 쉽다. 첫 행은 변수명에, 둘째 행은 대상 부동산에 예약되는데, 시장분석만 하는 경우에도 그러하다. 스프레드시트를 데이터 프레임으로 읽어들이면 시트의 첫 행은 names 속성이 되고, 대상 부동산이 있는 둘째 행이 데이터 프레임의 첫 행이 된다.

RCA 거래사례비교표의 열에는 식별용 열이 있고, 그다음에 1 차 변수, 외부 변수, 합성 변수가 결합되어 독립변수로 MARS 에 투입된다. MARS 는 이 변수들 가운데 거래가격 (종속변수 또는 목표변수라고도 한다) 에 유의미한 영향을 준다고 판단되는 것들을 출력한다. 2 차 회귀에서는 새로운 변수 역할을 하는 변수 쌍이 나올 수 있다. 3 차 회귀에서는 이른바 독립변수 역할을 하는 변수 삼중항도 얻을 수 있는데, 물론 이는 밑에 깔린 변수들과 높은 상관을 갖는다.

8 변수

MARS 를 실행하는 입력 측에는 다음과 같은 단순 변수만 있다.

1. MLS 스프레드시트에서 적재된 부동산 특성.
2. MLS 변수를 다듬거나 기존 변수의 변환을 추가하기 위해 VE 가 만든 파생 변수. 예컨대 VE 는 거래일 (Date of Sale) 을 감정평가 기준일에서 거래일을 뺀 값인 “DaysOffMarket” 으로 변환하기를 선호할 수 있다.
3. VE 는 주택담보대출 금리와 같은 제 3 자 자료원의 데이터를 들여올 수 있다. 또는 MLS 데이터에 위도, 경도, 표고 데이터가 없다면 다른 자료원에서 가져올 수 있다. 이러한 변수는 “외부 변수” 라 부를 수 있다.

MARS 의 출력 측에서 우리가 다루는 것은 다음 중 하나에 해당하는 항들의 조합으로 이루어진 가격 모형이다.

1. 입력 변수 하나만을 포함하며 그 변수의 가치기여에 대한 식을 제공하는 항. 이를 1 차 항 또는 (출력) 변수라 부른다. 항 자체가 가치기여 값으로 취급되며 이름이 부여된다.
2. 입력 변수 두 개를 포함하며 그 두 변수의 상호작용에 대한 가치기여의 식인 항. 이를 2 차 항 또는 (출력) 변수라 부를 수 있다.
3. 입력 변수 세 개를 포함하며 그 세 변수 사이 상호작용에 대한 가치기여의 식인 항. 이 논문에서는 이를 3 차 항 또는 (출력) 변수라 부를 수 있다.

그리하여 MARS 의 출력 변수는 잔차 가치기여를 포함한 가치기여들이다. 이들은 다시 보정치 계산이나 그래프 작성과 같은 거래사례비교표 계산에서 변수로 취급된다.

변수에는 두 가지 기본 온톨로지로 나뉘는 여러 유형이 있다. (1) 변수의 기원, (2) 변수의 용도, 즉 비잔차 대 잔차.

8.1 변수 기원 온톨로지

1. 1 차 변수
 - 1.1 MLS 회귀 변수
 - 1.2 외부 회귀 변수
 - 1.3 파생 회귀 변수
2. 팩터 변수 (factor variable):

팩터 변수를 다루려면 그것을 식별할 수 있는 함수가 필요하다. 팩터 변수는 Excel 프로젝트 설정 워크북에 지정되므로, 전달받은 변수가 팩터 유형으로 분류되어 있으면 참을 반환하는 함수를 만든다.
3. 상호작용 변수: 차수가 2 이상으로 설정된 경우 MARS 가 모형을 생성하면서 만들어 낸다. 다만 해석의 어려움, 과적합의 가능성, 자유도에 미치는 영향과 그로 인해 견고한 모형을 위해 필요한 부동산 레코드의 증가를 고려하면 권장할 수 있는 상호작용 변수는 최대 3 개다.
 - 3.1 2 차 변수: 흔히 나타나며, GLA:BathRms, GLA:LotSize, DateOfSale:BathRms, Latitude_Longitude 가 좋은 예다.
 - 3.2 3 차 변수: 덜 흔한 상호작용으로, 신증을 요하며 과적합의 소지가 있다.
4. 잔차 분해 변수

8.2 변수 용도 온톨로지

1. MARS 가 생성한 가치기여

1.1 1 차 변수

1.1.1. MLS 회귀 변수

1.1.2. 외부 회귀 변수

1.1.3. 파생 회귀 변수

1.2 상호작용 변수

1.2.1. 2 차 변수

1.2.2. 3 차 변수

1.3 잔차 분해 변수

1.4 집계 변수

2. 공학자가 생성한 변수 및 값

2.1 잔차 분해 변수

2.1.1. 디자인/미관

2.1.2. 조경

2.1.3. 기능적 효용

2.1.4. 조망

2.1.5. 상태

2.1.6. 품질

2.1.7. ...

이 변수 집단들에 대한 표기법과 상세한 설명은 다음 절에서 다룬다.

프로토콜 3.1: 팩터 변수

팩터 변수, 줄여서 팩터는 상호작용을 포함한 모든 항에 나타날 수 있다. MARS 회귀분석과 그래프 작성에서 특별한 처리를 요하므로 중요하다.

문자값을 갖는 변수는 필연적으로 팩터이지만, 수치 변수는 MARS 가 팩터로 인식하도록 팩터 변수로 표시해 주어야 한다. 예컨대 MLS 지역은 흔히 번호로 식별된다. 특히 샌프란시스코 베이 에어리어의 근린 구역은 흔히 고유한 정수로 식별된다. 이 값들을 팩터로 MARS 에 넘기면, MARS 는 다른 구역과의 관계와 무관하게 특정 구역이 그 경계 안의 부동산 거래에 평균적으로 기여하는 값이 있는지를 찾아내려 한다. 이와 달리 일반 변수의 경우 MARS 는 변수값이 변함에 따라 가치기여에도 어떤 연속성이 있다고 가정한다. 이는 중요한 구별이다. 더욱이 이러한 팩터 변수는 모형의 항에서 팩터가 아닌 변수와 함께 나타날 수 있어 그래프 작성을 어렵게 한다. 구체적으로, 팩터가 포함된 단일 변수 항을 그래프로 그릴 때는 그러한 팩터를 모두 한데 모아 그 기여를 양이든 음이든 막대그래프로 단순히 그린다.

팩터가 항에서 다른 변수와 함께 나타나는 경우, 팩터 변수는 특정 부동산에 대해 값을 하나만 가질 수 있으므로 언제나 하나의 가능한 값으로 나타난다. 어떤 구역의 값이 662 라면 정의상 그 구역이 다른 값을 가질 수 없다. 이것이 혼란스럽다면 사정을 더 복잡하게 만드는 점도 있다. MARS 가 생성한 모형에는 팩터에 대한 단일 변수명이 없다. MARS 는 입력으로 받은 팩터 변수의 어근 이름을 취한 다음, 가능한 값들을 덧붙여 만든 변수명으로 새 열들을 만든다. 그래서 MlsArea 같은 Excel 스프레드시트의 단일 열에 대해 MlsArea660, MlsArea661 등 가능한 값마다 데이터 프레임 열이 여러 개 생성된다. MlsArea661 같은 변수는 661 구역의 부동산에 대해 값 1 을, 그렇지 않으면 0 을 갖는다. 그런 다음 MARS 는 그 열들을 불리언 변수처럼 취급하여 회귀분석한다. 이렇게 변형된 이름이 MARS 가 제공하는 모형에 사용된다.

깔끔하고 단순한 것을 좋아하는 의뢰인에게 그래프를 (가능한 한) 더 이해하기 쉽게 만들려면 그 모형 이름들을 원래 이름으로 되돌리고 싶을 것이다. MlsArea661=1 대신 MlsArea=662 와 같이 보게 되는 것이다. 그러면 그래프 머리글에 “MLS Area = 662 에 대하여” 와 같은 문구를 적어 항을 그래프로 그릴 수 있다. 이렇게 하여 팩터 변수 M 개를 가진 N 차 항은 (N-M) 차 항으로 그려지고, 팩터 차원은 제목 속으로 사라진다.

한 항에 팩터 변수가 둘 이상 있을 수 있으므로, 값의 조합마다 그래프가 하나씩 필요할 수 있다는 점에 유의하라. 서로 다른 값 3 개를 갖는 변수 x 와 값 2 개를 갖는 변수 y 가 팩터가 아닌 변수 z 와 같은 항에 나타난다면 $2 \cdot 3 = 6$ 개의 그래프가 필요하다. 예를 들면 “MlsArea” + “Architectural Style” 이 있는데, 샌프란시스코의 빅토리아식 주택은 어느 구역에 있느냐에 따라 가치가 더하거나 덜할 수 있다. 다행히 대부분의 지역에서는 (건축양식이 6 가지 이상 있을 수 있더라도) 모형에 쓸 만큼 가치기여자로서 중요한 건축양식은 통상 1 3 가지뿐이다.

8.3 2 차 및 3 차 항

2 차 또는 3 차 회귀에서는 출력 모형의 일부 항에 변수가 각각 둘 또는 셋 포함될 수 있다. 예를 들면 다음과 같다.

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

이러한 항의 이름은 변수명을 알파벳순으로 정렬하여, 대개 밑줄을 넣어 이어 붙여 짓는다는 점에 유의하라. 3 차 항에 대해서도 모형이 산출한 변수명 삼중항에 같은 방식을 적용한다. 예컨대 “BathRms_GLA_LotSz” 이다.

이 변수들은 RCA 프로그램이 상호작용 변수를 포함한 모든 변수의 부동산별 기여를 계산하는 3 단계에서 1 차 입력 변수에 추가된다.

8.3.1 보정 변수

RCA 프로그램은 가치기여가 계산된 뒤 3 단계에서, 어떤 특성에 대한 대상 부동산의 기여에서 그에 대응하는 비교사례의 가치기여를 빼는 방식으로 모든 변수와 부동산에 대한 보정치도 계산한다.

8.3.2 집계 변수

집계 변수는 변수들의 부분집합을 대체하기 위해 정의되며, 그 가치기여와 보정치를 모두 더하여 보고서에 쓸 공통 변수로 모은다. 예컨대 마감된 주거면적과 마감되지 않은 주거면적을 다음과 같이 더 세분한 변수로 나누어 두었을 수 있다.

1. 지상 합법 마감면적
2. 지하 합법 마감면적
3. 지상 불법 마감면적
4. 지하 불법 마감면적
5. 지상 미마감면적
6. 지하 미마감면적

실제로 이런 구분이 의미 있는 지역이 있다.

MARS 회귀분석은 이 모든 변수에 대해 회귀를 수행하여 그 기여와 보정치를 제공하는 모형을 생성한다. 그러나 양식 보고서를 위해서는 이렇게 서로 다른 기여들을 더 적은 수의 변수로 결합하거나 집계해야 할 것이다. 집계의 세부는 부록에 제시해야 한다. 더 자주 있는 예는 상호작용 보정치를 1 차 변수 보정치 중 하나로 집계하는 것이다. 예컨대 GLA_Lotsize 에 대한 보정치를 GLA 로 또는 LotSize 로 집계하거나, 집계 시트에서 정한 대로 둘 사이에 50:50 으로 나눌 수도 있다.

8.3.3 특수 변수 및 표기법

π_i	부동산 $i = 1, 2, \dots, n_p$ 의 실제 순 거래가격. 대상 부동산의 경우 π_0 는 가치 결론을 나타낸다 (증명 I 참조)
α_i	부동산 $i = 1, 2, \dots, n_p$ 의 보정 후 순 거래가격
ρ_i	부동산 $i = 0, 1, \dots, n_p$ 에 대한 MARS 추정 순 거래가격
β	모형의 기초가치
$\varphi_{i,j}$	부동산 $i = 0, 1, \dots, n_p$ 에 대한 회귀 변수 $j = 1, 2, \dots, n_r$ 의 가치기여
$\delta_{i,j}$	부동산 $i = 1, 2, \dots, n_p$ 에 대한 회귀 변수 $j = 1, 2, \dots, n_r$ 의 보정치
ϵ_i	부동산 $i = 0, 1, \dots, n_p$ 의 총 잔차
$\epsilon_{i,c}$	부동산 $i = 0, 1, \dots, n_p$ 에 대한 잔차 구성요소 c , $c = 1, 2, \dots, n_c$ 의 가치기여
$\xi_{i,c}$	부동산 $i = 0, 1, \dots, n_p$ 에 대한 잔차 구성요소 c , $c = 1, 2, \dots, n_c$ 의 보정치

여기서

n_p = 비교사례 부동산 수

n_r = 회귀 변수 수

n_c = 잔차 분해 변수 수

8.3.4 잔차의 분해

MARS 회귀분석이 모형을 만들고 나면, MARS 가 선택한 입력 변수들로부터 추정 거래가격을 생성한다. 그리고 그 값을 대응하는 순 거래가격에서 빼서 각 부동산의 잔차값을 얻는다. 이제 RCA 프로토콜은 VE 가 최종 비교사례 부동산 — 통상 6 12 개의 부동산 리스팅 — 을 정하고 나면, 각 부동산 리스팅을 하나씩 살펴 계산된 잔차를 가치에 영향을 준다고 판단되는 가장 유력한 특성 (잠재변수) 들로 분해할 것을 요구한다. 이때 무작위적 데이터 오류의 가능성은 고려하지 않는다. 예컨대 어떤 부동산에 대해 다음과 같은 분해 변수를 둘 수 있다.

1. 디자인
2. 기능적 효용
3. 시장가 초과/미달 거래
4. 상태

5. 품질

부동산마다 잔차 분해 변수가 얼마간 다를 것이므로, 이들을 하나의 순서집합으로 모아야 한다.

희소 특성의 기여 추정. 가치를 지닌 특성 가운데 회귀 경로에 태우기에는 지나치게 드문 것들이 있다. 780 건의 거래 중 3 건에만 있는 작업장은 방어 가능한 계수를 얻을 수 없고, 전진 단계는 그러한 변수를 아예 건너뛰는 것이 보통이다. 이들은 잔차 특성이며, 그 기여는 잔차 공간에서 추정할 수 있다. 즉 해당 특성 없이 모형을 적합시킨 다음, 모형의 잔차를 제외한 열 (들) 에 대해 통상최소제곱으로 회귀시키는 것이다. 이진 플래그의 경우 계수는 일시금 형태의 기여 (달리 말해 플래그가 붙은 거래들의 평균 잔차 편차) 이고, 수치 열의 경우에는 단위당 기여다. 상관된 여러 부대시설을 동시에 추정하는 경우 (예컨대 부속 건물 면적을 작업장 플래그와 함께), 회귀분석은 같은 금액을 두 개의 분해 행에 이중으로 계상하지 않고 그 몫을 둘 사이에 나눈다. 관측치가 약 열 개에 미치지 못하는 추정치는 참고용일 뿐이므로 감가상각 원가나 넓은 표본 (예컨대 public remarks 에서 복원한 플래그) 으로 뒷받침해야 한다. 잔차 제약 자체가 최종적으로 배정되는 값의 상한을 지운다. earthUI(v0.11 이상) 는 이 방법을 *Residual Contribution*(잔차 기여) 탭으로 직접 구현하고 있으며, 동시 추정, 거래별 잔차 증거, 작업 파일용 문서화 노트를 함께 제공한다.

표기: 위 표의 표기법을 사용하면 다음을 얻는다.

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

여기서 n_c 는 잔차 분해 구성요소의 개수다.

VE 는 자신의 판단으로 잔차를 주어진 구성요소들로 분해해야 한다. 이 때문에 최종 가치 결론에 편이가 끼어들 수 있다고 생각할 수도 있다. 그러나 그렇지 않다. 분해된 보정치의 합계가 잔차와 같다면 가치 결론에 영향을 주지 않는다는 수학적 증명이 있기 때문이다. 분해의 목적은 가치를 확정하는 데 있지 않고, 다른 부동산 거래의 잔차와 견주어 그 잔차가 왜 그러한 값인지를 설명하는 데 있다. 공학자의 편이가 최종 가치 결론에 영향을 줄 수 있는 유일한 지점은 대상 부동산의 잔차 추정이다. 물론 대상 부동산에는 잔차를 계산하는 데 필요한 실제 거래가격이 없기 때문이다.

부동산마다 잔차 가치기여의 집합이 있으나, 이를 모든 부동산에 걸친 하나의 집합 C 로 병합해야 한다.

다음과 같이 두면: $A_i = \{\{\epsilon_{i,c}\}_{c=1}^{n_c}\}$

$$C = \{x \mid x \in \bigcup_{i=0}^{n_p} A_i\}$$

여기서 $[\epsilon_{i,c}]$ 는 값 $\epsilon_{i,c}$ 에 대한 변수명이고, A_i 는 그러한 변수들의 집합이며, n_c 는 각 ϵ_i 에 들어 있는 값의 개수다.

따라서 대상 부동산과 비교사례의 데이터를 별개의 열에 두면, 각 부동산에 대한 모든 회귀 모형 변수의 가치기여 아래에 그에 대응하는 n_c 개의 잔차 가치기여 집합 C 가 보여야 한다.

9 RCA 의 핵심 증명

다음과 같이 두자.

n_p = 부동산 수

n_r = 회귀 변수 수

n_c = 잔차 구성요소 변수 수

증명 I: 모든 비교사례의 보정 후 거래가격은 대상 부동산의 추정 거래가격과 같다.

임의의 비교사례 i 에 대하여, 보정 후 거래가격 α_i 는 순 거래가격 π_i 에 보정치들을 더한 값이다. 보정치란 기초가치, 회귀 변수 $j = 1, 2, \dots, n_r$, 그리고 잔차 각각에 대한 대상 부동산과 비교사례의 가치기여 차액을 말한다. 실제 거래가 없는 대상 부동산의 경우 순 거래가격 π_0 는 가치 결론 $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$ 로 정의한다. 그러면 다음과 같다.

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

증명 II: 모든 잔차 구성요소의 가치기여 합계가 비교사례의 잔차와 같다는 제약 아래에서 잔차 분해를 바꾸더라도 RCA 프로토콜의 최종 가치 결론에는 아무런 영향이 없다.

이는 거의 자명하다. 그럼에도 수학을 생각해 보지 않으면 놓치기 쉬운데, 전통적 감정평가사는 가치기여가 아니라 보정치에 주목하기 때문이다. 여기서 상쇄되는 것은 바로 가치기여이며, 이는 수십 년간 간과되어 온 사실이다. 어떤 비교사례 $k(k > 0)$ 에 대하여, 총 잔차 보정치는 n_c 개의 분해된 잔차 구성요소의 합과 같고, 이는 다시 가치기여 차액들의 합과 같으며, 이는 곧 대상 부동산의 총 잔차에서 비교사례의 총 잔차를 뺀 값과 같다. 그리고 이는 만들어진 잔차 구성요소의 개수나 그것이 무엇을 나타내는지와 무관하다. 대상 부동산과 선택된 비교사례 모두에서 잔차 구성요소들의 값이 총 잔차와 같기만 하다면 ξ_k 는 달라지지 않는다.

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

10 MARS 모형

10.1 모형 변수 및 상호작용 “ g ” 함수

RCA 방법의 핵심 원리 하나는 “설명적 추론”이 무엇보다 중요하다는 것이다. 안타깝게도 특히 상호작용과 팩터 변수가 쓰이는 경우, 여러 가지기여 식에 대해 의미 있는 그래프를 제공하기란 쉽지 않을 수 있다. 그래서 이러한 그래프를 만드는 명확하고 엄격한 프로토콜을 제시하는 것이 긴요하다고 본다. 이를 여기서 제시하되, 이 프로토콜은 그래프를 생성하는 R 또는 Python 코드를 작성하는 개발자를 겨냥한 것이며 코딩에 관여하지 않는 이들에게는 표면적인 관심사에 그칠 수 있음을 이해하기 바란다.

MARS가 생성한 가격 모형은 기초 상수에 각각 변수 1 3 개를 담은 여러 항을 더한 것이다. 이를 분석하기 위해 먼저 모형을 개별 항으로 나눈다. 그런 다음 동일한 변수 집합을 공유하며 하나의 상호작용을 나타내는 여러 항을 하나의 상호작용 식으로 결합한다. 회귀의 최대 차수가 3 이라고 하자. 기초 상수, n_1 개의 1 차 식, n_2 개의 2 차 상호작용 식, n_3 개의 3 차 상호작용 식으로 이루어진 3 차 회귀 모형을 생각해 보자. 이렇게 하면 단일 변수이거나 다변수 상호작용인 네 개의 구별되는 식 집단이 생기며, 각각은 두 가지 요인에 따라 다르게 그려진다. (1) 팩터 변수 (범주형 변수)를 포함하는지 여부, 그리고 (2) 그 안에 들어 있는 서로 다른 변수의 개수. 구체적으로는 다음과 같다.

- 1 차 식 (변수 1 개) 은 2D 로 그린다.
- 2 차 식 (변수 2 개) 은 3D 로 그린다.
- 3 차 식 (변수 3 개) 은 3D 그래프의 집합으로 그린다.
- 식에 팩터 변수가 포함되면 서로 다른 팩터 하나마다 그래프의 차원이 하나씩 줄어든다 (예컨대 한 변수가 팩터이면 3D 그래프가 2D 가 된다).

각 함수 g 의 색인은 그것을 어떻게 그려야 하는지를 판정하는 데 필요한 정보를 제공하며, 서로 다른 변수의 개수와 팩터 변수의 존재 여부를 반영한다.

10.2 g 함수의 색인

각 함수 g 는 $f_{j,k}g_k^j$ 로 색인되며, 여기서:

- $f_{j,k}$ 는 첫 번째 색인이다 (g 앞의 위 첨자).
 - j 는 0에서 3까지의 값을 가지며 집단을 식별한다.
 - * $j = 0$: 기초 상수 (통상적 색인 규칙의 예외).
 - * $j = 1$: 1 차 식.
 - * $j = 2$: 2 차 식.
 - * $j = 3$: 3 차 식.
 - k 는 집단 안에서 특정 함수의 색인이다 (예컨대 집단 j 의 첫 함수는 $k = 1$).
 - $f_{j,k}$ 의 값은 그 함수 안에 있는 서로 다른 팩터 변수의 개수를 나타낸다.
 - * $f_{j,k} = 0$: 팩터 변수 없음.
 - * $f_{j,k} > 0$: 팩터 변수의 개수 (예컨대 1, 2, 3).
- g 에 뒤따르는 색인:
 - g 뒤의 위 첨자 j (예: g^j)는 명료성을 위해 집단 번호를 반복한다.

- g 뒤의 아래 첨자 k (예: g_k) 는 $f_{j,k}$ 의 k 와 일치하여 집단 안에서 함수를 식별한다.

f 와 g 의 색인은 수치상 일치하지만 그 역할은 다르다. $f_{j,k}$ 는 그래프 작성 판단에 결정적인 팩터 변수의 개수를 세는 반면, g 의 j 와 k 는 그 집단과 위치를 규정한다.

10.3 그래프의 차원

함수 $f_{j,k}g_k^j$ 에 대한 그래프의 차원 d 는 다음과 같이 계산한다.

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

여기서:

- j 는 서로 다른 변수의 개수 (상수의 경우를 제외하면 집단 번호) 다.
- $f_{j,k}$ 는 팩터 변수의 개수다.
- d 는 그래프로 그릴 연속 (비팩터) 변수의 개수이며 최대 3 이다 (최대 차수가 3 이므로). 연속 변수가 d 개인 함수는 $d + 1$ 차원으로 그린다. $d = 1$ 이면 2D 그래프, $d = 2$ 면 3D 그래프 (또는 2D 열지도), $d = 3$ 이면 3D 그래프의 집합이다.

예를 들면:

- $j = 3$ (변수 3 개) 이고 $f_{j,k} = 0$ (팩터 변수 없음) 이면 $d = 3 - 0 = 3$ 이므로, 변수 하나씩을 고정한 3D 그래프의 집합이 필요하다.
- $j = 3$ 이고 $f_{j,k} = 1$ (팩터 변수 1 개) 이면 $d = 3 - 1 = 2$ 이므로, 모형에 나타나는 팩터 변수의 값마다 3D 그래프 (또는 2D 열지도) 가 하나씩 필요하다.
- $j = 0$ (상수) 이면 값이 하나뿐이므로 통상 그래프가 필요하지 않다.

$d = 3$ 인 경우 (예컨대 팩터 변수가 없는 3 차 식) 함수 전체를 시각화하려면 여러 개의 3D 그래프가 필요할 수 있으며, 각각은 변수 하나를 서로 다른 값에 고정하여 함수 거동의 한 부분을 보여준다.

10.4 최종 모형 식

MARS 모형은 네 개의 식 집단 (상수, 1 차, 2 차, 3 차) 으로 나뉜다. 색인 $f_{j,k}g_k^j$ 는 다음을 알려준다.

- j : 집단과 변수의 개수.
- $f_{j,k}$: 팩터 변수의 개수.
- $d = j - f_{j,k}$: 그래프의 차원.

이 구조는 각 g 함수를 어떻게 시각화할지 이끌어 주며, 그래프 작성 방법이 식의 복잡성과 변수 유형에 부합하도록 보장한다.

MARS 모형이 제공하는 거래가격 추정치 $\hat{P}$ 는 다음과 같이 표현할 수 있다.

$$\hat{P} = f_{0,1}g_1^0 + \sum_{q=1}^{n_1} f_{1,q}g_q^1(x_*)$$

$$+ \sum_{r=1}^{n_2} f_{2,r}g_r^2(x_*, y_*)$$

$$+ \sum_{s=1}^{n_3} f_{3,s}g_s^3(x_*, y_*, z_*) \quad (15)$$

$n_1 = 1$ 차 항의 수
 $n_2 = 2$ 차 항의 수
 $n_3 = 3$ 차 항의 수

10.5 모형을 그래프로 나타내는 데 필요한 차트의 수

모형 안의 모든 함수를 시각화하려면 그래프가 몇 개나 필요한가? 답은 변수, 그 상호작용, 그리고 그것이 팩터 변수인지 여부에 달려 있다.

VE 는 R/earth 를 위한 Excel 설정 파일에서 변수와 상호작용을 제한할 수 있으며, 이는 어떤 상호작용이 허용되는지에 영향을 준다. 실용적인 접근은 R/earth 로 하여금 가능한 모든 상호작용을 평가하게 하여 거래가격 추정에 유의미한 기여를 하는 것들을 고르게 하되, 과적합이나 공선성 같은 문제를 함께 점검하는 것이다. 일부 상호작용은 배제될 수 있고, 이는 더 높은 차수의 상호작용에도 영향을 준다 (예컨대 3 차 상호작용에서 한 쌍이 제외되면 그 조합 전체가 제외된다).

팩터 변수의 경우 (예컨대 값이 12 개인 MLS Area 가 GLA 와 상호작용하는 경우) 그래프 작성성이 더 복잡해진다. 구역마다 하나씩 12 개의 그래프가 필요할 수 있다. 여기에 값이 6 개인 Architectural Style 같은 또 다른 팩터가 MLS Area 및 GLA 와 상호작용한다면 최대 $12 \cdot 6 = 60$ 개의 그래프가 필요할 수 있다. 그러나 R/earth 가 그 모든 조합을 유의미하다고 보는 일은 드물다. 통상은 모형이 짚어 주는 몇 개의 핵심 상호작용 (예컨대 특정 구역과 양식) 만이 그래프로 그럴 가치가 있다.

필요한 그래프 수에 대한 간략한 지침은 다음과 같다.

1. 단일 변수 함수 (비팩터): 그래프 1 개.
2. 단일 변수 함수 (팩터): 모든 팩터 값을 합친 막대그래프 1 개.
3. 기초 함수: 그래프 불필요.
4. 2 차 상호작용 (팩터 없음): 그래프 1 개 (예컨대 거래가격을 색으로 나타내는 2D 열지도).
5. 2 차 상호작용 (팩터 1 개): 모형에 나타나는 팩터 값마다 그래프 1 개 (예컨대 값이 2 개면 그래프 2 개).
6. 2 차 상호작용 (팩터 2 개): 팩터 조합의 행렬을 담은 그래프 1 개 (예컨대 구역 3 개 $\times$ 양식 2 개 = 셀 6 개).
7. 3 차 상호작용 (팩터 없음): 여러 개의 그래프 (예컨대 욕실 수 = 1, 2, 3, 4 에 대해 4 개).
8. 3 차 상호작용 (팩터 1 개): 모형에 나타나는 팩터 값마다 그래프 1 개.
9. 3 차 상호작용 (팩터 2 개): 모형에 나타나는 팩터 값 조합마다 그래프 1 개.
10. 3 차 상호작용 (팩터 3 개): 서로 다른 값의 수가 가장 적은 팩터의 값마다 그래프 1 개를 그리고, 나머지 두 팩터는 색 행렬로 나타낸다.

더 복잡한 상황에서는 고급 그래프 기법이 필요할 수 있다.

더 정교한 그래프를 요하는 다른 가능성들도 있다.

MARS 가 생성한 완전한 모형의 예는 [그림 1](#)에 있다.

11 데이터의 특성

이 절에서는 MARS 를 다룬다. MARS 는 전적으로 자신이 받는 데이터에 의존하며, 그 데이터의 대부분은 어떤 MLS(Multiple Listing Service) 나 세무 평가관에게서 나온다. 다만 데이터를 관리하고 개선하여 이윤을 붙여 파는 다른 데이터 공급자를 거칠 수도 있다. 어떤 지역에서는 상당히 신뢰할 만하고, 다른 지역에서는 그렇지 않다. 결정적인 사안에서는 부동산 조사자가 직접 관찰하고 측정해야 하는데, 그 정확성 또한 의문의 대상이 될 수 있다.

감정평가공학자는 데이터가 무엇을 나타내는지, 그리고 그것이 나타내야 할 바를 얼마나 잘 나타내는지 이해해야 한다. 여기서 다음과 같은 여러 쟁점이 따라 나온다.

- 어떤 변수를 측정값으로 분류하는 기준은 무엇인가?
- 변수가 측정값이 아니라면 달리 무엇일 수 있는가?
- 받은 데이터가 정확하지 않다면 회귀분석이 어떻게 쓸모 있을 수 있는가?

MLS 리스팅 서비스와 세무 평가관 데이터에서 회귀분석용으로 얻는 데이터의 대부분은 측정값이나 개수를 함의하는 수치 데이터다. 일부는 참/거짓의 논리값이다.

11.1 잔차의 목적

잔차는 회귀 모형에 포함되지 않은 변수들이 더한 가치의 간접적 척도 역할을 한다. 헤도닉(hedonic) 문헌의 용어로 하자면 이러한 측정되지 않은 특성은 잠재변수다. 즉 거래가격에 영향을 주지만 직접 측정되지 않는 변수이며, 대개는 그것을 객관적으로 측정할 실용적 방법이 없기 때문이다. 그 값은 대신 추론되어야 하는데, 실무에서 잠재변수의 값은 흔히 주관적 판단으로 추정된다. RCA 에서는 그 판단이 잔차 순위표에 맞을 내린다. 부동산을 잔차 또는 제곱피트당 잔차로 정렬함으로써 매력이 가장 낮은 부동산에서 가장 높은 부동산에 이르는 객관적 순위를 얻는다. 지역 시장을 깊이 이해하는 숙련된 분석자가 MARS(다변량 적응 회귀 스플라인) 모형을 만든다는 전제 아래 이 순위 매기기 방법은 효과적이다. 추가로 요구되는 사항은 다음과 같다.

- 데이터와 모형의 품질: 잘 구축된 회귀 모형과 양질의 데이터는 믿을 만한 잔차의 전제조건이다. 샌프란시스코 베이 에어리어에서 나의 목표는 R^2 약 75-80%, 교차검증 $R^2(CV\ R^2)$ 55-70% 의 모형을 개발하는 것이다.
- 과적합을 피하고 R^2 를 제한하라: SF 베이 에어리어처럼 가격대가 높고 성숙한 대부분의 시장에서 감정평가공학자는 R^2 상한을 약 80% 로 잡아야 한다. 주거용 부동산의 경쟁력 있는 가치 가운데 상당 부분이 계량 불가능하여 회귀 모형에 직접 넣을 수 없는 주관적 매력에 있기 때문이다. R^2 가 80% 를 크게 웃돈다면 과적합된 모형일 수 있으므로 더 면밀히 살펴야 한다. 다만 한 가지 유의할 점이 있다. 과적합 없이도 R^2 가 90% 를 넘는 시장 지역이 있고, 감정평가가 어려워 잘해야 50-60% 또는 그 이하에 머무는 지역도 있다.
- 특정 지역에서의 높은 R^2 : 다만 택지개발지구나 그 밖의 균질한 주거지역처럼 부동산의 동질성 덕분에 놀라울 만큼 흔히 높은 R^2 값이 나오는 지역도 있다.

11.2 근사 회귀변수로서의 상태 및 품질 등급

Fannie Mae 의 상태 및 품질 등급 척도 (C1-C6, Q1-Q6) 에는 특별한 언급이 필요하다. 이 척도들은 참된 측정 특성으로 취급하기에는 너무 거칠고, 감정평가사가 부여하는 것이므로 실무상 주관적 등급 — 그 자체로 잠재변수 — 이다. 두 층위를 명시적으로 구별해 두는 것이 유용하다. 하나는 회귀변수로 쓰이는 감정평가사의 거친 등급, 즉 상태-조립도 (condition-1) 와 품질-조립도 (quality-1) 이고, 다른 하나는 잔차에서 복원되는 세밀한 잠재변수, 즉 상태-정밀도 (condition-2)

와 품질-정밀도 (quality-2) 다. VE 는 감정평가사의 조립도 등급을 근사치로 받아들여 그에 대해 회귀분석을 수행할 수 있으며, 그렇게 얻은 근사적 가치기여는 세밀한 잠재변수의 초기 값 배정에 유용하다. 그렇게 하는 경우의 프로토콜은, 이 두 조립도 변수의 회귀 추정 가치기여를 모형 추정치에서 빼내어 — 거래사례비교표의 회귀 구역 아래 측정된 속성에서 옮겨내어 — 잔차에 다시 더하는 것이다. 그런 다음 이를 잔차에서 제거하여 그에 대응하는 세밀한 잠재변수 `res_condition1` 과 `res_quality1` 에 배정한다. 이어서 MLS 의 `public remarks(public_remarks)` 와 관련 문자열 목록에서 추출한 추가 정보로 이들 잠재변수 값을 미세하게 다듬는다. 최종 결과는 세밀한 가치기여가 조립도 가치기여를 대체하는 것이다. 거친 등급은 근사 모형을 위한 발판 역할만 할 뿐이며, 상태와 품질에 대해 최종적으로 보고되는 값은 잔차에서 도출된 세밀한 값이다.

11.3 데이터 오류

소셜 미디어의 언급들에 비추어 볼 때, 캘리포니아 MLS 데이터는 미국의 다른 여러 지역에 비해 대체로 품질이 높다. 통상 MLS 데이터는 성숙한 대도시권에서는 양질일 것으로, 농촌 지역에서는 그보다 못할 것으로 예상해야 한다.

여러 이유로 데이터에 간간히 오류가 있기는 하나, 데이터가 충분하다면 이러한 오류는 대체로 서로 상쇄되어 가격 모형에 쓰이는 평균에는 미미한 영향만 미친다.

그러나 특정 지역에서는 체계적 편이가 발생할 수 있다. 예컨대 많은 주택이 애초에 2 층을 온전히 갖추어 지어졌으나, 일부 매수인에게 2 층 방을 없애 1 층의 한두 방에 높은 천장을 만드는 선택권이 주어져 총 주거면적 (GLA) 이 크게 줄어든 근린지역을 생각해 보라. 그러한 변경이 어떤 이유로든 이후 평가관 기록에 반영되지 않았다고 하자. 그 결과 실측값이 2,000 제곱피트인데도 MLS 에는 2,400 제곱피트로 나타날 수 있다. 그러한 경우 MARS 회귀분석을 수행하기 전에 MLS 데이터를 바로잡아야 한다. 비교사례에 대해 이 정보를 얻으려면 얼마간의 조사 노력이 필요할 수 있다는 점에 유의해야 한다. 불일치 사항은 후일의 참조를 위해 기록해 두어야 한다.

이상을 고려하면, 양질의 MLS 지역에서는 잔차로 흘러드는 오류가 디자인, 상태, 품질과 같은 주관적, 미측정 특성의 영향에 비해 미미하다고 보는 것이 합당하다. 이는 RCA 방법의 정확도를 추정할 때 중요한 고려사항이다. 예컨대 어느 시장 지역의 R^2 값이 80% 라면 가격의 약 20% 가 CQA 잔차에 귀속된다. 따라서 대상 부동산을 잔차 순위표에 놓는 데서 생기는 $\pm 10\%$ 의 오차는 최종 가치 결론에서 $\pm 10\% \cdot 20\% = \pm 2\%$ 의 오차로 전파된다. 정확도는 이렇게 높이는 것이다! 많은 경우 $\pm 1\%$ 를 훨씬 밑도는 정확한 가치 추정이 가능하다.

그러므로 잔차를 MARS 회귀분석에서 모형화되지 않은 변수들의 집합적 가치에 대한 간접 지표로 취급하는 것은 합당하다. 다만 감정평가공학자가 체계적 측정 편이가 어디에 왜 존재하는지를 식별할 만큼 충분한 지역 시장 경험을 갖추고 있어야 한다. 이 접근은 부동산 데이터의 한계와 미묘함을 인정하면서도 분석이 실제 부동산 가치를 정확히 반영하도록 보장한다.

실례: Sea Ranch 에서 측정되지 않은 가치결정요인 측정하기

어느 시장 지역에서 달성 가능한 R^2 가 그 시장의 외견상 동질성에서 기대할 법한 수준보다 낮을 때, 이 틀이 제시하는 대응은 그 한계를 받아들이는 것이 아니라 빠진 변수를 찾아내어 측정하는 것이다. 소노마 카운티의 Sea Ranch 지역사회가 이를 직접적으로 보여준다.

측정하라, 받아들이지 말라는 대응의 구체적 사례다. 소노마 카운티의 Sea Ranch 지역사회는 바다까지의 거리가 주된 가치결정요인인 해안 개발지다. 다른 특성이 같다면 절벽에서 수백 피트 떨어진 주택은 내륙으로 4 분의 1 마일 떨어진 주택과 상당히 다른 가격을 받는다. 이 지역사회의 MLS 데이터에는 바다까지의 거리가 변수로 기록되어 있지 않다. 사용 가능한 MLS 변수들로 적합시킨 초기 MARS 모형은 그 시장의 외견상 동질성이 뒷받침했어야 할 수준보다 낮은 설명력을 냈고, 잔차 구조는 데이터에서 보이는 어떤 특성으로도 부동산을 충분히 정렬하지 못했다. 이는 시장이 진정으로 잡음이 많다는 것이 아니라 중요한 결정요인이 모형에서 빠져 있다는 신호였다.

해법은 직접 측정이었다. 저자는 해안선을 기준으로 부동산 배치를 보여주는 지역사회 지도를 구

해 수백 채의 주택에 대해 바다까지의 거리를 측정했다. 측정된 거리를 MLS 데이터 집합에 변수로 추가하고 MARS 회귀분석을 다시 실행했다. 새 변수는 모형에서 유의미한 것으로 선택되었고, 설명력은 크게 높아졌으며, 그에 따라 잔차 순위표는 이전까지 그것을 오염시켰던 잠재적 바다-거리 신호가 아니라 남은 주관적 특성 — 조망, 상태, 품질 — 과 부합하는 양상으로 부동산을 정렬하게 되었다. MLS 변수가 없거나 신뢰할 수 없는 구릉지 근린지역에서는 표고를 같은 방식으로 다루어 왔다.

이 방법론적 원리는 일반적이다. 달성 가능한 R^2 가 그 시장의 외견상 동질성이 뒷받침해야 할 수준에 못 미치면, 첫 번째 가설은 중요한 변수가 모형에서 빠져 있다는 것이며, 적절한 대응은 그것이 무엇인지 찾아내어 직접 측정치를 수집하는 것이다. 자료원으로는 지역사회 지도, GIS 데이터, 세무 평가관 기록, 건축 당국 기록, 위성 영상, 현장 조사 등이 있다. 그러한 측정 작업의 비용은 실재하지만 한정되어 있다. 특정 시장 지역에 대해 변수를 한 번 수집해 두면 그 지역의 이후 업무에 그대로 이어 쓸 수 있다. RCA 는 새로운 변수를 깔끔하게 흡수하는데, 이 틀은 어떤 변수가 MARS 회귀분석에 들어가는지에 무관하기 때문이다. 새 변수가 실재하는 신호를 담고 있는지는 잔차 구조가 VE 에게 알려 주며, 가치기여는 그에 따라 조정된다.

대안적 대응 — 낮은 R^2 를 받아들이고 더 넓은 오차범위를 보고하는 것 — 은 빠진 결정요인이 진정으로 측정 불가능할 때 (매수인 개인의 선호, 일시적 시장 심리, 특이한 거래 사정) 에는 방어 가능하지만, 단지 수집되지 않았을 뿐일 때는 그렇지 않다. 모형이 약할 때 VE 가 먼저 할 일은 빠진 신호가 후자인지를 묻는 것이다.

11.4 정확성이란 무엇인가

감정평가에서 완전한 정확성을 갖추려면 다음이 필요하다.

1. 시장가치에의 부합: 대상 부동산의 가상 거래가 주어진 시장가치 정의에 100% 부합할 것
2. 정확하고 견고한 모형 개발: 지역에 따라 R^2 약 75 80%, CVR2 대략 55 70% 의 고품질 MARS 모형이 개발될 것
3. 충실한 데이터 오류 분석: 측정 변수의 있음직한 데이터 오류가 모두 파악되어 통제되고 잔차로 흘러드는 부분이 거의 없을 것
4. 체계적 편이의 반영: 데이터의 체계적 편이가 모두 반영되고 보정될 것
5. 잔차 순위 분석: 잔차 순위상 부동산들이 최하위에서 최상위까지 특성의 합리적 양상을 따르는지 확인할 것. 상속재산 처분 거래, 숏세일, 경매 거래와 같은 이상치는 조사하고 설명하여 필요하면 제외할 것
6. 대상 부동산의 위치 결정: 잔차 순위를 충분히 이해한 상태에서 대상 부동산을 정확히 자리매김할 것. 매력이 뚜렷이 높은 부동산과 뚜렷이 낮은 부동산을 각각 찾아 그 사이에 대상 부동산을 놓고 점수를 매긴다. 그런 다음 위아래 부동산 사이의 간격을 가능한 한 좁혀 나간다.
7. 설명적 추론: 정확성을 위해 설명적 추론을 결코 희생하지 말 것. 감정평가사는 적어도 통상의 지력을 갖춘 사람에게 대상 부동산과 비교사례 사이의 가치 차이를 설명할 수 있어야 한다. 예컨대 R^2 와 CVR2 가 최적인 좋은 MARS 모형은 측정된 특성의 가치기여를 설명하는 데 쓸 수 있고, 감정평가공학자의 잔차 분해는 그 밖의 특성의 가치기여를 설명하는 데 쓸 수 있다.
8. 과적합: 단독으로든 조합으로든 특정 부동산 한둘을 식별할 수 있는 특성을 사용하지 않음으로써 모형의 과적합을 피할 것. 특히 그 값이 소수의 부동산을 특정할 수 있는 팩터 변수의 사용을 피하라. 규모 있는 데이터 집합에서 근린지역을 식별하는 팩터 변수는 적절하지만, 수영장이 거의 없는 지역에서 수영장 보유 부동산을 식별하는 변수는 신중하

게 써야 한다. 정확한 GIS 좌표, 필지번호, 주소처럼 개별 부동산을 특정하는 특성은 절대 허용되지 않는다.

회귀 R² 가 80% 라면 대상 부동산 가치의 약 20% 가 잔차에서 비롯되리라 예상한다. 다만 이는 근사일 뿐이다. CQA-잔차 곡선으로부터 우리는 잔차 분포의 대부분이 CQA 의 낮은 쪽과 높은 쪽 극단에 있음을 안다. CQA-잔차 곡선과 낮은 CQA 값 및 높은 CQA 값에서의 가파른 기울기에 비추어 보면, 이 구간에서는 잔차의 추정이 덜 정확하리라 예상할 수 있다. 그러나 이는 이 구간에 부동산이 상대적으로 드물다는 점과 인접한 부동산 사이의 매력 차이가 더 크다는 점으로 상쇄된다. 그리하여 중간 구간에서는 $\pm 5\%$ 범위의 영향을 말할 수 있는 반면, 극단에서는 다소 좁은 $\pm 2 \sim \pm 3\%$ 범위를 더 유념하게 된다.

RCA 의 정확도를 계산하는 일은 CQA-잔차 곡선의 비선형성 때문에 일반적으로 다소 어렵다. 그러나 합리적인 잔차 순위표가 있다면 사안별로는 가능하다. CQA 점수의 범위는 0.0 에서 10.0 이다. 좋은 MARS 모형이 있다면 잔차 순위표는 매력이 가장 낮은 것에서 가장 높은 것까지 뚜렷한 양상을 보여야 하며, 전체 범위의 $\pm 5\%$ 이내, 즉 CQA 점수로 ± 0.5 이내로 평정하기가 어렵지 않을 것이다. 회귀분석에 부동산 100 건이 들어갔다면 이는 대상 부동산을 10 건의 무리 안에서 정확히 자리매김하는 것에 해당한다. 그러니 표 3의 Burlingame 거래 데이터를 참조하여, VE 가 추정한 CQA 가 7.5, 5.1, 1.9 이고 대상 부동산에 대한 그의 가치 결론이 \$3,000,000 일 때의 근사적 정확도 (즉 오차) 를 계산해 보자. 표 3에는 사용된 수백 건 가운데 표본만 실려 있으므로 필요한 곳에서는 (선형성을 가정하여) 잔차를 보간해야 한다. 분명히 해 두자면, 여기서는 VE 가 각 경우에 100 건뿐인 잔차 순위표에서 대상 부동산의 순위를 ± 5 건 범위 안으로 매우 확신한다고 가정한다.

CQA= 7.5 ± 0.5					
CQA	잔차	절대차	예상 오차	가상 거래가격	오차율
8.00	\$257,112				
7.50	\$197,113	\$59,999			
7.00	\$154,615	\$42,498	$\pm \$51,249$	\$3,000,000	$\pm 1.71\%$

CQA=5.1 ± 0.5					
CQA	잔차	절대차	예상 오차	가상 거래가격	오차율
5.60	\$50,911				
5.10	\$19,559	\$31,352			
4.60	-\$28,207	\$47,766	$\pm \$39,559$	\$3,000,000	$\pm 1.32\%$

CQA= 1.9 ± 0.5					
CQA	잔차	절대차	예상 오차	가상 거래가격	오차율
2.40	-\$228,411				
1.90	-\$276,381	\$47,970			
1.40	-\$326,938	\$50,557	$\pm \$49,264$	\$3,000,000	$\pm 1.64\%$

끝으로, 정확성을 검증하는 유일하게 객관적인 방법은 엄격한 프로토콜 (이 논문은 그 큰 줄기만 부분적으로 시사할 뿐이다) 을 세우고 그 준수 여부를 검증하는 것이다. 실제 거래가격은 시장의 불완전성 때문에 시장가치에서 벗어날 수 있으므로 이것이 필요하다. 감정평가와 이후의 거래를 비교하면 정확성에 관한 어느 정도의 시사를 얻을 수 있다. 그러나 부동산 시장의 복잡성과 불안정성 때문에, 확립된 기준과 프로토콜에 기초하여 가치 추정치를 제공하려면 감정평가사나 고도로 자격을 갖춘 감정평가공학자의 전문성이 필요하다. 그 기준과 프로토콜은 유능한 감정평가공학자들이 독립적으로 거의 동일한 가치 결론에 이르도록 보장할 만큼 엄격해야 한다. 좋은 데이터가 있는 대부분의 경우 이는 달성 가능한 목표일 것이다.

11.5 이상치로서의 대상 부동산

부동산의 시장가치를 추정하려면 최근 거래사례의 분석이 필요하므로, 현대화와 수선이 이루어지지 않은 노후한 대상 부동산의 여러 특성을 최근 거래사례가 잘 대표하는지 물어보는 것은 마땅하다. 최근 거래사례가 통상 매도가격을 높이려고 개보수를 거친 주택인 샌프란시스코 베이 에어리어에서는 특히 그러하다. 그리하여 비교 거래사례의 평균적 품질과 상태는 꽤 오랫동안 시장에 나오지 않았던 노후 주택보다 대체로 우월하다. 이러한 가치 차이는 대상 부동산의 잔차 추정, 달리 말해 그 CQA 점수에서 포착되어야 한다. 이는 대상 부동산을 평가할 때 데이터 품질과 정확성과 관련하여 중요한 고려사항이다.

11.6 예상 밖 데이터를 다루는 데서 CVR2의 중요성

MARS 회귀분석은 무작위로 선택된 부동산 부분집합에 $R/earth$ 를 여러 차례 반복 적용하는 데 기초하는데, 이 부분집합은 통상 MLS 데이터 집합의 80% 에서 90% 에 해당하며 훈련 집합이라 불린다. 나머지 10% 20% 의 부동산이 검정 집합이 된다. 이 단계에서 훈련 데이터 집합으로 MARS 모형을 만들고, 그 모형으로 검정 데이터 집합에 속한 부동산의 거래가격을 예측한 뒤 실제 거래가격과 비교한다. 이 방법론은 아직 관측되지 않은 부동산의 거래가격을 예측할 수 있는 견고한 모형을 개발하는 가장 효과적인 전략으로 데이터 마이닝 분야에서 인정받고 있다.

그러나 CVR2 지표의 일차적 효용은 실제 시장 지역 거래에 관한 회귀 모형의 사실적 정확성을 반영하는 데 있다기보다, 특정 기준일 현재의 가상적 부동산 거래에 관한 모형의 예측력을 최적화하는 데 있다. 그러한 부동산은 훈련 데이터 집합에 포함되어 있지 않으며, 훈련 데이터와 검정 데이터 어디에도 없는 새롭거나 예상치 못한 특성 조합을 보일 수 있다.

이러한 맥락에서, 예컨대 감정평가 기준일 한 달 전에 매물로 나왔고 지난 15 년간 개보수가 없었던 부동산의 거래가격은 어떻게 추정할 수 있을까? 반복적 교차검증을 거치지 않은 모형보다는 강한 CVR2 를 보이는 모형이 가장 정확한 추정치를 제공할 가능성이 높다. 또한 데이터 집합에 개보수, 수선, 현대화의 정도가 다양한 비교사례가 충분히 포함되어 있는 것이 필요하다.

궁극적으로, 비교사례 데이터 집합에서 여러 부동산 특성의 가치기여를 정확히 예측하는 견고한 모형을 갖고 있다면, 잔차로 정렬된 그 부동산들의 순위표 안에 새로운 대상 부동산을 효과적으로 자리매김할 수 있다. 이렇게 위치를 정하면 해당 부동산에 적절한 CQA 점수와 잔차를 배정할 수 있다.

12 잔차

“잔차 제약 접근법”이라는 이름이 말해 주듯, 이 방법론에서 잔차는 중심적이다. 여기서 잔차란 어느 부동산의 순 거래가격과 회귀분석이 예측한 거래가격의 차액을 말한다.

양의 잔차는 총 주거면적 (GLA) 이나 대지면적과 같은 통상의 측정 가능한 속성에 근거한 추정가격보다 그 부동산이 비싸게 팔렸음을 시사하며, 이는 평균보다 매력이 클 수 있음을 함의한다. 음의 잔차는 더 싸게 팔렸음을 뜻하며, 매력이 덜하거나 포착되지 않은 다른 부정적 속성이 있음을 시사할 수 있다.

큰 주택일수록 잔차도 클 가능성이 높으므로, 보다 정규화된 비교를 위해 GLA 기준 제곱피트당 잔차를 쓸 수 있다. 이 문제에 대한 더 깊은 논의는 다른 논문으로 미룬다.

12.1 순위와 점수

측정 변수에 대한 MARS(다변량 적응 회귀 스플라인) 모형을 만든 뒤 각 부동산에 대해 잔차를 계산한다. 이 잔차는 관측된 순 거래가격과 모형이 예측한 거래가격의 차액이다. 잔차/SF, 즉 잔차를 주택의 주거면적으로 나눈 값을 값을 열을 하나 더 만들어야 한다. 대체로 특정 부동산의 잔차는 주거면적이 클수록 크다는 이해에서다. 근린지역에 따라서는 다른 통계량이 더 잘 맞을 수도 있으나, 그것은 다음 기회의 주제다.

가정 7: “잔차에 의한 순위”

이 논문에서 “잔차에 의한 순위”라는 표현은 잔차에 의한 순위, 잔차/SF에 의한 순위, 또는 감정평가공학자가 해당 시장 지역에 가장 적합하다고 판단하는 잔차의 다른 함수에 의한 순위를 뜻하는 것으로 읽어야 한다. SF 베이 에어리어에서의 저자의 경험에 따르면, 특히 GLA(총 주거면적)가 유난히 작거나 큰 대상 부동산을 다룰 때는 잔차/SF가 대체로 가장 믿을 만하다.

다음으로 부동산을 잔차가 가장 작은 (음의 값이 가장 큰) 것에서 가장 큰 것 순으로 정렬한다. 그런 다음 각 부동산에 0.0에서 10.0 범위의 상태-품질-매력 (CQA) 점수를 부여한다. 이 점수는 다음과 같이 계산한다.

CQA 점수는 전체 부동산 가운데 그 부동산 잔차의 백분위 순위에 대응한다. 구체적으로는 해당 부동산보다 잔차가 낮은 부동산의 비율을 구한 뒤 그 비율을 10으로 나누어 계산한다. 예컨대 어떤 부동산의 잔차가 전체의 50%보다 높다면 그 점수는 5.0이 된다.

반올림 때문에 잔차가 가장 큰 부동산 하나가 10.0을 받을 수 있다. 그러나 논리적으로 자기 자신을 포함한 100%의 부동산보다 잔차가 높을 수는 없으므로 만점 10.0은 이론상 불가능하다. 반면 최저점 0.0은 가능한데, 적어도 한 부동산은 자기보다 잔차가 낮은 부동산이 없기 때문이다. 비교 대상 부동산이 많으면 반올림 때문에 둘 이상이 0.0을 받을 수도 있다는 점에 유의하라.

따라서 CQA 점수는 어느 부동산의 속성이 모형의 기대에 얼마나 부합하는지를 직관적으로 보여주는 척도이며, 점수가 높을수록 모형의 예측에 더 잘 부합함을 뜻한다.

CQA 점수가 비교사례의 매력에 대한 간접 지표로 여겨질 수 있는 것은 거래 조건이 그 시장가치에 상응할 때뿐임을 인식하는 것이 필요하다. 감정평가공학자는 거래에 대한 포괄적 설명과 철저한 조사를 통해 잔차값에 시장가 초과 또는 미달 요소가 포함되어 있는지를 확인할 책임이 있다. 시장가를 크게 넘거나 밑도는 거래가 잦은 것은 아니지만, 강제 매각, 상속재산 처분, 무능력, 담합 등의 사정에서 일어날 수 있다.

표 3은 MLS 데이터의 준현실적 발제를 잔차 및 CQA 점수와 함께, 총액 기준과 제곱피트당 기준으로 제시한다. 이 연구에서 분석된 부동산의 총수는 약 600건이었다. 회귀에 투입된 여러 변수가 표에 표시되지 않았다는 점은 유의할 만하다. 달리 말해 여기 제시된 잔차는 거래일과 같은 다른 변수의 영향을 MARS 회귀분석이 이미 고려한 뒤의 값이다.

표 3: 잔차로 정렬한 비교 거래사례

SalePrice	Estimated SP	Residual	CQA	Residual/SF	CQA SF	GLA	SaleDate
\$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-잔차 곡선

어느 근린지역에 대해 CQA 대 잔차/SF 를 그리면 예외 없이 [그림 2](#)에 보이는 CQA-잔차 특성 곡선이 나온다. 이 그림은 2022 년경 캘리포니아 Burlingame 의 데이터를 나타낸다. CQA 점수를 잔차값에 대해 그릴 때도 대체로 비슷한 곡선이 관찰되지만, 나는 대상 부동산의 총 주거면적 (GLA) 과 관련하여 CQA 를 맥락화하는 데 유용하다는 점에서 잔차/SF 를 선호한다. 잔차/SF 에 GLA 를 곱하면 어느 부동산의 CQA 를 그에 대응하는 잔차값으로 환산할 수 있다.

감정평가공학자는 부동산의 규모, 구체적으로는 총 주거면적 (GLA) 이 잔차의 크기에 유의미한 영향을 주는지를 신중히 판정해야 한다. 이 판정은 부동산 가치를 계산하는 가장 정확한 방법을 정할 수 있으므로 결정적이다. 예컨대 고급 주방기와 호화로운 욕실 설비는 주택의 잔차값을 크게 높일 수 있다. 그러나 주방과 욕실의 크기와 배치가 대체로 균일한 근린지역에서는 이러한 업 그레이드가 GLA 와 강하게 상관하지 않을 수 있다. 그것이 잔차값에 미치는 영향은 차지하는 공간보다는 그것이 나타내는 품질에 더 관계될 수 있다.

반대로 부동산의 외관 매력을 눈에 띄게 높이는 값비싼 외부 요소와 자재는 부동산의 규모와 더 직접적인 관계를 가질 수 있다. 큰 주택일수록 광범위한 조정, 우수한 지붕 자재, 넓은 파티오 같은 더 정교한 외부 구성을 수용할 수 있고, 이는 자연스럽게 부동산의 전반적 매력과 그에 따른 잔차값을 높인다.

그러므로 감정평가공학자가 잔차를 분석할 때 이러한 동학을 고려하는 것은 필요하다. 고품질 설비와 특성이 더하는 내재적 가치를 판정하는 데 그치지 않고, 그러한 개선이 부동산의 물리적 규모와 어떻게 상호작용하는지도 이해해야 한다. 이러한 포괄적 접근은 감정평가가 부동산의 질적 속성과 양적 속성을 모두 정확히 반영하도록 보장한다.

따라서 CQA-잔차 곡선과 CQA-잔차/SF 곡선을 모두 그려 그 차이를 살펴볼 것을 권한다. 그 차이가 시장 지역의 양상을 시사할 수 있다. 감정평가공학자는 프로그램으로 두 잔차를 모두 계산한 뒤 가중평균을 취하여 대상 부동산에 배정할 잔차를 산출할 수 있다. 차이가 클 수도, 크지 않을 수도 있다.

12.3 CQA 곡선의 특성

상태-품질-매력 (CQA) 곡선은 특정 지역 안에서 부동산 가치가 매력, 상태, 시장 동학에 따라 어떻게 변동하는지를 이해하는 정교한 모형을 제공한다. 이 모형은 부동산 시장에 영향을 주는 경제적, 물리적 요인의 상호작용을 포착하며, 시장 부문별 감정평가의 미묘한 실상을 반영한다.

12.3.1 CQA 곡선 구성요소의 분석

1. 곡선 좌측의 지수적 감쇠:

CQA 점수가 2 부근에서 0 으로 낮아짐에 따라 부동산은 급속한 노후화의 특징을 보인다. 곡선의 이 부분은 지수적 감쇠 함수로 잘 표현되며, 적절한 유지관리와 개보수 없이 부동산이 얼마나 빠르게 매력과 가치를 잃는지를 부각한다. 목재의 부식, 색의 바램, 금속의 녹과 같은 물리적 열화가 이러한 하락을 가속하며, 정기적으로 관리되지 않는 주택에서 특히 그러하다.

2. 중간 구간의 선형적 증가:

시장의 대부분이 통상 이 범주에 들며, 점수 2 부근에서 8 까지 매력이 거의 선형으로 증가한다. 이 구간은 대체로 표준적으로 관리되고 있으나 반드시 특출한 품질이나 입지를 갖추지는 않은 부동산을 반영한다. 이는 주택의 광범위한 중간층을 대표하며, 표준적 특성과 전반적 상태를 근거로 경쟁적으로 가격이 매겨지는 주택시장의 대다수를 아우른다.

3. 곡선 우측의 지수적 증가:

여기서는 점수 8 에서 10 까지 곡선이 지수적으로 상승하며, 시장의 매력이 높은 쪽 끝을 반영한다. 이 구간의 부동산은 흔히 더 부유한 이들의 것으로, 그들은 뛰어난 유지관리뿐 아니라

부동산의 매력과 가치를 크게 높이는 개선에도 투자한다. 지수적 증가는 그러한 고품질 부동산의 희소성과 부유한 매수인들 사이의 높은 수요를 반영한다. 그리고 가격이 낮은 주택이 매력 면에서는 매우 높게 평정될 수 있고, 크고 비싼 주택이 낮게 평정될 수도 있다는 점을 이해하라. 이러한 동학은 MARS 를 전문적으로 활용할 때 드러나며, 대부분의 감정평가사는 대개 이를 알아채지 못한다.

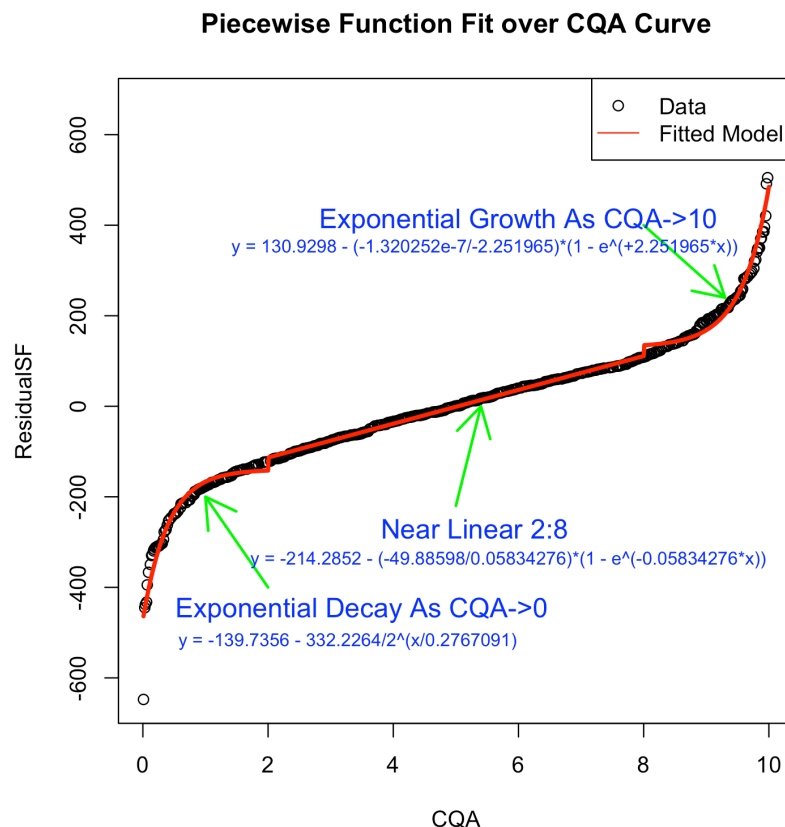
12.3.2 시장 동학과 사회적 함의

CQA 곡선은 지역마다 다른 상대적 부와 경제적 격차의 이야기를 은연중에 서술한다. San Mateo 카운티 같은 곳에서 Burlingame 과 Hillsborough, Woodside, Atherton 같은 더 부유한 근린지역 사이의 부동산 가치 편차는 상당한 사회경제적 분단을 보여준다. 이 격차는 부동산 유지관리에서 시장 평가에 이르는 모든 것에 영향을 주며, 궁극적으로 매수인 구성과 부동산 시장의 추세에도 영향을 미친다.

나아가, 비교적 부유한 사람이 스스로 편안하게 여기고 여러 이유로 살고 싶어 하는 중간 등급 근린지역에 고품질 주택을 마련한다는 일화는 부동산 경제의 또 다른 결정적 측면 — “과잉건축 (overbuilt)” 부동산이라는 개념 — 을 부각한다. 그러한 부동산은 매우 매력적이지만, 더 높은 가격대의 시장 지역에서라면 받았을 값을 치를 수 있는 현지 매수인을 찾지 못할 수 있다.

CQA-잔차 곡선은 시장 지역마다 고유하므로, 분석을 할 때 간과해서는 안 된다.

그림 2: CQA-잔차 특성함수



12.4 대상 부동산의 잔차

비교사례와 달리 대상 부동산에는 객관적으로 확정된 잔차가 없으므로, 감정평가공학자 (VE) 가 이를 추정해야 한다. 이 추정은 잔차 제약 접근법 (RCA) 프로토콜에 어느 정도의 주관성을 들여 오며, 개인의 판단이 결정적 역할을 하는 중대한 분기점이 된다.

상태-품질-매력 (CQA) 점수를 추정하기 위해 VE 는 순위화된 부동산들을 면밀히 분석하여 순위 안에서 식별 가능한 양상을 찾는다. 이 분석에는 각 부동산에 딸린 MLS 사진의 상세한 검토가 포함되는 경우가 많다. 600 건 같은 큰 데이터 집합에서 이를 남김없이 검토하기란 현실적이지 않다. 그 대신 VE 는 실무적 추론을 적용하여 전체 순위 안에서 대상 부동산의 있음직한 위치를 어렵다. 대대적인 수선이 필요한 부동산, 이른바 “픽서 (fixer)” 는 대체로 하위에 놓이고, 개보수와 매력이 뛰어난 부동산은 상위에 오르는 경향이 있다.

그런 다음 VE 는 더 좁혀 들어가는 접근을 취하여, 순위 안에서 “건너뛰며” 대상 부동산의 매력에 근접하는 비교사례를 찾는다. 범위가 좁혀질수록 VE 는 품질, 상태, 디자인 미감 (도장, 목공, 석공, 창호 등의 요소를 포함), 기능적 효용, 그 밖의 관련 특성 등에서 부동산들 사이의 유사성에 더욱 주의를 기울인다.

VE 의 판단이 근본적으로 시장이 형성한 순위에 이끌린다는 점을 이해하는 것이 결정적이다. 이로써 절차의 주관적 요소가 시장의 실상에 맞을 내리게 되며, 대상 부동산에 대한 공정하고 정확한 평가를 지향하게 된다.

12.5 RCA 가치 결론

12.5.1 일반적 고려사항

CQA 점수의 판정은 부동산의 추정 잔차값을 확정하는 데 결정적인데, 이 둘이 본질적으로 연결되어 있기 때문이다. CQA 없이도 부동산을 평가할 수는 있으나, 이 지표는 해당 부동산이 시장 지역의 다른 부동산과 견주어 어디에 서 있는지를 보여줌으로써 유용한 맥락을 제공한다. 예컨대 CQA 점수 3.2 는 그 부동산이 인근 비교 부동산의 32% 보다 매력이 크다는 뜻이다. 부여된 점수에 이의가 있다면, 이의 제기자는 순위표를 참조하여 사진 증거나 그 밖의 실질적 데이터에 근거해 제안하는 조정을 정당화해야 한다. 순위 체계가 견고하고 감정평가공학자 (VE) 가 그 틀 안에서 부동산을 정확히 자리매김했다면 유의미한 불일치가 생길 가능성은 낮다.

보고 단계에서는 통상 가치 결론을 뒷받침할 최근의 적절한 비교 거래사례 6 12 건을 선정한다. 이 비교사례와 대상 부동산의 차이를 분석하고 변수마다 그에 맞게 보정한다. 그런 다음 이 보정치를 비교사례의 순 거래가격에 적용하여 보정 후 거래가격을 도출한다. 수학적으로 밝힌 바와 같이, 이 보정 후 가격들은 대상 부동산의 추정 거래가격으로 수렴해야 한다.

12.5.2 검토, 감사, 이의

대상 부동산의 추정 잔차가 확정되면 모든 비교사례에 대한 잔차 보정치를 계산할 수 있다. 이 보정치는 VE 가 적절하다고 판단한 특정 CQA 변수에 맞추어 이후 수작업으로 분해할 때의 제약으로 작용한다. 이러한 상세 분해는 설명을 위한 것일 뿐 최종 평가액에는 영향을 주지 않는다는 점을 강조해 둘 필요가 있으며, 이는 증명 II 가 확인한 바다.

감정평가 가치 결론에 만족하지 못한 이용자의 이의와 관련하여, 전통적 감정평가사는 가격이 더 높거나 낮은 이런저런 부동산이 왜 이용자가 기대한 만큼 가치에 영향을 주지 않는지 설명하느라 시간을 허비하는 것이 보통이다. RCA 에서는 가능한 모든 비교 거래사례 (통상 100 600 건) 가 대체로 가격 결론과 정확히 같은 값으로 평가되므로, 임의의 부동산에 대한 보정치를 보여주는 데 품이 거의 들지 않는다. RCA 절차 전체는 대상 부동산의 잔차 점수 매기기를 제외하면 자동화되어 있고 객관적이다. 그리하여 RCA 의 또 다른 이점은 VE 의 주관적 판단이 지닌 불완전성에 대한 비판이 주로 그 한 영역에 국한된다는 것이며, 이는 효과적이고 효율적으로 관리할 수 있다. 또

한 R^2 가 약 80% 라고 할 때, 대상 부동산 잔차 추정에서의 $\pm 10\%$ 오차는 최종 가치 추정치의 $\pm 2\%$ 오차에 해당한다.

많은 경우 대상 부동산이 잔차 순위표에 매우 잘 들어맞아 정확도가 $\pm 1\%$ 이내일 것으로 기대할 수 있다.

추정 잔차와 대상 부동산의 MARS 추정 거래가격의 합이 잔차 제약 접근법 (RCA) 가치 결론이라 불리는 최종 가치 결론을 이룬다. 이 절차는 부동산 감정평가에 대한 포괄적이고 투명한 접근을 보장하여 평가의 정확성과 공정성에 대한 신뢰를 키운다.

13 결론

잔차 제약 접근법은 세 가지 실질적 기여를 갖는 틀이다. 첫째, 소수의 비교사례 집합이 아니라 시장 전체에 대해 모형을 적합시키는 데서 도출되는 구조적 반정박 (anti-anchoring) 성질 (증명 I). 둘째, 잔차 분해가 가치 결론을 흐트러뜨리지 않으면서 설명의 층위 역할을 할 수 있게 하는 불변성 결과 (증명 II 와 그에 이어지는 논평). 셋째, 지수-선형-지수의 특징적 형태로 여러 시장 지역에 걸쳐 되풀이하여 나타나는 것으로 보이는 경험적 규칙성 — CQA-잔차 특성곡선. 이 틀의 한계 또한 그에 못지않게 실질적이다. 잔차 분해는 불변이지만 식별적이지는 않고, CQA 위치 결정 단계는 아직 정식 평정자 간 신뢰도 연구로 검증되지 않은 VE 의 판단에 의존하며, 정확성에 관한 주장은 대안적 방법과의 벤치마킹을 기다리고 있다.

이 논문이 제시하는 것은 하나의 틀과 프로토콜이다. 아직 제시하지 못했으나 현재 진행 중인 더 큰 연구 프로그램이 제시하는 것은, 방어 가능한 기본값과 함께 그 프로토콜을 소프트웨어로 구현한 실행 가능하고 감사 추적 가능한 파이프라인이다. 그 파이프라인 — MARS 기저함수 단계를 위한 CRAN 의 earthUI 패키지, 그 기저 위에서 정칙화된 구조 모형을 다루는 glmnetUI 패키지, 공간 평활을 위한 mgcvUI 패키지 — 은 이번 호의 관련 논문들이 다루는 주제이며, 방법으로서의 RCA 에서 산출물로서의 RCA 로 건너가는 다리다. 감정평가 규제에서의 교리-실무 간극 (Craytor, 2026, 이번 호) 은 그 다리가 왜 중요한지를 알려 준다. 감정평가의 산출물은 경제적으로 미래 지향적 위험 투입값으로 소비되며, 현재가치 의견을 감사 가능하고 재현 가능하게 만드는 틀은 그러한 미래 지향적 용도와 정직하게 대면하기 위한 선결조건이다. RCA 는 그 더 큰 프로젝트의 한 수 (手) 일 뿐이다. 프로젝트 전체가 아니며, 이 논문도 그렇게 주장하지 않는다.

참고문헌

- 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) 패키지를 위한 대화형 인터페이스

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

William Bert Craytor

제 1 권 제 1 호 · 2026 년 여름호

발행일: 2026 년 7 월 1 일

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

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

이 오픈 액세스 논문은 [크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스 \(CC BY 4.0\)](#)의 조건에 따라 배포됩니다.

초록. earthUI 는 다변량 적응 회귀 스플라인 (MARS, multivariate adaptive regression splines) 을 구현한 R `earth` 패키지의 그래픽 사용자 인터페이스 (graphical user interface) 이다. 일반 예측 모형화, 부동산 감정평가, 시장지역 분석이라는 세 가지 용도 모드를 제공하며, 데이터 가져오기, 모형 구성, 적합, 그리고 `earth` 의 진단 정보와 그래픽 출력의 해석에 이르기까지 사용자를 안내한다. 이 논문은 earthUI 의 데이터 형식 요건, 모형화 작업 흐름, 출력 화면, 그리고 전체 기능 참조를 문서화한다.

키워드: MARS, 다변량 적응 회귀 스플라인, `earth`, R, 그래픽 사용자 인터페이스, 부동산 감정평가

발행인 주 (2026 년 7 월): 이 논문은 earthUI 0.10.0 을 다룬다. 인쇄 직전에 earthUI 0.11.0 이 자매 응용 프로그램인 `glmnetUI` 와 `mgcvUI`(이어지는 두 편의 논문에서 설명한다) 를 흡수하여 MARS, 엘라스틱 넷, GAM 작업 흐름을 하나의 응용 프로그램으로 통합하였다. 여기에 서 설명하는 인터페이스와 작업 흐름은 이번 통합의 영향을 받지 않는다.

목차

1	서론	3
1.1	earthUI 란 무엇인가?	3
1.2	하나의 패키지, 세 개의 루틴	4
1.3	세 가지 용도 모드	4
1.4	부동산 전용 기능	4
1.5	earthUI 와 vProlog 설치	5
1.6	시작하기	5
2	MLS 입력 데이터 요건	7
2.1	파일 형식과 구조	7
2.2	감정평가 모드에 필요한 열	7
2.3	특수 열 명명 규약	8
2.4	데이터 품질과 완전성	8
2.5	대상 부동산 행의 배치	8
3	General 용도 모드	9
3.1	개요	9
3.2	사이드바 작업 흐름	9
3.3	주 패널 탭	9
3.4	설정 지속성	10
3.5	다크 모드	10
3.6	로컬 및 지역 설정	10
3.6.1	지원 국가	11
3.6.2	용지 재정의	11
3.6.3	기본값 저장	11

4 Appraisal 모드	12
4.1 대상 부동산 행의 처리	12
4.2 기준일과 Sale Age	12
4.3 특수 열 지정	12
4.4 잔차 기여도 (회귀 특성 추정)	13
4.5 RCA 보정 개요	13
5 Market Area Analysis 모드	15
5.1 Appraisal 모드와의 차이	15
5.2 시장 모드를 쓸 때	15
6 변수 선택	16
6.1 목표 (반응) 변수	16
6.2 예측변수 표	16
6.3 자료형 감지와 재정의	16
6.4 Factor 및 Linear 플래그	17
6.5 특수 열 유형 참조	17
6.6 다중 목표변수	18
7 매개변수 선택	19
7.1 기본 매개변수	19
7.2 전진 단계	19
7.3 허용 상호작용	20
7.4 가지치기	20
7.5 교차검증	20
7.6 부분집합 필터링	20
7.7 권장 매개변수 값	21
7.7.1 전진 단계 매개변수	21
7.7.2 가지치기 매개변수	22
7.7.3 교차검증 매개변수	22
7.7.4 분산 모형	23
7.7.5 Earth 출력 파일	23
7.8 설정 기본값	23
8 Earth 모형 적합	24
8.1 Fit 버튼	24
8.2 적합 모달과 추적 출력	24
8.3 적합 로그	24
9 데이터 내려받기	25
9.1 출력 열	25
9.2 열 순서와 Excel 형식	25
9.3 CQA 점수	26
9.4 정렬과 대상 부동산 행	26
9.5 mgcvUI 자동 내보내기	26
9.6 glmnetUI 가져오기	26
9.6.1 주요 이점	27
9.6.2 중요: 가중치 열이 일치해야 한다	27
9.6.3 작업 흐름	27

10 RCA 계산과 내려받기	28
10.1 RCA 대화 상자 열기	28
10.2 CQA 점수 보간	28
10.3 출력 열	28
10.4 RCA Adjustments 탭	29
11 거래사례비교표	30
11.1 개요와 목적	30
11.2 비교사례 선택 모달	30
11.3 표 배치	30
11.4 매매가격, 양허, Net SP	31
11.5 묶음 행 (위치, 부지, 연령)	31
11.6 모형 변수 행	31
11.7 CQA Residual 및 잔차 특성 행	31
11.8 보정 매매가격 수식	32
11.9 시트 보호와 편집 가능 셀	32
11.10 Excel 에서 표 다루기	32
12 보고서 내려받기	33
12.1 Generate Quarto Report	33
12.2 Convert Quarto Report	33
12.3 보고서 내용	33
13 후처리 모듈: 엘라스틱 넷과 GAM	35
13.1 왜 보조 적합인가?	35
13.2 실행과 상태 공유	35
13.3 엘라스틱 넷 루틴	35
13.4 GAM 루틴	35
13.5 실무적 작업 흐름	35
14 데모 데이터셋: Appraisal_1.csv	37
14.1 설명	37
14.2 열	38
14.3 권장 빠른 시작	38
15 실험적 기능: Prolog 텍스트 추출과 도출 규칙	40
15.1 무엇을 하는가	40
15.2 요구 사항: SWI-Prolog 가 아니라 vProlog(Trealla)	40
15.3 활성화와 실행	41
15.4 도출 규칙	41
15.5 버전이 매겨진 활성 데이터 파일	41
15.6 프로그램 인터페이스와 함께 제공되는 파일	42
15.7 관련 실험: 기후 지역 사전정보	42
15.7.1 캘리포니아를 넘어: 관할권 팩과 시장지역 모듈	42

1 서론

1.1 earthUI 란 무엇인가?

earthUI 는 다변량 적응 회귀 스플라인 (MARS, multivariate adaptive regression splines) 을 구현한 R earth 패키지의 그래픽 사용자 인터페이스 (graphical user interface) 이다 (Friedman 1991).

모형화 엔진은 Stephen Milborrow 의 **earth** 패키지이며 (Milborrow 2024), Milborrow (n.d.[a]) 및 Milborrow (n.d.[b]) 에 상세히 문서화되어 있다. earthUI 는 로컬 Shiny 응용 프로그램으로 실행되며, 로그인도 서버도 계정도 없다. R 에서 실행한 뒤 데이터셋 (CSV 또는 Excel) 을 가져오고, 모형을 구성하여 대화형으로 적합시킨다.

이 응용 프로그램은 데이터 가져오기, 변수 구성, 모형 적합, 진단 그림, 변수 중요도, 모형 방정식, 그리고 HTML, Word, PDF 형식으로 내려받을 수 있는 보고서에 이르는 완전한 작업 흐름을 제공한다.

1.2 하나의 패키지, 세 개의 루틴

현재 릴리스 기준으로 earthUI 패키지는 두 개의 후처리 모듈 (Post Processing Module) 까지 포함하는데, 이전 릴리스에서는 glmnetUI 와 mgcvUI 라는 별도의 패키지였다. earthUI 를 설치하면 세 가지 루틴이 모두 설치되며, 각각은 자체 로컬 Shiny 응용 프로그램으로 실행된다:

- **earthUI::launch()** — 이 논문에서 설명하는 earth(MARS) 응용 프로그램으로, 포트 7878 을 사용한다. earth() 가 중심 감정평가 엔진이다.
- **earthUI::launch_glmnet()** — 엘라스틱 넷 후처리 모듈 (포트 7879) 로, earth 모형의 기저 함수 (basis function) 를 가져와 라쏘/엘라스틱 넷 정규화 아래에서 재추정한다 (아래 “glmnetUI Import” 참조).
- **earthUI::launch_mgcv()** — GAM 후처리 모듈 (포트 7880) 로, earth 모형의 매듭점 (knot) 위치를 mgcv 평활항의 출발점으로 사용한다 (아래 “mgcvUI Auto-Export” 참조).

세 루틴은 하나의 프로젝트 트리, 하나의 설정 데이터베이스, 그리고 지원 기반 구조의 사본 하나를 공유하므로, 한 응용 프로그램에서 연 프로젝트는 곧바로 다른 응용 프로그램에서도 사용할 수 있다. glmnet 과 mgcv 루틴은 서로 경쟁하는 작업 흐름이 아니라 earth 모형의 가치 결론을 뒷받침하는 증거이다. 의도된 방식은 먼저 earth 모형을 적합시키고 다듬은 다음, 보강이 필요할 때 같은 프로젝트에 대해 모듈을 실행하는 것이다.

1.3 세 가지 용도 모드

earthUI 는 작업을 프로젝트 단위로 조직한다. 프로젝트란 입력 파일, 출력물, 설정을 고정된 위치에 함께 보관하는 이름 붙은 개체이다. 모든 프로젝트는 생성 시 새 프로젝트 마법사에서 선택하는 **Purpose**(용도) 를 지닌다. 세 가지 모드가 있다:

- **General**(일반) — 모든 유형의 모집단이나 데이터셋에 대한 Earth 회귀분석. 기본 모드이다. 도메인 특화 기능 없이 전체 MARS 모형화 작업 흐름을 제공한다.
- **For Appraisal**(감정평가용) — 부동산 감정평가에 맞춘 Earth 회귀분석. 대상 부동산 처리, 특수 열 지정, 잔차 제약 접근법 (RCA, Residual Constraint Approach), 거래사례비교표 생성 등 단일 부동산 감정평가에 특화된 기능을 추가한다.
- **Market Area Analysis**(시장지역 분석) — 시장지역 연구에 맞춘 Earth 회귀분석. 단일 대상 부동산을 평가하는 대신 정의된 시장 내 부동산 집단을 분석하는 기능을 추가한다.

세 모드 모두에서 핵심 모형화 엔진은 동일하다. 사용자는 언제나 Earth(MARS) 모형을 적합시킨다. 용도 설정은 어떤 추가 도구와 인터페이스 요소를 사용할 수 있는지를 제어할 뿐이다.

용도는 프로젝트별로 고정된다. 프로젝트의 용도는 국가, 주, 카운티, 도시와 함께 프로젝트 생성 시점에 잠긴다. 다른 프로젝트로 전환하면 earthUI 는 깨끗한 기본 상태로 초기화되고 (가져온 데이터, 모형 결과, 탭, 변수 구성, earth 매개변수가 모두 지워진다), 이어서 해당 프로젝트의 용도에 저장된 구성을 복원한다.

1.4 부동산 전용 기능

For Appraisal 또는 **Market Area Analysis** 를 선택하면 earthUI 는 부동산 분석을 위해 설계된 여러 기능을 활성화한다:

- **특수 열 지정** — 각 예측변수에는 contract_date, dom, concessions, latitude, longitude, living_area, lot_size, actual_age, effective_age, area, site_dimensions,

`display_only` 같은 특수 역할을 부여할 수 있다. 이 지정은 적합, 출력, 거래사례비교표 생성 과정에서 해당 열이 어떻게 다루어지는지를 제어한다. 전체 참조는 제 6 장을 보라.

- **위도와 경도의 반올림** — 위도 또는 경도로 지정된 열은 과적합을 막기 위해 자동으로 소수점 셋째 자리까지 반올림된다.
- **Sale Age 열** — 어떤 열이 `contract_date` 로 지정되고 기준일 (Effective Date) 이 주어지면, `earthUI` 는 `sale_age` 열 (매매일과 기준일 사이의 일수) 을 계산하여 예측변수로 대체한다.
- **RCA 계산 (감정평가 모드 전용)** — 감정평가 모드에서는 모형을 적합시킨 뒤 `earthUI` 가 잔차 제약 접근법 (RCA) 출력을 계산할 수 있으며, 이는 비교사례별 보정치, 순보정/총보정 요약, 그리고 대상 부동산의 보정 매매가격을 산출한다.
- **거래사례비교표 (감정평가 모드 전용)** — 비교사례 보정, 잔차 특성 분해, 보정 매매가격에 대한 수식을 포함한 다중 시트 Excel 통합문서를 생성한다. 제 11 장을 보라.

팁: *General* 용도 모드는 금융, 과학, 공학, 부동산 등 어떤 데이터셋에도 적용된다. 감정평가 모드와 시장 모드는 부동산 전문가를 위한 편의 기능을 더할 뿐이다.

1.5 earthUI 와 vProlog 설치

`earthUI` 는 CRAN 에 공개되어 있으므로 표준 설치인 R 에서 한 줄이면 된다:

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

이 논문은 현재 릴리스를 설명한다. 이 글을 읽는 시점에 CRAN 이 아직 이전 버전을 제공하고 있다면 개발 버전을 GitHub 에서 설치할 수 있다 (패키지가 저장소의 `pkg/` 폴더에 있으므로 `subdir` 에 유의하라):

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

vProlog(선택 사항). 실험적인 비고 처리 기능 (이 논문 끝부분의 “Prolog 텍스트 추출” 참조) 은 `vProlog` 패키지, 즉 자체 완결형 내장 Prolog 엔진을 사용한다. 이는 의도적으로 필수 의존성이 아니며, `earthUI` 는 `vProlog` 없이도 설치되고 실행되며 Prolog 기능만 `vProlog` 가 있을 때까지 비활성 상태로 남는다. `vProlog` 는 저자의 `r-universe` 저장소를 통해 배포되며, Windows 와 macOS 용 사전 빌드 바이너리를 제공한다 (컴파일러 도구 모음이 필요 없다):

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

시스템에 Prolog(SWI-Prolog 등) 를 별도로 설치할 필요는 없다. Trealla 엔진이 패키지 자체에 컴파일되어 들어 있다. 이 논문의 나머지 내용은 모두 `vProlog` 없이도 동작한다.

1.6 시작하기

`earthUI` 를 사용하려면:

1. 위에서 설명한 대로 CRAN 에서 패키지를 설치한다.
2. **응용 프로그램을 실행한다.** R 에서 `earthUI::launch()` 를 실행하거나, 명령줄에서 `Rscript -e 'earthUI::launch()'` 를 실행한다. 응용 프로그램은 웹 브라우저의 포트 7878 에서 열린다. 브라우저에서 `http://localhost:7878` 로 직접 접속할 수도 있다. (후처리 모듈도 같은 방식으로 시작한다: `launch_glmnet()` 은 포트 7879, `launch_mgcv()` 는 포트 7880 을 사용한다.)
3. 사이드바의 1 절에서 프로젝트를 만들거나 연다. 새 프로젝트 마법사는 프로젝트 이름, Purpose(General, For Appraisal, Market Area Analysis 중 하나), 위치 (국가, 주, 카운티, 도시), 그리고 선택적으로 최초 데이터 파일을 받아 프로젝트의 `in/` 폴더로 복사한다.

4. 2 절에서 데이터를 가져온다. 2 절에는 활성 프로젝트의 `in/` 폴더에 있는 CSV 및 Excel 파일이 나열된다. Finder 나 탐색기로 직접 추가한 파일은 Refresh 를 누르면 나타난다.
5. 변수를 구성한다 — 목표변수와 예측변수를 선택하고 자료형을 설정하며 특수 열 역할을 지정한다.
6. **Earth** 매개변수를 설정한다 — degree, nprune, 부분집합 필터를 비롯한 `earth()` 인수들.
7. 모형을 적합시킨다 — “Fit Earth Model” 을 클릭하고 주 패널에서 결과를 검토한다.
8. 내보낸다 — 예측값을 Excel 로 내려받거나, Quarto 보고서 묶음을 생성하거나, (감정평가 모드에서는) RCA 보정치와 거래사례비교표를 계산한다.

구성은 3 절과 4 절의 “Save current as default” 버튼을 통해 프로젝트와 용도를 키로 하여 프로젝트 데이터베이스에 서버 측으로 저장된다 (제 3 장 “설정 지속성” 참조).

2 MLS 입력 데이터 요건

부동산 감정평가 및 시장 분석 작업 흐름에서 입력 데이터는 대개 종합 부동산 정보 시스템 (MLS, Multiple Listing Service) 내보내기에서 나온다. 이 장에서는 기대되는 파일 구조와 earthUI 가 사용할 수 있는 열을 설명한다.

2.1 파일 형식과 구조

earthUI 는 CSV 와 Excel(.xlsx, .xls) 파일을 받는다. 가져오는 시점에 열 이름은 자동으로 snake_case 로 변환된다. 예를 들어 “Living SqFt” 는 living_sqft 로, “Contract Date” 는 contract_date 로, “Sale Price” 는 sale_price 로 바뀐다. 이 정규화 덕분에 작업 흐름 전체에서 열 참조가 일관되게 유지된다. 가져오기에 사용되는 CSV 구분자와 소수점 기호는 로컬 설정에 따라 결정된다 (제 3 장 “로컬 및 지역 설정” 참조).

데이터 파일은 부동산 한 건당 한 행, 속성 하나당 한 열로 이루어진 평면 표여야 한다. 파일의 첫 행에는 반드시 열 머리글이 있어야 한다.

2.2 감정평가 모드에 필요한 열

earthUI 는 어떤 열 구성에서도 동작하지만, 전체 감정평가 작업 흐름 (RCA 보정 + 거래사례비교표) 은 MLS 내보내기에 다음 열들이 있을 때 더 잘 작동한다:

열	특수 유형	용도
Sale Price	(목표변수)	모형의 반응변수
Contract Date	contract_date	sale_age(기준일로부터의 일수) 계산에 사용
Listing Date	listing_date	DOM 열이 없을 때 계약일과 함께 DOM 계산에 사용
Days on Market	dom	시장 노출 일수. 거래사례비교표에 표시
Concessions	concessions	매매 양허. 거래사례비교표에서 Net SP = Sale Price - Concessions
Living Area (SF)	living_area	제곱피트당 잔차 (residual_sf, cqa_sf) 사용 가능
Lot Size	lot_size	거래사례비교표의 “Site Size Dimensions” 행에 묶임
Site Dimensions	site_dimensions	거래사례비교표에서 부지 면적과 함께 묶임
Latitude	latitude	소수점 셋째 자리로 반올림. 근접성 계산과 Location 그룹에 사용
Longitude	longitude	소수점 셋째 자리로 반올림. 근접성 계산과 Location 그룹에 사용
Area ID	area	거래사례비교표의 “Loc: Long Lat Area” 행에 묶임
Actual Age	actual_age	거래사례비교표의 “Actual Age Effective Age” 행에 묶임
Effective Age	effective_age	거래사례비교표에서 실제 연령과 함께 묶임
Parcel Number	(없음)	거래사례비교표의 APN 행에 표시
Listing ID	(없음)	거래사례비교표의 APN 행에 표시
Address	display_only	거래사례비교표의 Address 행에 표시. 모형에서는 제외

스프레드시트의 열 이름은 외국어여도 무방하다. “특수” 이름이 영어인 것은 R 프로그램이 해당 열을 특별하게 처리할 수 있도록 하기 위함이다. 그 밖의 경우 주어진 열 이름이 회귀 모형, 그래프, 그리고 (감정평가를 수행한다면) 중간 거래사례비교표에 그대로 나타난다.

모든 열이 필수인 것은 아니다. earthUI 는 상황에 맞추어 동작한다. 어떤 열이 없으면 그에 대응하는 기능이 단순히 생략된다. 예를 들어 concessions 열이 지정되지 않았다면 거래사례비교표의 Net SP 행에는 양허 공제 없이 매매가격이 표시된다. 다만 부동산 가격 모형에서 받아들이기 만한 적합도를 얻으려면 다음 열들이 매우 권장된다:

1. **Sale Age** — 계약 매매일과 감정평가 또는 분석의 기준일 사이의 일수. 여러 해에 걸친 매매 이력을 사용할 때, 특히 5 년을 넘는 기간에 대해서는 `sale_age` 가 매매가격 추정에서 중심적 역할을 하는 경우가 많다. 실제로 이 변수가 워낙 중요해서 그것이 없으면 earthUI 가 아예 모형을 만들어 내지 못하는 경우도 흔하다.
2. **Living Area** — “Living Sqft”, “GLA”(총 주거 면적) 등 여러 이름으로 불린다. 매매가격을 결정하는 또 하나의 주요 요인이다.
3. **Total Bath Count** — 전체, 4 분의 1, 절반, 4 분의 3 욕실을 모두 합한 수. 예를 들어 전체 욕실 둘과 반욕실 하나는 2.5 가 된다.
4. **Garage Bays** 또는 **Garage Area** — 차고 주차 칸 수 또는 차고 면적.
5. **Lot Size** — 부동산의 토지 면적. 보통 제곱피트나 에이커 단위이다.
6. **Longitude, Latitude**, 그리고 가능하다면 **Area ID**. 이들에 대한 보정은 거래사례비교 표에서 하나의 Location 보정으로 통합된다.

2.3 특수 열 명명 규약

earthUI 는 열 이름이 아니라 특수 유형 지정으로 열을 식별한다. MLS 내보내기에서 열 이름은 원하는 대로 붙여도 되며, 중요한 것은 변수 구성 표 (제 6 장) 에서 올바른 특수 유형을 부여하는 것이다.

예를 들어 MLS 가 주거 면적을 “GLA”, “Living SqFt”, “liv_area”, “gross_living_area” 중 어느 이름으로 내보내더라도, 가져오기 (그 과정에서 snake_case 가 된다) 이후 Special 드롭다운에서 `living_area` 로 지정하기만 하면 된다. 그러면 earthUI 는 원래 이름과 무관하게 그 열을 제곱피트당 잔차 계산과 거래사례비교표 묶음에 사용한다.

2.4 데이터 품질과 완전성

- **결측값 (NA)**: 예측변수나 목표변수 열에 NA 가 있는 행은 적합 전에 자동으로 제거된다. 내부적으로 `na.action` 은 항상 `na.fail` 로 설정되며, NA 제거는 `earth()` 호출 이전에 처리된다.
- **날짜 열**: 날짜는 R 이 해석할 수 있는 형식이어야 한다 (예: 2025-06-15, 06/15/2025, June 15, 2025). earthUI 는 값의 80% 이상이 성공적으로 해석되면 날짜 열로 자동 감지한다. 두 자리 연도 (예: 06/15/25) 는 가장 가까운 연도로 읽히므로 최근 MLS 내보내기는 올바르게 해석된다. 네 자리 연도는 항상 문자 그대로 받아들여진다.
- **숫자 열**: 매매가격, 주거 면적, 부지 면적, 양허 등의 필드는 숫자여야 한다. MLS 가 통화 기호나 천 단위 구분자와 함께 가격을 내보낸다면 (예: “\$350,000” 또는 “350.000,00”), 가져오기 전에 정리해야 할 수 있다.
- **요인 열**: 지역 ID, 양식, 상태 같은 범주형 변수는 관리 가능한 수의 고유값을 가져야 한다. earthUI 는 요인을 자동 감지하지만 감지 결과를 사용자가 재정의할 수 있다.

팁: 가져오기 후 예측변수 표의 T/F/NA 열을 확인하라. 각 열의 참/거짓/NA 개수를 보여 준다. 결측값이 30% 를 넘는 열은 빨간색으로 표시된다. NA 가 많은 열은 제외하거나 가져오기 전에 데이터를 정리하는 것을 고려하라.

2.5 대상 부동산 행의 배치

Appraisal 모드에서는 1 행이 반드시 대상 부동산이어야 한다. 나머지 행은 모두 비교 매매사례이다. 대상 부동산의 매매가격은 비워 두거나 (NA) 어떤 값을 넣어도 되며, earthUI 는 적합 시 이를 무조건 NA 로 취급한다.

Market Area Analysis 모드에서는 1 행에 대상 부동산을 두는 것이 선택 사항이며, 어떤 행도 자동으로 제외되지 않는다. 대상 행 (또는 다른 어떤 행) 을 적합에서 제외하려면 지정된 가중치 열에 0 값을 주거나 부분집합 필터를 사용하라.

General 모드에서는 특별한 행 처리가 없으며 모든 행이 동등하게 취급된다.

3 General 용도 모드

3.1 개요

General Purpose 모드는 earthUI 를 실행했을 때의 기본 모드이다. 부동산에 국한되지 않고 어떤 데이터셋에 대해서도 완전한 MARS 모형화 작업 흐름을 제공한다. 과학 데이터, 재무 분석, 공학 연구를 비롯하여 변수 사이의 비선형 관계와 상호작용을 탐색하려는 어떤 회귀 문제에도 earthUI 를 쓸 수 있다.

General 모드에서는 인터페이스가 부동산 전용 기능 (특수 열, sale age, 좌표 반올림, RCA) 을 생략한다. 사이드바는 변수 선택, 매개변수 구성, 모형 적합, 내보내기에 집중하도록 간소화된다.

3.2 사이드바 작업 흐름

사이드바는 번호가 매겨진 접이식 절들로 조직되어 있어 프로젝트 선택부터 내보내기까지 안내한다:

1. Project — 기존 프로젝트를 고르거나 새 프로젝트 마법사로 새로 만든다. 마법사는 용도, 국가, 행정 위치를 받으며 이는 생성 후 모두 잠긴다. 활성 프로젝트가 입력 파일을 읽어 오는 위치와 모든 출력이 기록되는 위치를 결정하며, 별도의 출력 폴더 입력란은 없다.

2. Import Data — 활성 프로젝트의 in/ 폴더에서 CSV 또는 Excel 파일을 고른다. Refresh 링크는 직접 추가한 파일을 찾기 위해 폴더를 다시 훑는다. 시트가 여럿인 Excel 파일에는 시트 선택기가 나타난다. 열 이름은 자동으로 snake_case 로 변환된다.

3. Variable Configuration — 목표변수 선택기 (다중 목표변수 지원), 기준일 (Effective Date), 그리고 Include, Factor, Linear, Latent 확인란이 있는 예측변수 표. 자세한 내용은 제 6 장을 보라.

4. Earth Call Parameters — earth() 함수의 모든 인수: degree, penalty, nk, 가지치기 방법, 교차검증 등. 전체 매개변수 참조는 제 7 장을 보라.

Execute Prolog Processing — 4 절과 5 절 사이에 있는 선택적이고 번호가 없는 절로, 설정에서 실험적 Prolog 기능을 켜지 않으면 비활성 상태이다. 15장을 보라.

5. Fit Earth Model — 난수 시드 입력란과 모형을 비동기적으로 실행하는 버튼. 적합에 관한 자세한 내용은 제 8 장을 보라.

6. Download Estimated Target Variable(s) & Residuals — (감정평가 및 시장 모드에서는 “Download Estimated Sale Prices & Residuals” 라는 제목이 된다) 예측값, 잔차, CQA 점수, g 함수별 기여도를 Excel 파일로 내보낸다. 제 9 장을 보라.

7-8. RCA Adjustments and Sales Grid — 감정평가 모드 전용이다. 제 10 장과 제 11 장을 보라.

9. Generate Quarto Report — 자체 완결형 Quarto 보고서 묶음 (.qmd 원본과 자산) 을 프로젝트의 출력 폴더에 기록한다. 감정평가 모드가 아닐 때는 번호가 7 이다. 제 12 장을 보라.

10. Convert Quarto Report — 임의의 .qmd 파일을 HTML, Word, PDF 로 렌더링한다. 제 12 장을 보라.

3.3 주 패널 탭

적합 후 주 패널에는 열 개의 탭이 제공된다:

탭	내용
Data	가져온 데이터의 미리보기. 감정평가 모드에서는 대상 부동산 표와 비교 매매사례 표로 나뉜다.
Equation	적합된 모형 방정식을 LaTeX/MathJax 로 표시하며, 각 기저 함수와 그 계수를 보여 준다.
Summary	주요 지표 (R^2 , GRSq, GCV, RSS, CV R^2) 와 계수 표.
Variable Importance	예측변수 중요도 점수의 막대그래프와 순위 표.
Contribution	g 함수 기여도 그림. 드롭다운에서 g 함수를 선택한다. 일변량 예측변수는 선 그림으로, 이변량 상호작용 ($\text{degree} \geq 2$) 은 3 차원 투시도와 등고선 그림으로 표시된다.
Correlation	예측변수 상관관계 히트맵.
Diagnostics	세 가지 그림: 잔차 대 적합값, 정규 Q-Q, 실제값 대 예측값.
RCA Adjustments	(감정평가 전용) 비교사례 전반의 Residual Adj %, Net Adj %, Gross Adj % 히스토그램. 7 단계 (Calculate RCA Adjustments) 가 완료된 뒤에 채워진다.
Residual Contribution	모형이 사용하지 않은 열들에 대해 모형 잔차를 결합 회귀시켜, 드물거나 제외된 특성의 시장 기여도를 추정한다 (subsection 4.4 참조).
ANOVA	각 기저 함수의 RSS 기여도를 보여 주는 ANOVA 분해 표.
Earth Output	가지치기 단계 세부 정보를 포함한 'earth::summary()'의 원시 텍스트 출력.

3.4 설정 지속성

earthUI 는 프로젝트와 용도를 키로 하여 프로젝트 데이터베이스 (regProj 루트의 projects.sqlite) 에 구성을 서버 측으로 저장한다. 저장은 버튼 방식이다. 3 절의 “Save current as default” 버튼은 목표변수, 기준일, 예측변수 구성을 저장하고, 4 절의 버튼은 earth 호출 매개변수와 Allowed Interactions 행렬을 저장한다. 각 버튼은 자기 몫만 갱신하므로 서로 덮어쓰지 않으며, **적합은 자동 저장하지 않으므로** 저장된 기본값을 덮어쓸 걱정 없이 자유롭게 실험할 수 있다. 한 프로젝트의 여러 입력 파일 (작은 시험 추출본, 전체 데이터셋) 은 하나의 구성을 공유하고, 설정이 프로젝트 트리 안에 있으므로 regProj 폴더를 동기화하거나 백업하거나 공유할 때 설정도 함께 따라간다.

3.5 다크 모드

오른쪽 위 모서리의 달/해 아이콘을 클릭하면 밝은 테마와 어두운 테마가 전환된다. 테마 설정은 로컬 저장소에 보관되어 세션이 바뀌어도 유지된다.

3.6 로캘 및 지역 설정

earthUI 는 국가 기반 로캘 체계를 통해 국제적인 숫자, 날짜, CSV 형식 관행을 지원한다. 상단 메뉴 막대의 **Settings**(톱니바퀴) 메뉴에 있는 **Country** 드롭다운으로 지원되는 31 개 국가의 사전 설정을 선택한다. 각 사전 설정은 다음을 구성한다:

- **CSV 구분자** — 미국/영국/일본은 쉼표 (,), 쉼표를 소수점 기호로 쓰는 유럽 대부분 지역은 세미콜론 (;).
- **소수점 기호** — 마침표 (.) 또는 쉼표 (,).
- **천 단위 구분자** — 쉼표 (미국/영국/일본), 마침표 (독일/이탈리아/스페인), 공백 (핀란드/프랑스/폴란드/발트 3 국/우크라이나/러시아), 또는 아포스트로피 (스위스).
- **날짜 형식** — MM/DD/YYYY(미국), DD/MM/YYYY(유럽 대부분), YYYY-MM-DD(스웨덴/리투아니아/일본/캐나다).
- **용지 크기** — Letter(미국/캐나다/멕시코) 또는 A4(그 외 모든 지역).

3.6.1 지원 국가

국가	CSV 구분자	소수점	천 단위	날짜	용지	
미국, 캐나다 (csv)	,	.	,	MDY/YMD	Letter	
영국, 아일랜드, 호주, 뉴질랜드	,	.	,	DMY	A4	
독일, 오스트리아, 벨기에, 네덜란드, 이탈리아, 스페인, 포르투갈, 덴마크, 브라질	;	,	.	DMY	A4	
프랑스, 폴란드, 체코, 노르웨이, 라트비아, 에스토니아	;	,	(공백)	DMY	A4	
핀란드, 스웨덴	;	,	(공백)	DMY/YMD	A4	
스위스	;	,	'	DMY	A4	
우크라이나, 러시아	;	,	(공백)	DMY	A4	
튀르키예	;	,	.	DMY	A4	
일본, 대한민국	,	.	,	YMD	A4	
멕시코	,	.	,	DMY	Letter	
리투아니아	;	,	(공백)	YMD	A4	

3.6.2 용지 재정의

국가 선택기 아래의 **Paper** 드롭다운 (Letter 또는 A4) 을 쓰면 국가를 바꾸지 않고도 보고서 페이지 크기를 재정의할 수 있다. CSV 구분자, 소수점 기호, 날짜 형식 순서는 선택한 국가의 사전 설정을 따르며, 국가를 바꾸면 용지 재정의는 그 국가의 기본값으로 초기화된다.

3.6.3 기본값 저장

Settings 메뉴에서 **Save** 를 클릭하면 로컬 설정이 전역으로 저장된다. 이 기본값은 어떤 프로젝트를 열든 이후 모든 세션에 적용된다. 프로젝트별 설정 (목표변수, 예측변수, 매개변수) 은 프로젝트 데이터베이스에 따로 저장되며, 로컬 기본값은 사용자 수준 SQLite 데이터베이스를 통해 모든 프로젝트에 걸쳐 유지된다. 이 두 층 구조는 여러 나라의 데이터를 다루는 감사 법인 같은 조직을 염두에 둔 것이다. 가장 자주 쓰는 국가를 기본값으로 정해 두고 필요할 때 조정하면 된다.

팁: 그림 축과 기울기 레이블의 숫자 형식은 로컬에 맞추어 자동으로 적응한다. 독일 로컬은 천 단위에 마침표를 쓰고 (200.000), 핀란드는 공백을 (200 000), 스위스는 아포스트로피를 (200'000) 쓴다. 통화 기호는 표시되지 않는다. earthUI 는 통화 중립적이다.

4 Appraisal 모드

Purpose 로 **For Appraisal** 을 선택하면 earthUI 는 단일 부동산 감정평가에 맞게 스스로를 구성한다. 제 3 장에서 설명한 모든 기능은 그대로 사용할 수 있으며, 이 장에서는 감정평가 전용 추가 사항만 다룬다.

4.1 대상 부동산 행의 처리

감정평가 모드에서는 데이터셋의 1 행이 대상 부동산이고 나머지 행은 모두 비교 매매사례이다. 입력 파일을 그에 맞게 구성해야 한다 (제 2 장 참조). 대상 부동산의 매매가격은 비워 두거나 어떤 값을 넣어도 되며, earthUI 는 적합 시 자동으로 NA 로 취급한다.

가져오기 후 Data 탭은 두 부분, 즉 **Subject Property**(1 행) 와 **Comparable Sales**(2 행 이후) 로 나뉜다. 1 행은 항상 모형 적합에서 제외되며, “Skipping row 1 (subject). Fitting on N rows.” 라는 알림이 이를 확인해 준다. 적합 후에도 모형은 대상 행에 대한 예측값을 생성하며, 출력에서 `est_<target>` 로 표시된다.

4.2 기준일과 Sale Age

Effective Date(기준일) 입력란은 모든 용도 모드의 변수 구성 절에 나타나며 (기본값은 오늘 날짜) — 감정평가뿐 아니라 일반 분석과 시장 분석에도 기준일이 있다 — 실제로 사용되는 것은 감정평가 모드와 시장 모드뿐이다. 이들 모드에서 Special 열 드롭다운으로 어떤 열을 `contract_date` 로 지정하면, earthUI 는 `sale_age` 열, 즉 각 매매의 계약일과 기준일 사이의 정수 일수를 계산한다. 이 열이 원래의 날짜 열을 예측변수로 대체한다.

계약일을 지정한 뒤 처음 Fit 을 클릭하면 earthUI 가 `sale_age` 열을 만들고, 이를 포함시키려면 Fit 을 한 번 더 클릭하라고 알려 준다.

4.3 특수 열 지정

감정평가 모드와 시장 모드에서는 변수 구성 표의 각 예측변수마다 **Special** 드롭다운이 나타난다. 특수 유형의 전체 목록과 그 효과는 다음과 같다:

특수 유형	효과
no	기본값 — 특별한 처리 없음
contract_date	기준일로부터 sale_age 를 자동 계산하도록 함
listing_date	dom 열이 없을 때 계약일과 함께 DOM 계산에 사용
dom	시장 노출 일수 열을 식별. 거래사례비교표에 표시
concessions	매매 양허를 식별. 거래사례비교표의 Net SP 수식 (Sale Price – Concessions) 에 사용
latitude	값을 소수점 셋째 자리로 반올림. 하버사인 근접성과 거래사례비교표의 Location 그룹에 사용
longitude	값을 소수점 셋째 자리로 반올림. 하버사인 근접성과 거래사례비교표의 Location 그룹에 사용
area	거래사례비교표의 "Loc: Long Lat Area" 행에 묶임
living_area	내려받기 출력에서 제곱피트당 잔차 (residual_sf, cqa_sf) 를 사용 가능하게 함
lot_size	거래사례비교표의 "Site Size Dimensions" 행에 묶임
site_dimensions	거래사례비교표에서 부지 면적과 함께 묶임
actual_age	거래사례비교표의 "Actual Age Effective Age" 행에 묶임
effective_age	거래사례비교표에서 실제 연령과 함께 묶임
display_only	내보내기에는 포함되지만 모형 적합에서는 제외되는 열 (여럿 지정 가능)
remarks	(실험적) 선택적 Prolog 단계가 특정 열로 구문 분석하는 자유 텍스트 비고 열 (여럿 지정 가능). 15장 참조
list	(실험적) 선택적 Prolog 단계가 원-핫 TRUE/FALSE 열로 분할하는 섹션 구분 열 (여럿 지정 가능). 15장 참조

특수 유형당 열은 하나만 허용된다 (display_only, remarks, list 는 예외). 어떤 특수 유형을 두 번째 열에 부여하면 첫 번째 열에서 자동으로 해제된다. 변수 이름 옆에는 부여된 특수 유형을 보여 주는 작은 파란색 배치가 나타난다.

4.4 잔차 기여도 (회귀 특성 추정)

회귀 경로에 담기에는 너무 드문 특성 — 몇 건의 매매에만 나타나는 작업실 같은 것 — 은 RCA 잔차 층에 속한다. **Residual Contribution** 탭은 모형이 전혀 사용하지 않은 열들에 대해 적합한 모형의 잔차를 보통최소제곱으로 회귀시켜 그 시장 기여도를 추정한다. 여러 열을 선택하면 결합하여 추정하므로, 서로 상관된 편의 시설 (예: outbuilding_sqft 와 pr_has_workshop) 이 중복 계산되지 않고 기여를 나누어 갖는다. 이진 플래그는 총액 추정치를 내고 플래그가 붙은 각 매매의 잔차를 증거로 열거하며, 숫자 열은 단위당 기여도를 낸다. 보고되는 표준오차는 1 표준오차이다 (통상적인 유의성 기준으로는 $t \geq 2$). 관측치가 열 개 미만인 항목에는 생성되는 작업 파일 메모에 자동으로 CAUTION 이 붙는데, 이 메모는 방법, 추정치, 유의 사항을 한 문단으로 정리해 보고서에 쓸 수 있게 한다. 후보 목록에서 제외되는 것은 모형이 실제로 사용한 예측변수뿐이다. 적합에 제공되었으나 전진 단계에서 선택되지 않은 회귀 플래그는 여전히 후보로 남는데, 모형이 그 가치를 전혀 흡수하지 않았기 때문이다.

4.5 RCA 보정 개요

Calculate RCA Adjustments & Download 버튼 (사이드바 7 절, 감정평가 모드에서 적합 후에만 표시) 은 대상 부동산 대비 각 비교사례의 시장 도출 보정치를 계산한다. 전체 RCA 작업 흐름은 제 10 장에서 설명한다.

RCA 보정치를 계산한 뒤에는 **Generate Sales Grid & Download** 버튼 (사이드바 8 절) 을 사용할 수 있게 된다. 거래사례비교표 작업 흐름은 제 11 장에서 설명한다.

답: 대상 부동산에 부여하는 *CQA* 점수는 잔차 분포 가운데 얼마만큼을 대상 부동산에 귀속시킬지를 결정한다. 5.00 점은 대상 부동산을 비교사례의 중앙값에 놓는다.

5 Market Area Analysis 모드

Purpose 로 **Market Area Analysis** 를 선택하면 earthUI 는 감정평가 모드와 동일한 부동산 전용 기능 (특수 열, sale age, 좌표 반올림) 을 제공하되, 단일 대상 부동산을 평가하기보다 부동산 집단을 분석하는 쪽으로 방향을 맞춘다.

5.1 Appraisal 모드와의 차이

- 대상 부동산 자동 제외 없음 — 감정평가 모드와 달리 시장 모드는 1 행을 적합에서 자동으로 제외하지 않는다. 모든 행이 포함된다. 대상 행을 적합에서 빼려면 지정된 가중치 열에 0 값을 주거나 (데모 데이터셋처럼) 부분집합 필터를 사용하라.
- **RCA** 및 거래사례비교표 절 없음 — “Calculate RCA Adjustments & Download” 와 “Generate Sales Grid & Download” 단계는 제공되지 않는다. 시장 모드는 비교사례별 보정이 아니라 모형 적합과 출력에 집중한다.
- **사이드바 번호** — RCA 와 거래사례비교표 단계가 없으므로 Generate Quarto Report 절의 번호가 9 가 아니라 7 이 된다 (Convert Quarto Report 절은 그대로 10 이다).

5.2 시장 모드를 쓸 때

Market Area Analysis 모드는 다음과 같은 경우에 적합하다:

- 가치 결정 요인을 파악하기 위해 근린지역이나 시장지역의 회귀 모형을 구축할 때
- 면적, 연령, 부지 면적, 위치 같은 변수가 부동산 집단 전반의 매매가격에 어떤 영향을 주는지 분석할 때
- 시장 상황 분석이나 근린지역 획정을 뒷받침할 자료를 준비할 때
- 1 행에 대상 부동산이 들어 있는 데이터셋을 다루면서 그것을 적합에 포함할지 제외할지 선택하고 싶을 때

팁: 시장 모드는 특수 열 기능 (sale age, 좌표 반올림) 은 쓰고 싶지만 RCA 보정 작업 흐름은 필요 없는 일반적인 부동산 회귀분석에도 유용하다.

6 변수 선택

사이드바 3 절 — **Variable Configuration** — 은 어떤 열이 모형에 참여하고 그 열들이 어떻게 다루어지는지를 정하는 곳이다.

6.1 목표 (반응) 변수

3 절 상단의 **Target (response) variable(s)** 드롭다운에는 데이터셋의 모든 열이 나열된다. 표준적인 단일 반응 모형에는 열 하나를, 다중 반응 모형에는 여러 열을 선택한다. 목표변수를 여럿 선택하면 모형은 `earth(cbind(y1, y2, ...) ~ .)` 를 사용하여 모든 반응변수를 동시에 적합시킨다.

목표변수로 선택된 열은 예측변수 목록에서 자동으로 제외된다.

6.2 예측변수 표

목표변수 선택기 아래에는 남은 모든 열을 다음 필드와 함께 나열하는 표가 있다:

열	설명
Variable	열 이름 (잘린 경우 전체 이름이 툴팁에 표시) 과 그 앞의 $\times/+$ 토글 (아래 참조). 감정평가/시장 모드에서는 파란색 배지가 부여된 특수 유형을 보여 준다.
Type	자료형 드롭다운: <code>numeric, integer, character, logical, factor, Date, POSIXct</code>
Include	확인란 — 이 열을 모형의 예측변수로 포함한다
Factor	확인란 — 이 열을 범주형 변수로 취급한다
Linear	확인란 — 선형 향으로만 들어가도록 강제한다 (힌지 함수 없음)
Latent	확인란 — 이 열을 모형에서 완전히 배제한다 (<code>earth</code> 와 <code>glmnetUI</code> / <code>mgcvUI</code> 로 내보내는 데이터 모두에서 제외), 잠재변수 작업을 위해 남겨 둔다
Special	드롭다운 (감정평가/시장 전용) — 아래 특수 열 유형 참조
T/F/NA	열마다 세 개의 개수: 논리형은 TRUE/FALSE/NA, 숫자형은 $\neq 0 = 0/NA$. 결측값이 30% 를 넘으면 빨간색으로 표시된다

표 위의 안내 줄은 약어를 설명한다: “Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`).”

삭제 대신 소프트 비활성화. 각 변수 이름 옆의 $\times$ 토글은 해당 행을 회색으로 만들고 그 열을 데이터에서 지우지 않은 채 모형에서 빼낸다. 토글은 초록색 $+$ 로 바뀌어 열을 다시 활성화한다. 열이 파괴되는 일은 결코 없다.

6.3 자료형 감지와 재정의

`earthUI` 는 가져오기 시점에 자료형을 자동으로 감지한다. 숫자, 정수, 논리, 요인, 날짜 열이 인식된다. 날짜처럼 보이는 문자 열 (값의 80% 이상이 일반 날짜 형식으로 해석되는 경우) 은 `Date` 로 분류된다.

Type 드롭다운을 바꾸어 어떤 감지 결과든 재정의할 수 있다. 열을 `character` 나 `factor` 로 바꾸면 Factor 확인란이 자동으로 선택된다. 자료형을 바꾸면 그 열이 `earth()` 함수에 전달되는 방식이 달라진다.

6.4 Factor 및 Linear 플래그

- **Factor** — 선택하면 해당 열이 범주형 변수로 취급된다. 연속적인 측정값이 아니라 이산적인 집단을 나타내는 `style`, `area_id`, `grade` 같은 열에 적절하다. 요인 열은 지시 (더미) 변수로 모형에 들어간다.
- **Linear** — 선택하면 해당 열이 선형으로만 모형에 들어가며, 힌지 (구간별 선형) 함수가만 들어가지 않는다. 어떤 변수가 목표변수와 엄밀히 선형 관계에 있음을 아는 경우에 유용하다.

6.5 특수 열 유형 참조

감정평가 모드와 시장 모드에서 **Special** 드롭다운은 다음 선택지를 제공한다. 각 유형은 최대 한 열에만 부여할 수 있다 (`display_only`, `remarks`, `list` 는 여럿 허용):

날짜 및 시간 유형:

- `contract_date` — `sale_age` 자동 계산을 촉발한다. 원래의 날짜 열은 매매일과 기준일 사이의 일수를 나타내는 정수 열로 대체된다.
- `listing_date` — 명시적인 `dom` 열이 지정되지 않은 경우 시장 노출 일수 (`DOM` = 계약일 - 등록일) 를 계산하는 대체 수단으로 사용된다.
- `dom` — 시장 노출 일수 열을 식별한다. 거래사례비교표의 `APN` 행과 `Date of Sale` 행에 표시된다.

금액 유형:

- `concessions` — 매매 양허 (매도인 크레딧, 매수인 유인책 등) 를 식별한다. 거래사례비교표에서 순매매가격을 계산하는 데 쓰인다: $\text{Net SP} = \text{Sale Price} - \text{Concessions}$.

규모 및 위치 유형:

- `latitude` — 과적합을 막기 위해 값이 자동으로 소수점 셋째 자리까지 반올림된다. 하버사인 근접성 계산 (대상 부동산에서 각 비교사례까지의 거리) 에 사용되며 거래사례비교표의 `Location` 행에 묶인다.
- `longitude` — 위도와 동일한 반올림 처리를 받는다. 근접성 계산과 `Location` 그룹에 사용된다.
- `area` — 대개 근린지역 또는 지역 식별자이다. 거래사례비교표의 “`Loc: Long | Lat | Area`” 행에서 위도, 경도와 함께 묶인다.
- `living_area` — 내려받기 출력에서 제곱피트당 잔차 계산 (`residual_sf`, `cqa_sf`) 을 가능하게 한다.
- `lot_size` — 거래사례비교표의 “`Site Size | Dimensions`” 행에 묶인다.
- `site_dimensions` — 거래사례비교표에서 부지 면적과 함께 묶인다 (예: “75x120”).

연령 유형:

- `actual_age` — 거래사례비교표의 “`Actual Age | Effective Age`” 행에 묶인다.
- `effective_age` — 거래사례비교표에서 실제 연령과 함께 묶인다.

표시 유형:

- `display_only` — 해당 열은 Excel 내보내기에는 포함되지만 모형 적합에서는 완전히 제외된다. 주소 필드, MLS 번호, 필지 ID 등 예측변수가 되어서는 안 되는 참조 데이터에 사용하라. 여러 열에 이 지정을 부여할 수 있다.

텍스트 추출 유형 (실험적):

- `remarks` — 선택적 Prolog 단계가 특성 열로 구문 분석할 자유 텍스트 열 (예: `public_remarks`). 해당 기능을 설정에서 켜기 전까지는 아무 일도 하지 않는다. 여러 열을 지정할 수 있다. 15장을 보라.
- `list` — 같은 단계가 원-핫 `<col>_<value>` TRUE/FALSE 열로 분할할 심표 구분 열 (예: `existing_structures`). 여러 열을 지정할 수 있다. 15장을 보라.

6.6 다중 목표변수

목표변수를 둘 이상 선택하면 다중 반응 Earth 모형이 적합된다. 여러 목표변수를 선택하면:

- 모형이 모든 반응변수를 동시에 예측하며 목표변수들 사이에서 기저 함수를 공유한다
- **wp (response weights)** 버튼이 활성화되어 각 목표변수에 숫자 가중치를 부여할 수 있다
- 분산 모형 (`varmod.method`) 은 비활성화된다 (다중 반응에서는 지원되지 않음)
- 결과 탭은 반응변수별 지표, 방정식, 진단 그림을 표시한다

팁: `display_only` 로 지정된 열은 데이터셋에 남아 *Excel* 내보내기에 나타나지만 예측변수 표에서는 완전히 제외된다. *ID* 열, 주소, 기타 참조 필드에 사용하라.

7 매개변수 선택

사이드바 4 절 — **Earth Call Parameters** — 은 `earth()` 함수가 받는 모든 인수에 접근할 수 있게 한다. 각 매개변수에는 툴팁 설명이 딸린 파란색 도움말 아이콘 (?) 이 있다. 매개변수는 접이식 하위 절로 조직되어 있다.

7.1 기본 매개변수

매개변수	기본값	설명
<code>subset</code>	(비어 있음)	행 필터 표현식. 아래 “부분집합 필터링” 참조.
<code>weights</code>	NULL	사례 (행) 가중치를 담은 열 선택기. 숫자 열만 나열된다.
<code>wp</code>	NULL	다중 목표 모형의 반응 가중치. 버튼을 누르면 목표변수마다 숫자 입력란이 있는 대화 상자가 열린다 (기본 값은 각각 1.0).
<code>keepxy</code>	꺼짐	x, y, subset, weights 를 모형 객체에 보존한다.
<code>trace</code>	0	적합 출력의 추적 수준 (0 에서 5 까지).
<code>glm</code>	없음	선택적 GLM 분포족: <code>gaussian</code> , <code>binomial</code> , <code>poisson</code> .
<code>degree</code>	1	최대 상호작용 차수. degree 를 ≥ 2 로 설정하면 교차검증이 자동으로 켜지고 Allowed Interactions 행렬이 나타난다.
<code>penalty</code>	2	매듭점당 GCV 벌점. 값이 클수록 더 단순한 모형이 만들어진다.

7.2 전진 단계

매개변수	기본값	설명
<code>nk</code>	자동	가지치기 전 최대 항 수. 적재된 데이터에 대한 권장 값으로 자동 입력된다 (아래 “권장 매개변수 값” 참조).
<code>thresh</code>	0.001	전진 단계 임계값. 값이 작을수록 더 많은 항이 허용된다.
<code>minspan</code>	자동	매듭점 사이의 최소 간격. 권장 값으로 자동 입력된다. 음수 값은 예측변수당 최대 매듭점 수를 정한다.
<code>endspan</code>	자동	끝 간격 — 매듭점에서 데이터 경계까지의 최소 거리. 권장 값으로 자동 입력된다.
<code>newvar.penalty</code>	0	새 변수를 도입하는 데 대한 벌점 (기존 예측변수의 재사용을 유도한다).

매개변수	기본값	설명
fast.k	20	빠른 MARS 알고리즘에서 고려할 부모 항의 수.
fast.beta	1	빠른 MARS 의 노화 계수를 제어한다.

7.3 허용 상호작용

degree 가 2 이상으로 설정되면 기본 매개변수 아래에 상호작용 행렬이 나타난다. 이는 예측변수 쌍마다 하나씩의 확인란으로 이루어진 상삼각 격자이다. 확인란이 선택되어 있으면 두 예측변수의 상호작용이 허용되고, 해제하면 그 특정 상호작용이 금지된다.

예측변수 이름 (행 또는 열 레이블) 을 클릭하면 그 변수의 모든 상호작용이 한꺼번에 전환된다. 상단의 **Allow All** 과 **Clear All** 확인란은 일괄 제어를 제공한다. 이 행렬은 고정 머리글을 사용하므로 스크롤해도 행과 열 레이블이 계속 보인다.

정보 알림은 다음을 상기시킨다: “Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).”

7.4 가지치기

매개변수	기본값	설명
pmethod	backward	가지치기 방법: backward , none , exhaustive , forward , seqrep , cv .
nprune	NULL	가지치기 후 최대 항 수. 자동 선택을 원하면 비워 두라.

7.5 교차검증

매개변수	기본값	설명
nfold	10	교차검증 분할 (fold) 수. 0 으로 설정하면 교차검증이 꺼진다.
ncross	20	교차검증 반복 횟수.
stratify	켜짐	각 분할이 비슷한 반응 분포를 갖도록 교차검증 표본을 층화한다.

degree 가 ≥ 2 일 때는 상호작용으로 인한 과적합을 막는 데 도움이 되도록 교차검증이 자동으로 켜진다 (nfold 가 10 으로 설정된다).

7.6 부분집합 필터링

subset 텍스트 입력란은 모형 적합에 사용할 행을 걸러 내는 R 표현식을 받는다. 표현식을 직접 입력하거나 (예: `sale_age < 365 & area_id == 460`) **Build filter...** 버튼으로 시각적으로 구성할 수 있다.

Build Filter 대화 상자 — “Build filter...” 를 클릭하면 안내식 대화 상자가 열린다. 각 조건 행에는 열 드롭다운, 연산자 (<, >, <=, >=, ==, !=), 그리고 열 유형에 맞추어 바뀌는 값 입력란이 있다. 숫자에는 숫자 입력란, 날짜에는 날짜 선택기, 문자/요인 열에는 고유값 드롭다운이 제공된다. 조건은 AND(&) 또는 OR(|) 연결자로 결합된다. 하단의 미리보기가 표현식과 일치하는 행의 수를 보여 준다. **Apply** 를 클릭하면 표현식이 텍스트 입력란에 삽입된다.

날짜 열은 수동 표현식에서 `as.Date("...")` 또는 `as.POSIXct("...")` 래퍼를 사용해야 한다. Build Filter 대화 상자는 이를 자동으로 처리한다.

부분집합 필터링은 비파괴적이다. 제외된 행도 데이터셋에 남아 있으며 Excel 내보내기에서 예측값을 받는다.

팁: OR 조건을 AND 와 결합할 때는 괄호로 묶으라. 괄호가 없으면 $A \& B / C$ 는 $(A \& B) / C$ 로 평가되며, 이는 의도한 바가 아닐 수 있다.

7.7 권장 매개변수 값

earthUI 는 4 절의 주요 매개변수 아래에 권장 값을 표시한다. 이 권장값은 적합에 쓰이는 행의 수 (n) 와 선택된 예측변수의 수 (p) 에 따라 반응적으로 갱신된다. 공식은 Friedman 의 MARS 논문, earth 의 내부 알고리즘, 그리고 경험적 시험에서 도출되었다. **nk**, **minspan**, **endspan** 입력란은 데이터가 적재될 때 권장 값으로 미리 채워지며, 수동으로 편집한 값은 그대로 보존된다.

7.7.1 전진 단계 매개변수

nk(가지치기 전 최대 항 수):

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

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 의 기본값인 $\min(200, \max(20, 2p)) + 1$ 은 데이터셋 크기를 고려하지 않아 큰 데이터셋에서는 지나치게 작을 수 있고, 전진 단계가 모든 예측변수를 충분히 탐색하기 전에 제약을 받게 된다. $\lfloor n/10 \rfloor$ 항은 관측치 10 개당 대략 한 항을 허용하는데, 이는 과다 모수화를 피하기 위한 표준적인 경험 법칙이다. 상한 100 은 수확 체감을 막는다.

minspan(매듭점 사이의 최소 관측치 수):

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

endspan(데이터 경계로부터의 최소 관측치 수):

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

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

n	minspan	endspan
1500	16	16

earth 가 자동 계산하는 minspan 은 Friedman 의 식 43 을 사용한다: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$. 이는 $\ln(np)$ 에 비례하여 커지므로 큰 데이터셋에서는 지나치게 느리게 증가한다. $n=1500$ 에서는 약 ~ 7 밖에 되지 않아 연속 예측변수마다 대략 245 개의 후보 매듭점 위치를 허용하는데, 이는 너무 많아 전진 단계가 잡음을 적당하게 만든다. 권장 공식은 큰 n 에서 예측변수당 후보를 약 100 개로 맞추면서, 작은 n 에서는 (그 영역은 이미 충분히 검증되었으므로) earth 의 자동 계산을 따른다. endspan 은 minspan 보다 약간 크게 설정하여 경계 보호를 추가하는데, 이는 데이터의 꼬리가 얇을 때 (부동산에서는 흔한 일이다) 도움이 된다.

penalty(매듭점당 GCV 벌점):

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

이는 earth 자체의 기본값이다. 가법 모형에는 2, 상호작용 모형에는 3(더 큰 탐색 공간을 상쇄하기 위해 벌점을 높인다).

newvar.penalty(새 예측변수 도입에 대한 벌점):

권장값: 예측변수들이 서로 상관되어 있을 때 (예: 주거 면적과 침실 수, 욕실 수, 부지 면적) 0.1. 이는 전진 단계가 상관된 대체 변수를 새로 들이기보다 이미 모형에 있는 예측변수를 재사용하도록 편향시킨다. 작동 방식은 이렇다. 새 예측변수의 RSS 개선분에 $1/(1 + \text{penalty})$ 가 곱해진다. **newvar.penalty** = 0.1 이면 새 변수는 기존 변수에 매듭점을 하나 더 두는 것보다 RSS 를 최소 10% 더 개선해야 한다. 그 결과 최종 계수 추정치에는 영향을 주지 않으면서 (선택 이후 벌점은 제거된다) 예측변수가 더 적은 단순한 모형이 만들어진다. 예측변수들이 상관되어 있지 않다면 0 으로 두라. earth 는 0 이 아닌 **newvar.penalty** 와 사례 가중치를 함께 지원하지 않으므로, 가중치 열이 지정되면 earthUI 가 이 입력을 비활성화한다 (0 으로 강제).

7.7.2 가지치기 매개변수

pmethod: 권장값은 **backward** 이다. 기본값인 GCV 기반 후진 가지치기는 결정론적이어서 같은 데이터가 언제나 같은 모형을 산출하며 난수 시드에 의존하지 않는다. 이는 재현성, 특히 감정 평가 업무에서 중요하다. **cv** 방법은 가지치기에 교차검증을 사용하므로 시드 의존성이 생긴다.

nprune: 권장값은 비워 두는 것 (NULL) 이다. GCV 가 최적 항 수를 선택하게 하라. **nprune** 을 설정하면 GCV 의 판단을 무시하는 엄격한 상한이 부과된다.

7.7.3 교차검증 매개변수

nfold(교차검증 분할 수): 표본 크기 구간별 권장값:

n (적합 행 수)	nfold
0-299	5
300-399	6
400-499	7
500-599	8
600-699	9
700+	10

분할이 많을수록 각 분할이 더 많은 데이터로 학습되어 (편향이 줄어든 추정치) 좋지만, 대신 검정용으로 남겨 두는 분할이 작아진다. 위 표는 표본이 커짐에 따라 표준적인 10 분할까지 단계적으로 올라가되 각 검정 분할이 안정적인 지표를 내기에 충분한 크기 (대략 50 행 이상) 를 유지하도록 하며, 작은 표본에서는 5 분할로 되돌아간다. 작은 표본에서는 CVR^2 추정치가 여전히 잡

음이 많으므로 `ncross` 를 높여 (예: 15–20) 분산을 평균해 내라. 상한을 10 으로 둔 것은 분할을 하나 더할 때마다 `earth` 적합이 통째로 한 번 더 필요한 데 비해 진단상의 이득은 무시할 만하기 때문이다. `pmethod = "backward"`에서는 교차검증이 모형 자체에는 영향을 주지 않고, 진단용 CVR^2 를 계산하고 분산 모형에 쓸 잔차를 제공할 뿐이다.

`ncross`(교차검증 반복 횟수): 표본 크기 구간별 권장값:

n (적합 행 수)	<code>ncross</code>
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

반복은 분할 구성의 운을 평균해 내어 CVR^2 를 안정시킨다. 표본이 클수록 단 한 번의 교차검증만으로도 이미 더 안정적이므로 반복이 덜 필요하다. 반복된 교차검증의 잔차는 분산 모형에도 투입된다.

7.7.4 분산 모형

`varmod.method`: 권장값은 `lm` 으로, Milborrow 의 기본값이다. 예측구간을 위한 강건한 선형 분산 모형이다 (산포는 잡음이 많은 잔차로부터 추정되므로 단순한 모형이 더 잘 일반화된다). 데이터에서 `lm` 이 수렴하지 않으면 `rlm` 이나 `const`(등분산, 항상 수렴) 를 시도하거나 `earth` 로 되돌아가라 (비선형 산포도 다룬다). 이 기능에는 `nfold > 1` 과 `ncross > 1` 이 필요하다.

7.7.5 Earth 출력 파일

적합에 성공할 때마다 `earthUI` 는 출력 폴더에 `<filename>_earth_output_<timestamp>.txt` 를 기록한다. 이 파일에는 모형 항, 요약 통계량 (R^2 , $GRSq$, $CVR Sq$), 분산 모형 세부 정보, 추적 로그가 들어 있다. 적합마다 파일이 하나씩 만들어져 매개변수 구성을 비교할 수 있는 누적 기록이 된다.

7.8 설정 기본값

4 절 상단의 라디오 버튼이 어떤 기본값을 적재할지 제어한다:

- **Use last settings for input file** — 이 특정 파일에 대해 지난번에 사용한 설정을 복원한다
- **Use default settings** — 저장해 둔 사용자 지정 기본값을 적용한다
- **Earth defaults** — 모든 매개변수를 공장 초기값으로 되돌린다

4 절 하단의 **Save current as default** 를 클릭하면 현재 `earth` 매개변수와 Allowed Interactions 행렬이 저장된다. 3 절의 대응 버튼은 목표변수, 기준일, 예측변수 구성을 저장한다. 둘 다 프로젝트 데이터베이스에 프로젝트별, 용도별로 저장된다.

8 Earth 모형 적합

8.1 Fit 버튼

사이드바 5 절에는 **Fit Earth Model** 버튼과 **Random seed** 입력란이 있다. 시드는 교차검증 분할 배정을 재현 가능하게 만든다. 드롭다운에서 최근에 사용한 시드 열 개를 불러올 수 있고, 사용된 시드는 적합 후 Earth Output 탭에 기록된다. 적합 전에 earthUI 는 구성을 검증한다. 목표 변수가 선택되어 있어야 하고 예측변수가 최소 하나는 포함되어야 한다. 감정평가/시장 모드에서는 위도와 경도 열이 반올림되고, 기준일과 계약일 (지정된 경우) 로부터 sale age 가 계산된다.

8.2 적합 모달과 추적 출력

Fit 을 클릭하면 어두운 모달 오버레이가 나타나며 다음을 담고 있다:

- 머릿글 — “Fitting Earth Model” 과 함께 초 단위 경과 타이머, 그리고 배경 적합을 취소하는 **Abort** 버튼
- 추적 로그 — `earth()` 함수의 실시간 출력을 보여 주는 스크롤 가능한 고정폭 텍스트 영역: 데이터셋 정보, 전진 단계 진행 상황, 교차검증 분할, 완료 시각
- 색 구분 — 기본 줄은 초록색, 교차검증 줄은 노란색, 오류는 빨간색

earthUI 는 `callr::r_bg()` 를 사용하여 모형을 비동기적으로 적합시키며, 이는 계산을 별도의 R 프로세스에서 실행한다. 덕분에 오래 걸리는 적합 중에도 응용 프로그램이 반응성을 유지한다. 300 밀리초 주기의 폴링 관찰자가 배경 프로세스의 출력을 읽어 추적 로그로 흘려보낸다.

적합이 끝나면 “Done in X.Xs” 메시지가 나타나고 모달에 닫기 버튼 (X) 이 추가되며 배경막이 제거되어 결과 탭이 필요에 따라 그려질 수 있게 된다. 이후 보이는 출력이 모두 그려지면 모달은 저절로 사라지며, 그전에 직접 닫을 수도 있다.

성공하면 Fit 버튼에 초록색 체크 표시가 붙는다. 오류가 발생하면 오류 메시지가 추적 로그에 표시되고 알림이 나타난다.

8.3 적합 로그

모든 적합은 성공이든 실패든 출력 폴더에 로그 파일을 기록한다:

- 파일 이름: `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- 내용: 타임스탬프, 적합 과정의 모든 추적 출력, 그리고 오류 메시지
- 위치: 활성 프로젝트의 출력 폴더

팁: `callr` 가 설치되어 있지 않으면 `earthUI` 는 간단한 진행 표시줄과 함께 동기 적합으로 되돌아간다. 완전한 추적 출력을 보려면 `callr` 를 설치하라.

9 데이터 내려받기

적합 후 사이드바 6 절 — General 모드에서는 “Download Estimated Target Variable(s) & Residuals”, 감정평가 및 시장 모드에서는 “Download Estimated Sale Prices & Residuals” 라는 제목이다 — 은 예측값과 진단 정보가 담긴 Excel 파일을 내려받는다. 이 출력은 7 단계 (RCA) 에서 대상 부동산에 CQA(Condition-Quality-Appeal, 상태-품질-선호도) 등급을 부여하는 데 사용된다. 출력은 `residual_sf` 와 `cqa_sf` 로 정렬되어 대상 부동산이 순위에서 어디에 놓이는지 판단하는 데 도움을 준다. 모형의 품질이 좋다면 부동산들은 회귀분석에 들어가지 않은 잔차 특성에 근거하여 선호도가 가장 낮은 것부터 가장 높은 것까지 정렬되어야 한다. 가운데 값은 대략 0 이어야 하고, 아래쪽 절반은 음수, 위쪽 절반은 양수여야 한다. 품질이 가장 나쁜 주택, 즉 “수리 대상” 은 순위 아래쪽에, 가장 좋은 주택은 위쪽에 나타날 것이다. 압류, 숯세일, 상속 관련 매매, 직장 이동 등으로 인한 급매처럼 이례적인 경우에는 대개 예외가 생긴다. 이런 이상값을 조사해 보면 대개 가격 이상을 설명하는 타당한 이유가 발견된다.

버튼 레이블은 용도 모드에 따라 달라진다:

- General/Market: **Download Output (Excel)**
- Appraisal: **Download Intermediate Output (Excel)**

파일 이름 형식은 `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx` 이다.

9.1 출력 열

목표변수마다 다음 열들이 덧붙는다:

열	설명
<code>est_<target></code>	모형 예측값. 예: <code>est_sale_price</code> (소수점 1 자리)
<code>residual</code>	실제값에서 예측값을 뺀 값 (소수점 1 자리)
<code>cqa</code>	상태-품질-선호도 점수 (소수점 2 자리, 범위 0-10)
<code>residual_sf</code>	잔차를 주거 면적으로 나눈 값 (지정된 경우, 소수점 1 자리)
<code>cqa_sf</code>	<code>residual_sf</code> 순위로부터 계산한 CQA(소수점 2 자리)
<code><variable>_contribution</code>	g 함수별 기여도 값 (소수점 1 자리, g 함수마다 한 열)
<code>basis</code>	절편 값의 기여분. 모든 부동산에서 동일하다 (소수점 1 자리)
<code>calc_residual</code>	실제값에서 (절편 + 모든 기여도) 를 뺀 값 — 검증용 열 (소수점 1 자리)

다중 목표 모형에서는 목표변수별 열을 구분하기 위해 `_<i>` 접미사가 붙는다.

9.2 열 순서와 Excel 형식

순위 열들은 출력 파일의 맨 왼쪽에 다음 순서로 놓인다: `residual_sf`, `cqa_sf`, `residual`, `cqa`. 덕분에 정렬된 목록을 훑어보며 대상 부동산이 CQA 분포에서 어디에 놓이는지 평가하기 쉽다.

이 열들에는 Excel 숫자 서식이 적용된다:

열	서식	예
<code>residual_sf</code>	숫자, 소수점 2 자리	12.34
<code>cqa_sf</code>	숫자, 소수점 2 자리	7.25
<code>residual</code>	숫자, 소수점 0 자리	15,200
<code>cqa</code>	숫자, 소수점 2 자리	6.80

9.3 CQA 점수

CQA(상태-품질-선호도) 점수는 각 행의 잔차를 다른 모든 잔차와 견주어 순위를 매긴다. 어떤 행의 CQA 는 그보다 작은 부호 있는 잔차를 갖는 행의 비율에 10 을 곱한 값이다. 그 결과 0-10 척도가 만들어지며 다음과 같이 해석한다:

- 높은 CQA($\approx 9-10$) — 그 부동산이 모형 예측보다 훨씬 비싸게 팔렸다 (큰 양의 잔차)
- 낮은 CQA($\approx 0-1$) — 그 부동산이 예측보다 훨씬 싸게 팔렸다 (큰 음의 잔차)
- CQA ≈ 5 — 잔차가 전체 잔차의 중앙값 근처에 있다

living_area 열이 지정되면 cqa_sf 가 제곱피트당 잔차에 근거한 동일한 순위를 제공한다.

9.4 정렬과 대상 부동산 행

감정평가 모드와 시장 모드에서는 비교사례 행이 residual_sf 내림차순으로 정렬된다 (주거 면적이 지정되지 않았다면 residual 기준). 대상 부동산 행 (1 행) 은 1 번 자리에 그대로 남고 정렬되지 않는다. General 모드에서는 행이 원래 순서대로 내보내진다.

이 정렬은 맨 왼쪽에 놓인 순위 열들과 함께, 감정평가사가 가장 과대 예측된 비교사례부터 가장 과소 예측된 비교사례까지 빠르게 훑어보고 대상 부동산에 부여된 CQA 점수가 그 분포의 어디에 놓이는지 평가할 수 있게 한다.

예측 데이터의 요인 수준은 predict() 를 호출하기 전에 학습 데이터와 정렬된다. 학습 때 보지 못한 요인 수준을 가진 행은 NA 예측값을 낸다.

팁: calc_residual 열은 (반올림 오차 범위 안에서) 언제나 residual 과 같아야 한다. 둘이 크게 다르다면 살펴볼 만한 계산 문제가 있다는 뜻이다.

9.5 mgcvUI 자동 내보내기

degree 가 ≤ 2 인 적합이 성공할 때마다 earthUI 는 전체 결과 객체를 .rds 파일로 프로젝트의 출력 폴더에 자동 저장한다. 파일 이름은 <datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds 형식을 따른다.

이 파일은 earthUI 패키지에 포함되어 있고 earthUI::launch_mgcv() 로 시작하는 mgcv 후처리 모듈이 readRDS() 로 읽어 들일 수 있다. GAM 루틴은 earth 모형의 매듭점 위치와 기저 함수를 GAM 평활항의 출발점으로 사용하여 MARS 에서 GAM 모형화로 매끄럽게 넘어갈 수 있게 한다.

degree 가 2 를 넘는 모형은 mgcvUI 가 쌍별 상호작용만 지원하므로 건너뛴다. earthUI 를 Trilogy 모드 (Trilogy UI 에서) 로 실행하면 추가로 Lock Model Output 절이 생겨, 특정 적합의 .rds 를 그 예측변수 기저 및 공유 CQA 설정과 함께 glmnetUI 와 mgcvUI 가 가져올 정본 earth 모형으로 잠글 수 있다.

9.6 glmnetUI 가져오기

GAM 루틴을 위해 저장된 것과 같은 .rds 파일을 엘라스틱 넷 후처리 모듈로도 가져올 수 있다. 이 모듈 역시 earthUI 패키지에 포함되어 있으며 earthUI::launch_glmnet() 으로 시작한다. glmnet 응용 프로그램에서 Section 2: Import from earthUI 로 이동한 뒤 Browse 버튼으로 .rds 파일을 선택한다.

glmnetUI 는 표준 earth::model.matrix() 방식을 사용한다. earth 모형의 기저 함수 (힌지, 상호작용, 선형 항) 가 glmnet 설계 행렬의 열이 된다. 그러면 glmnet 은 이 기저 열들에 대해 정규화 회귀를 수행하여 라쏘/엘라스틱 넷으로 선택하고 축소한다. 이는 earth 의 적응적 기저 구성과 glmnet 의 정규화를 결합한다.

9.6.1 주요 이점

- **해석 가능한 상호작용:** earth 의 힌지 기반 상호작용 (3 원 항 포함) 은 명확한 기하학적 의미를 지니며, 이는 법정이나 감사 상황에서 설명하기 어려운 glmnet 고유의 교차곱 상호작용과 대조된다.
- **자동 비선형성:** earth 의 힌지 함수는 원 예측변수를 쓰는 표준 glmnet 모형이 놓칠 비선형 관계를 포착한다.
- **변수 선택:** glmnet 은 기여하지 않는 힌지 항을 걸러 내며 earth 기저를 한 번 더 가지치기할 수 있다.

9.6.2 중요: 가중치 열이 일치해야 한다

경고

earthUI 가 가중치 열을 사용한다면 (예: 가중치 0 으로 대상 부동산 행을 제외하는 경우), glmnetUI 의 변수 구성에서도 동일한 가중치 열을 지정해야 한다. 가중치 설정이 다른 행 수가 맞지 않아 earth 기저를 glmnet 의 모형 행렬에 덧붙일 수 없다.

9.6.3 작업 흐름

1. earthUI 에서 earth 모형을 적합시킨다 (degree 는 상관없다).
2. .rds 결과 파일이 프로젝트의 출력 폴더에 자동 저장된다.
3. glmnetUI 에서 **Section 2: Import from earthUI** 를 열고 .rds 파일을 찾는다.
4. 두 응용 프로그램에서 같은 데이터 파일과 같은 가중치 열을 쓰고 있는지 확인한다.
5. 5 절에서 glmnet 매개변수를 구성한다 (earthUI 가져오기가 활성화되면 상호작용 행렬은 자동으로 비활성화된다).
6. **Fit Model** 을 클릭한다. earth 기저 열들이 Equation, Coefficients, Contributions 탭에 나타난다.

10 RCA 계산과 내려받기

Calculate RCA Adjustments & Download 버튼은 사이드바 7 절에 나타나며, 감정평가 모드에서 모형을 적합시킨 뒤에만 보인다. RCA(잔차 제약 접근법, Residual Constraint Approach)는 대상 부동산 대비 비교사례별 시장 도출 보정치를 산출한다.

10.1 RCA 대화 상자 열기

버튼을 클릭하면 다음을 담은 작은 모달 대화 상자가 열린다:

- **Score type** — **CQA** 또는 **CQA per SF** 를 고르는 라디오 버튼 (SF 선택지는 **living_area** 열이 지정된 경우에만 사용할 수 있다). “CQA per SF” 를 고르면, 대상 부동산에 부여한 CQA 점수에 근거하여 그 잔차 점수는 주거 면적에 해당 CQA_SF 점수에 대응하는 **residual_sf** 를 곱한 값이 된다.
- **CQA value** — 대상 부동산의 CQA 점수 입력란 (0.00–10.00). 이 점수는 비교사례들의 품질 분포에서 대상 부동산이 어디에 놓인다고 보는지를 나타낸다.
- **Generate** 버튼 — RCA 출력을 계산하여 Excel 로 내려받는다. 유효한 점수를 입력하기 전까지는 버튼이 비활성 상태이다.

10.2 CQA 점수 보간

Generate 를 클릭하면 earthUI 는 비교사례의 CQA/잔차 쌍으로부터 대상 부동산의 잔차를 보간한다:

1. 비교사례의 CQA 점수와 잔차를 CQA 오름차순으로 정렬한다
2. 선형 보간 (`stats::approx()`) 으로 입력한 CQA 값을 잔차에 대응시킨다
3. CQA per SF 를 사용하는 경우, 제곱피트당 잔차에 대상 부동산의 주거 면적을 곱하여 총 잔차로 되돌린다
4. 대상 부동산의 추정 가치를 다음과 같이 계산한다: 모형 예측값 + 보간된 잔차

10.3 출력 열

RCA Excel 파일 (<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx) 에는 모든 중간 출력 열에 더해 다음이 포함된다:

열	설명
subject_value	모형 예측값 + 보간된 잔차 (1 행에만)
subject_cqa	입력한 CQA 점수 (1 행에만)
<variable>_adjustment	대상 부동산의 기여도에서 비교사례의 기여도를 뺀 값 (g 함수별)
residual_adjustment	대상 부동산의 잔차에서 비교사례의 잔차를 뺀 값
net_adjustments	모든 보정치의 합 (기여도 + 잔차)
gross_adjustments	모든 보정치 절댓값의 합
residual_pct	비교사례 매매가격 대비 잔차 보정의 비율
net_adj_pct	비교사례 매매가격 대비 순보정의 비율
gross_adj_pct	비교사례 매매가격 대비 총보정의 비율
adjusted_sale_price	비교사례 매매가격 + 순보정

보정 열들은 각 비교사례에 대해 모형이 그 차이를 각 변수에 얼마나 귀속시키는지를 알려 준다. **adjusted_sale_price** 는 모형에서 도출된 모든 보정을 적용한 뒤의 비교사례 매매가격으로, 이렇게 얻은 보정 매매가격들은 대상 부동산의 추정 가치 주변에 모여야 한다.

10.4 RCA Adjustments 탭

RCA 출력이 계산되면 주 패널의 **RCA Adjustments** 탭에 비교사례 전반의 Residual Adj %, Net Adj %, Gross Adj % 분포 히스토그램이 표시되어 (다중 목표 모형에서는 목표변수마다 한 별) 보정치가 얼마나 촘촘히 모여 있는지 시각적으로 빠르게 확인할 수 있다.

팁: CQA 5.00 점은 대상 부동산을 비교사례의 중앙값에 놓는다. 5 보다 높은 점수는 대상 부동산이 평균 이상의 품질임을, 5 보다 낮은 점수는 평균 이하임을 뜻한다.

11 거래사례비교표

11.1 개요와 목적

거래사례비교표 (Sales Comparison Grid) 는 RCA 보정치를 계산한 뒤 감정평가 모드 (사이드바 8 절) 에서 생성되는 Excel 통합문서이다. 대상 부동산과 선택된 비교사례를 구조화된 표 형식으로 나란히 제시하며, Excel 수식이 각 비교사례의 보정치와 보정 매매가격을 자동으로 계산한다.

이 표는 감정평가사의 작업 파일을 위해 설계되었다. Earth 모형에서 도출된 회귀 기반 보정치와, 감정평가사가 CQA 잔차를 특정 부동산 특성에 배분할 수 있는 편집 가능한 셀을 결합한다. 출력 파일 이름은 SalesGrid_<YYYYMMDD_HHMMSS>.xlsx 이다.

11.2 비교사례 선택 모달

“Generate Sales Grid & Download” 를 클릭하면 포함할 비교사례를 선택하는 모달 대화 상자가 열린다:

- 권장 비교사례가 미리 선택되어 있다. 이는 총보정 비율이 매매가격의 25% 미만인 비교사례로, sale age 오름차순 (가장 최근 매매부터) 으로 정렬되어 최대 30 건까지 표시된다.
- 추가 비교사례가 그 아래에 기본적으로 선택되지 않은 상태로 나열된다. 총보정이 25% 이상이지만 여전히 선택할 수 있는 것들이다 (최대 50 건 표시, 총보정 비율 오름차순 정렬).
- 최대 **30** 건의 비교사례를 선택할 수 있으며, 최대 **10** 개 시트가 만들어진다 (시트당 비교사례 3 건).

선택을 확정하면 선택된 비교사례는 총보정 비율 오름차순으로 정렬되므로 가장 유사한 비교사례가 시트 1 에 나타난다.

11.3 표 배치

각 시트는 대상 부동산과 비교사례 3 건을 수용하는 20 열 배치를 갖는다. 각 개체의 열 구성은 다음과 같다:

개체	열 (개체당 5 개)
대상 부동산	Label, Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC)
각 비교사례	Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC), Adjustment

행은 위에서 아래로 다음과 같다:

1. 제목 — 시트 이름 (예: “Intermediate Sales Comparable Grid — Sheet 1 of 3”)
2. 머리글 — “Subject” 와 주소가 포함된 비교사례 열 머리글
3. Address — 부동산의 전체 주소
4. APN | MLS# | DOM | Subj.Prox — 필지 번호, 등록 ID, 시장 노출 일수, 그리고 대상 부동산으로부터의 하버사인 거리 (마일)
5. Sales Price | Concess. | Net SP — 매매가격, 양허, 그리고 순매매가격 수식
6. Regression Features 머리글 행
7. BASE VALUE — 모형 절편
8. Date of Sale | OffMkt | OnMkt — 계약일, sale age, DOM
9. 묶음 행 (조건부) — Location, Site Size, Age 그룹
10. 모형 변수 행 — 예측변수마다 한 행 (묶음 변수 제외)
11. 빈 구분 행
12. Residual Features 머리글 행
13. CQA|Residual — CQA 점수와 남은 잔차 수식

14. 잔차 특성 행 — 이름이 붙은 행과 감정평가사 입력용 빈 행
15. Total VC / Net Adjustment
16. Net Adjustment %
17. Gross Adjustment %
18. Adjusted Sale Price — 주식 행
19. 저작권 꼬리말

11.4 매매가격, 양허, Net SP

5 행은 각 비교사례에 대해 세 가지 값을 보여 준다:

- **Sale Price** — 데이터에 있는 비교사례의 매매가격
- **Concessions** — concessions 로 지정된 열의 값 (지정되지 않았으면 0)
- **Net SP** — Excel 수식: Sale Price – Concessions

대상 부동산 열은 매매가격을 알 수 없으므로 “N/A” 를 표시하며, 양허 값이 있으면 그 값을 보여 준다.

11.5 묶음 행 (위치, 부지, 연령)

특정 특수 유형이 지정되어 있고 그 변수들이 모형에 나타나면, earthUI 는 관련 변수를 결합한 묶음 행을 만든다:

Loc: Long | Lat | Area — longitude, latitude, area 중 어느 하나라도 모형에 있으면 나타난다. 구성 변수 각각의 사실값, 결합된 가치 기여도 (개별 VC 의 합), 그리고 비교사례에 대해서는 보정치 (대상 부동산의 결합 VC – 비교사례의 결합 VC) 를 보여 준다. VC 셀은 연한 파란색 배경으로 꾸며진다.

Site Size | Dimensions — lot_size 나 site_dimensions 가 모형에 있으면 나타난다. 구조는 Location 그룹과 같다.

Actual Age | Effective Age — actual_age 나 effective_age 가 모형에 있으면 나타난다. 구조는 동일하다.

묶음 행에 쓰인 변수는 아래의 개별 모형 변수 행에서 제외되어 보정 합계에서 중복 계산되는 일을 막는다.

11.6 모형 변수 행

묶음 행 아래 (묶음 행이 없으면 BASE VALUE 바로 아래) 에는 남은 모형 예측변수마다 한 행씩 다음을 보여 준다:

- 대상 부동산과 각 비교사례의 사실값 (실제 데이터 값)
- 가치 기여도 (VC) — 해당 행의 예측값에 대한 g 함수의 기여분
- 보정치 (비교사례 전용) — 대상 부동산의 VC 에서 비교사례의 VC 를 뺀 값

11.7 CQA|Residual 및 잔차 특성 행

CQA|Residual 행은 각 부동산의 CQA 점수를 보여 주며 남은 잔차, 즉 아직 특정 특성에 배분되지 않은 잔차 부분을 구하는 수식을 담고 있다. 수식은 다음과 같다:

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

그 아래에는 잔차 특성 행이 있다. 이름이 붙은 5 개 행 (Location, View/Appeal, Condition, Quality, Other) 과 빈 행 6 개이다. 이 셀들은 감정평가사가 모형이 포착하지 못한 특성의 가치 기여도를 입력하는 입력 셀이다. 감정평가사가 값을 채워 넣으면 남은 잔차 수식의 값이 자동으로 줄어든다.

각 비교사례의 보정 열에는 수식이 들어 있다: 대상 부동산의 특성 VC – 비교사례의 특성 VC.

11.8 보정 매매가격 수식

Adjusted Sale Price 행에는 Excel 수식이 들어 있다:

- **대상 부동산:** BASE VALUE 부터 마지막 잔차 특성 행까지의 모든 가치 기여도 셀의 합. 이는 대상 부동산에 대한 모형의 총 예측값에 감정평가사가 배분한 잔차 특성을 더한 값이다.
- **비교사례:** Net SP 에, 첫 묶음 행 또는 변수 행부터 마지막 잔차 특성 행까지의 모든 보정 셀의 합을 더한 값. 이는 회귀에서 도출된 보정과 감정평가사가 입력한 보정을 모두 적용한 뒤의 비교사례 매매가격이다.

11.9 시트 보호와 편집 가능 셀

각 시트는 수식과 데이터가 실수로 수정되는 것을 막기 위해 **보호**되어 있다. 보호는 다음을 잠근다:

- 모든 수식 셀 (Net SP, 남은 잔차, 보정치, 보정 매매가격 등)
- 모든 데이터 셀 (주소, 매매가격, 사실값, 모형에서 나온 가치 기여도)
- 레이블과 머리글

잠금이 풀린 (편집 가능한) 셀은 잔차 특성의 가치 기여도 입력란, 즉 CQA|Residual 행 아래에서 감정평가사가 내역을 입력하는 셀들뿐이다. 이 셀들은 편집 가능성을 나타내기 위해 연한 노란색 배경으로 꾸며져 있다.

11.10 Excel 에서 표 다루기

1. Excel 또는 호환되는 스프레드시트 프로그램에서 **파일을 연다**.
2. **회귀 보정치를 검토한다** — 모형에서 도출된 보정치는 미리 채워져 있고 잠겨 있다.
3. **잔차를 배분한다** — 잔차 특성 행 (노란색 셀) 에 남은 잔차 중 얼마만큼이 각 특성 (Location, View, Condition, Quality 등) 에 귀속되는지에 대한 판단을 입력한다. 배분함에 따라 남은 잔차 수식의 값이 줄어든다.
4. **보정 매매가격을 확인한다** — 잔차 특성 값을 입력하면 이 수식이 자동으로 갱신된다.
5. **여러 시트** — 비교사례를 3 건 넘게 선택했다면 시트 사이를 옮겨 다니라. 각 시트는 왼쪽 열에 같은 대상 부동산을 두고 서로 다른 비교사례 3 건을 보여 준다.

팁: 목표는 남은 잔차가 0 에 가까워질 때까지 CQA 잔차를 특성 행에 배분하는 것이다. 이는 각 비교사례의 회귀 추정치와 실제 매매가격의 차이를 설명할 수 있음을 보여 준다.

12 보고서 내려받기

모형을 적합시킨 뒤의 보고서 생성은 두 단계의 Quarto 작업 흐름이다. **Generate Quarto Report**(감정평가 모드에서는 사이드바 9 절, 그 외에는 7 절) 가 자체 완결형 보고서 묶음을 기록하고, **Convert Quarto Report**(10 절) 가 그것을 — 또는 다른 어떤 Quarto 파일이라도 — 최종 형식으로 렌더링한다.

12.1 Generate Quarto Report

Generate Quarto Report 를 클릭하면 프로젝트 출력 폴더 아래 `<base>_qmd/`에 자체 완결형 Quarto 묶음이 기록된다. 여기에는 값이 채워진 .qmd 원본, 미리 생성된 모든 그림 자산 (PNG 와 PDF), 보고서 데이터, 그리고 Word 참조 서식 파일이 들어 있다. 이 단계에서는 렌더링이 일어나지 않는다. 묶음은 직접 편집할 수도 있고, Quarto include 를 통해 다른 Quarto 원본과 결합할 수도 있으며, quarto preview 로 단독 미리보기를 할 수도 있다.

12.2 Convert Quarto Report

10 절은 파일 선택기 (기본값은 현재 프로젝트에서 가장 최근에 생성된 묶음이지만 어떤 .qmd 파일로도 찾아갈 수 있다) 와 형식 확인란을 제공한다:

- **HTML** — KaTeX 수식 렌더링, 내장 이미지, 목차를 갖춘 .html 파일. Flatly Bootstrap 테마와 Roboto Condensed 글꼴을 사용한다.
- **Word** — 편집과 배포에 적합한 .docx 파일. 목차와 쪽 번호를 포함하며, 일관된 서식을 위해 사용자 지정 참조 서식 파일을 사용한다.
- **PDF** — 전문가 수준의 출력을 위해 LuaLaTeX 으로 조판된다. 용지 크기는 로캘 설정 (Letter 또는 A4) 을 따른다. 큰 상호작용 행렬을 위한 가로 방향 페이지를 포함하며, 본문에는 Roboto Condensed, 수식에는 Latin Modern Math 를 사용한다. PDF 선택지는 LaTeX 배포판이 감지될 때에만 제공된다.

Convert 를 클릭하면 선택한 형식으로 렌더링된다. 렌더링은 배경에서 실행되며, 모달 대화 상자가 경과 시간과 Quarto 진행 상황을 보여 주는 동안에도 응용 프로그램은 반응성을 유지한다. 완료되면 알림이 파일 위치를 알려 준다.

모든 작업 (모형 적합, 데이터 내려받기, RCA 계산, 거래사례비교표 생성, 보고서 렌더링) 은 시작/종료 타임스탬프와 경과 시간과 함께 출력 폴더의 `<datafile>_earthui_log.txt` 에 기록되어 문제 해결과 성능 점검에 쓰인다.

12.3 보고서 내용

생성되는 모든 보고서에는 다음 절들이 들어 있다:

1. 데이터셋 설명 — 관측치 수, 목표변수, 사용된 예측변수, 범주형 예측변수
2. 모형 설정 — degree, 교차검증 여부, 항의 수, 모형에 들어간 예측변수
3. 허용 상호작용 (degree ≥ 2 인 경우에만) — 체크 표시가 있는 상호작용 행렬. 출력 형식에 맞게 서식이 갖추어진다
4. 결과 요약 — R^2 , GRSq, GCV, RSS, 그리고 (교차검증이 켜져 있었다면) $CV R^2$
5. 모형 방정식 — LaTeX 표기로 나타난 완전한 Earth 모형 방정식
6. 계수와 기저 함수 — 모든 항, 그 계수, 그리고 그것들을 정의하는 기저 함수의 표
7. 변수 중요도 — 예측변수 중요도 점수의 막대그래프와 순위 표
8. g 함수 기여도 — g 함수별 그림. 이변량 항은 선 그림, 이변량 상호작용은 3 차원 투시도와 등고선 그림
9. 상관 행렬 — 예측변수 상관관계 히트맵
10. 진단 — 잔차 대 적합값, 정규 Q-Q, 실제값 대 예측값 그림
11. ANOVA 분해 — 각 기저 함수의 RSS 기여도를 보여 주는 표

12. **Earth** 출력 — 가지치기 단계 세부 정보를 포함한 `earth::summary()` 의 원시 텍스트 출력

다중 반응 모형에서는 4-8 절과 10 절이 목표변수마다 반복된다.

팁: 보고서에는 *quarto R* 패키지와 *Quarto* 설치가 필요하다. *PDF* 출력에는 *LaTeX* 배포판도 설치되어 있어야 하며 (예: `tinytex::install_tinytex()`), 그렇지 않으면 *PDF* 확인란이 숨겨진다.

13 후처리 모듈: 엘라스틱 넷과 GAM

earthUI 패키지에는 두 개의 후처리 모듈이 함께 제공된다. 하나는 `glmnet` 기반의 엘라스틱 넷 루틴이고, 다른 하나는 `mgcv` 기반의 GAM 루틴이다. 각각은 이번 호의 별도 논문에서 깊이 다루므로, 이 절에서는 그것들이 earthUI 작업 흐름에서 무엇을 위한 것인지, 그리고 세 루틴이 어떻게 협력하는지를 설명한다.

13.1 왜 보조 적합인가?

가치 결론을 내는 것은 earth 모형이다. 후처리 모듈은 그것을 보강하기 위해 존재한다. 같은 프로젝트 데이터에 적합된, 방법론적으로 다른 엔진이 산출한 동일 대상 부동산에 대한 두 개의 추가 추정치인 것이다. earth 기저의 엘라스틱 넷 재추정과 독립적인 GAM 평활이 실질적으로 같은 가치에 도달한다면, 감정평가사의 결론은 더 이상 한 모형의 가정에만 기대지 않게 된다. 이는 심사와 소송 상황에서 의미 있는 이점이다. 반대로 결과가 어긋난다면 그 불일치 자체가 진단이 되어, 대개 데이터 부족, 지나치게 유연한 항, 또는 다시 살펴볼 만한 이상 비교사례를 가리킨다. 어느 모듈도 earth 모형보다 우위에 있지 않다. 이들은 보조 증거이며, 의도적으로 가중치 없이 제시된다.

13.2 실행과 상태 공유

각 루틴은 별도의 로컬 Shiny 응용 프로그램이다:

- `earthUI::launch_glmnet()` — 엘라스틱 넷 응용 프로그램, 포트 7879.
- `earthUI::launch_mgcv()` — GAM 응용 프로그램, 포트 7880.

세 응용 프로그램은 모두 같은 프로젝트 트리와 같은 설정 데이터베이스를 읽으므로, earthUI에서 열어 둔 프로젝트를 모듈에서도 곧바로 쓸 수 있다. earthUI가 그 프로젝트에 대해 저장하는 이월 필드, 즉 반응 (목표) 변수, 기준일, RCA 대상 부동산 CQA 유형과 값, 주거 면적 지정도 함께 넘어간다. 이 값들을 다시 입력할 필요는 없다. 모듈이 그것을 읽어 오므로 같은 값이 세 적합 모두에 흐른다.

13.3 엘라스틱 넷 루틴

`glmnet` 루틴은 earth 모형의 기저 함수 — 힌지, 상호작용, 선형 항 — 를 설계 행렬의 열로 가져와 라쏘/엘라스틱 넷 정규화 아래에서 재추정한다 (작동 방식과 가중치 열 관련 주의 사항은 위의 “`glmnetUI` 가져오기” 참조). 그 결과는 earth의 해석 가능하고 기하학적 의미가 분명한 항들을 유지하면서, 제 몫을 하지 못하는 항은 정규화가 축소하거나 제거하도록 한다. 자세한 내용은 이번 호의 `glmnetUI` 논문을 보라.

13.4 GAM 루틴

`mgcv` 루틴은 반대 방향으로 간다. earth의 기저를 재사용하는 대신, earth 모형의 매듭점 위치를 출발점으로 삼아 새로운 벌점 평활항을 적합시킨다 (위의 “`mgcvUI` 자동 내보내기” 참조). GAM은 다른 평활화 패러다임 아래에서 함수 형태를 다시 도출하므로, earth 모형과의 일치는 메아리가 아니라 진정으로 독립적인 보강이 된다. `degree`가 ≤ 2 인 earth 모형이 지원된다. 자세한 내용은 이번 호의 `mgcvUI` 논문을 보라.

13.5 실무적 작업 흐름

1. earthUI에서 만족스러울 때까지 **earth 모형**을 먼저 적합시키고 다듬는다. 성공한 적합 ($\text{degree} \leq 2$) 마다 모듈이 사용할 `.rds`가 자동으로 내보내진다.
2. 같은 프로젝트에 대해 **glmnet**과 **mgcv** 루틴을 열고 earth 결과를 가져와 적합시킨다. 한 번에 하나의 루틴만 다루라. 적합은 오래 걸리고 루틴들은 공유 프로젝트를 통해 협력하므로, 합리적인 리듬은 병렬이 아니라 순차이다.

3. 자유롭게 반복하라. 각 응용 프로그램은 전환하는 동안에도 결과를 유지하며, 도중에 보고서를 만들어야 할 이유는 전혀 없다.
4. 세 결과가 모두 만족스러울 때에만 각 루틴으로 돌아가 Quarto 보고서를 생성하라. 각 방법은 자신의 적합과 가치 결론을 프로젝트의 trilogy 기록에 등록하므로, 세 보고서가 하나의 비교 실행으로 추적된다.

팁: 모듈은 어느 수준에서도 선택 사항이다. 적합, RCA, 거래사례비교표, 보고서로 이어지는 *earthUI* 자체의 작업 흐름은 모듈 없이도 완결된다. 어려운 대상 부동산, 얇은 시장 부문, 또는 심사를 받게 될 가능성이 높은 업무처럼 결론에 보강이 필요할 때 모듈을 활용하라.

14 데모 데이터셋: Appraisal_1.csv

earthUI 에는 감정평가 작업 흐름을 살펴볼 수 있는 데모 MLS 데이터셋이 들어 있다. 프로그램 방식으로의 다음과 같이 불러온다:

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

또는 Shiny 응용 프로그램에서 이 파일을 프로젝트의 in/ 폴더로 복사하거나 (프로젝트를 만들 때 최초 데이터 파일로 첨부해도 된다) 2 절 (Import Data) 에서 선택한다.

14.1 설명

이 파일에는 모의 MLS 내보내기에서 나온 주거용 매매사례 1,501 건과 1 행의 대상 부동산 1 건 (총 1,502 행) 이 들어 있다. 데이터는 부동산 규모, 연령, 위치가 다양한 여러 지역으로 이루어진 시장의 단독주택 매매를 나타낸다.

이는 실제 데이터가 아니지만 북부 캘리포니아의 현실적인 근린지역에 기초하고 있다. 모든 식별 정보는 변경되거나 제거되었다.

14.2 열

열	자료형	특수 유형	설명
weight	numeric	—	관측치 가중치 (0 = 적합에서 제외)
id	numeric	display_only	내부 레코드 ID
property_id	numeric	display_only	MLS 부동산 식별자
listing_id	character	display_only	MLS 등록 번호
parcel_number	character	display_only	카운티 과세평가관 필지 번호 (APN)
street_address	character	display_only	부동산 주소
city_name	character	display_only	도시
postal_code	character	display_only	우편번호
county_name	character	display_only	카운티
contract_date	Date	contract_date	매매 계약일 (sale_age 계산)
sale_age	numeric	—	계약일부터 기준일까지의 일수 (미리 계산됨)
coe_date	Date	display_only	에스크로 종결일
listing_status	character	display_only	등록 상태 (예: "Sold")
sale_price	numeric	(목표변수)	매매가격 — 반응변수
rent	numeric	—	월 임대료 (다중 목표 모형용)
list_price	numeric	display_only	등록 가격
original_list_price	numeric	display_only	최초 등록 가격
living_sqft	numeric	living_area	총 주거 면적 (제곱피트)
beds_total	integer	—	침실 수
baths_total	numeric	—	총 욕실 수 (예: 2.5 = 전체 2 + 반 1)
lot_size	numeric	lot_size	부지 면적 (제곱피트)
area_id	integer	area	MLS 지역 식별자
area_text	character	display_only	지역명
age	numeric	actual_age	부동산 연령 (년)
year_built	integer	display_only	준공 연도
latitude	numeric	latitude	위도 (모형용으로 소수점 셋째 자리 반올림)
longitude	numeric	longitude	경도 (모형용으로 소수점 셋째 자리 반올림)
latitude6	numeric	display_only	전체 정밀도 위도
longitude6	numeric	display_only	전체 정밀도 경도
garage_spaces	integer	—	차고 주차 칸 수
fp_count	integer	—	벽난로 수
no_of_stories	numeric	—	층수
style	character	—	건축 양식
view	character	—	조망 유형 (예: "Neighborhood", "Hills")
days_on_market	integer	dom	시장 노출 일수
listing_date	Date	listing_date	등록일
sale_concessions	numeric	concessions	매도인 양허

14.3 권장 빠른 시작

1. earthUI 를 실행한다: `earthUI::launch()`
2. 새 프로젝트 마법사에서 Purpose 를 **For Appraisal** 로 하여 프로젝트를 만들고 **Appraisal_1.csv** 를 최초 데이터 파일로 첨부한다
3. 2 절 (Import Data) 에서 **Appraisal_1.csv** 를 선택한다
4. 목표변수로 **sale_price** 를 선택한다
5. 위 표에 나온 대로 특수 유형을 부여한다
6. 예측변수를 포함시킨다: **sale_age, living_sqft, baths_total, lot_size, area_id**(요인으로), **age, latitude, longitude, garage_spaces**
7. degree 를 1 로 설정하고 **Fit Earth Model** 을 클릭한다
8. 중간 출력을 내려받고 (6 단계) CQA 순위를 검토한다
9. CQA 점수 ~5.00 으로 RCA 보정치를 계산한다 (7 단계)
10. 거래사례비교표를 생성한다 (8 단계)

답: 데모 데이터셋은 자신의 *MLS* 데이터를 가져오기 전에 *earthUI* 의 모든 기능을 살펴보기에 좋은 방법이다. 위에 제시한 예측변수들은 대개 R^2 가 0.85 를 넘는 모형을 만들어 낸다.

15 실험적 기능: Prolog 텍스트 추출과 도출 규칙

0.10.0 버전은 MLS 내보내기의 서술형 텍스트 열 — 무엇보다도 “Public Remarks” 열, 그리고 중개인 비고와 방 특성이나 기존 구조물 같은 심포 구분 특성 목록 — 에서 모형에 바로 쓸 수 있는 특성을 추출하는 선택적이고 진정으로 실험적인 기능을 도입했다. 추출 엔진은 Prolog, 특히 한정절 문법 (DCG, Definite Clause Grammars) 이다. 비고 텍스트에서 “move in ready”, “3 car garage”, “owned solar” 같은 구절을 인식하는 규칙이 Prolog 문법 규칙으로 작성되어 있다.

이 기능은 아직 개발 중이다. 오늘도 추출에 부분적으로 쓸 수는 있다. 함께 제공되는 문법이 이미 유용한 특성 집합을 인식한다. 그러나 자신의 MLS 어휘에 맞게 조정하려면 Prolog, 그중에서도 DCG 를 읽고 쓸 수 있어야 한다. 그럴 형편이 아니라면 이 장은 통째로 건너뛰어도 된다. 이 기능은 기본적으로 꺼져 있으며, 그것 없이도 earthUI 의 나머지는 똑같이 동작한다.

경고

이것은 활발히 개발 중인 실험적 기능이다. 추출 문법은 완결되어 있지도, 실무 감정평가용으로 검증되어 있지도 않다. 추출된 열은 모형에서 신뢰하기 전에 원본 텍스트와 대조하여 검토해야 한다. 이 기능을 효과적으로 쓰려면 Prolog 와 DCG 에 대한 실무 지식이 필요하다. 이 기능이 하는 일은 모두 덧붙이기이고 되돌릴 수 있다. 문제가 생기면 설정 토글을 끄고 평소대로 계속하면 된다.

15.1 무엇을 하는가

MLS 등록 정보는 자유 텍스트와 구분자로 나뉜 목록 필드에 가치와 관련된 정보를 대단히 많이 담고 있다. 이 선택적 단계는 두 가지 특수 열 역할 (제 6 장) 을 통해 그 텍스트를 회귀분석에 쓸 수 있는 평범한 열로 바꾼다:

- **remarks** 열은 Prolog DCG 문법으로 구문 분석되어 추출된 속성마다 한 개의 특성 열이 만들어진다. 열 접두사는 원본 열 이름의 첫 글자들로 만들어진다. `public_remarks` 는 `pr_*` 열을, `agent_remark` 는 `ar_*` 열을 낳는 식이다.
- **list** 열 (예: `existing_structures = "Shed, Workshop"`) 은 원-핫 `<col>_<value>` TRUE/FALSE 열로 분할된다. 이 부분은 순수한 R 이며 Prolog 가 전혀 필요 없다.

함께 제공되는 문법 (`inst/prolog/mls_remarks.pl`) 은 현재 다음과 같은 것들을 추출한다: 1-5 범위의 `condition_score` 를 갖는 등급화된 상태 계층, 조망 플래그와 조망 유형, 주방 및 욕실 개조 플래그, 원목 및 기타 바닥재, 개선 목록과 개수, 차고 칸 수와 차고 면적, 수영장, 스파, 기타 옥외 시설, 벽난로 수, “40x50” 같은 치수에서 크기를 읽어 낸 부속 건물 (창고, 헛간, 작업실, 별채, 게스트하우스, 견사 등), RV 주차와 보관, 태양광의 소유 대 임대 여부와 배터리 저장, 부지 면적 (에이커-제곱피트 환산 포함), 대상 부동산과 결부된 매매 맥락상의 금액 (가격 인하, HOA 회비, 태양광 임대료 또는 상환액, Mello-Roos), 위치 단서, 그리고 $[-1, 1]$ 범위의 장점/우려 감성 점수.

15.2 요구 사항: SWI-Prolog 가 아니라 vProlog(Trealla)

비고 구문 분석은 자체 완결형 Trealla Prolog 엔진인 **vProlog** R 패키지 위에서 실행된다. 이 패키지는 CRAN 에 없다. 이 논문 앞부분의 “earthUI 와 vProlog 설치” 에 나온 대로 저자의 r-universe 저장소에서 설치하라 (Windows 와 macOS 용 사전 빌드 바이너리, 소스는 github.com/wcraytor/vProlog). 시스템에 Prolog(SWI-Prolog 등) 를 설치할 필요도 없고 사용하지도 않는다. vProlog 가 없으면 비고 구문 분석은 안내 메시지와 함께 건너뛰고 `list` 분할은 그대로 실행된다. 도우미 함수 `vProlog_available()` 은 엔진과 함께 제공되는 문법이 모두 있는지 알려 준다.

15.3 활성화와 실행

1. **Settings**(톱니바퀴) 를 열고 **Enable Prolog / list column processing** 을 선택한다. 이 토글은 기본적으로 꺼져 있고, “advanced feature — for users with Prolog expertise” 라는 안내가 붙어 있으며, 프로젝트별로 기억되고, 프로젝트 정보 대화 상자에 표시된다.
2. 예측변수 표에서 해당 열들의 **Special** 역할을 **remarks** 나 **list** 로 설정한다 (각 역할에 여러 열을 지정할 수 있다).
3. 사이드바의 4 절과 5 절 사이에 있는 **Execute Prolog Processing** 절을 사용한다. 두 단계를 제공한다:

1. **Expand lists & remarks** — 각 **remarks** 열에 DCG 문법을 실행하고 각 **list** 열을 원-핫으로 분할한다. 확장은 점증적이다. 마지막 실행 이후에 지정된 열만 처리되며, 지정된 것이 모두 확장되고 나면 버튼이 비활성화된다. 진행 표시줄이 구문 분석 진척을 보여 준다.

2. **Execute derivation rules** — 확장된 열들에 대해 사용자가 작성한 Prolog `derive/2` 규칙을 적용한다 (아래 참조). 규칙은 실행 전에 구문 검사와 예행 검증을 거친다.

15.4 도출 규칙

earthUI 는 데이터 행마다 그 행의 열들을 `cell(Name, Value)` 사실로 단언한 뒤 사용자의 `derive(Col, Val)` 규칙의 모든 해를 모은다. 서로 다른 `Col` 마다 새 열이 되며 열당 첫 번째로 찾은 값이 채택되므로, 뒤에 대체 절을 둘 수 있다. 도우미 술어 (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) 가 집계와 NA 처리를 지원하며, `drop/1` 사실은 규칙 실행 후 열을 버린다. 예를 들면:

```
% 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).
```

규칙은 프로젝트마다 하나씩 있는 `<project>_rules.pl` 파일에 들어 있고, 주석이 충실히 달린 서식 파일에서 출발하며, **Edit derivation rules...**를 통해 사이드바에 도킹할 수도 있는 떠 있는 이동 가능한 코드 편집기에서 편집한다. **Check** 버튼은 규칙을 실행하지 않고 검증한다 (구문 검사, 적재 검사, 그리고 데이터 1 행에 대한 예행 실행).

15.5 버전이 매겨진 활성 데이터 파일

두 단계 중 어느 쪽이든 실행하고 나면 증강된 데이터 프레임이 프로젝트의 입력 폴더에 `<original_name>_<YYMMDD_HHMMSS>.xlsx` 로 기록되어 프로젝트의 활성 데이터 파일이 된다. 프로젝트를 다시 열면 증강된 데이터가 적재되므로 확장을 다시 계산할 일이 없고, `glmnetUI` 와 `mgcvUI` 가 사용하는 것도 같은 파일이다. 원본 비교 열은 자동으로 모형에서 제외되며, 생성된 열은 예측변수로 자동 포함되지 않는다. 텍스트 구문 분석은 많은 열을 만들어 낼 수 있으므로, (아직) 회귀에 넣고 싶지 않은 열은 **Latent** 로 표시하거나 $\times$ 로 비활성화하라 (제 6 장).

15.6 프로그램 인터페이스와 함께 제공되는 파일

전체 파이프라인은 일괄 처리용으로 내보내져 있다. `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` 가 각 단계를 조율하며, 이는 `remarks_to_columns_()`, `split_list_column_()`, `apply_user_rules_()`, `validate_rules_()` 위에 세워져 있다. `inst/prolog/` 아래에 함께 제공되는 Prolog 원본은 `mls_remarks.pl`(DCG 문법과 특성 집계), `mls_analyze.pl`(일괄 구문 분석을 위한 파일 입출력 구동부), `mls_rules.pl`(도출 규칙에서 쓸 수 있는 도우미 술어), 그리고 `climate_regions_ca.pl`(아래 참조) 이다.

15.7 관련 실험: 기후 지역 사전정보

이와 짝을 이루는 실험적 기능은 데이터가 얇은 특성, 즉 비교사례 몇 건에만 나타나서 (이를 테면 증발식 냉방기) 회귀분석이 추정할 수 없는 특성을 다룬다. 세 개의 R 도우미 함수 — `climate_region_for()`, `market_area_profile()`, `climate_feature_priors()` — 가 프로젝트의 위치 (카운티 + 도시) 를 매수자가 인지하는 캘리포니아의 일곱 개 기후 지역 (캘리포니아 에너지위원회의 16 개 건축 기후대를 실용적으로 축약한 것) 가운데 하나로 분류하고, 지역별로 HVAC 및 냉방 특성 (난방, 에어컨, 증발식 냉방기, 전관 환기팬, 여과 장치) 에 대한 서열형 기여 가치 사전정보를 돌려준다. Prolog 모듈 `climate_regions_ca.pl` 은 도출 규칙 안에서 쓸 수 있도록 같은 분류를 그대로 반영한다. 양쪽 모두 하나의 정보 R 원본에서 생성되므로 서로 어긋날 수 없다. 이 사전정보는 특성 평가에 참고가 되고 추정 계수의 타당성 점검 역할을 하도록 의도된 것이다. 지역 배정 (특히 여러 지역에 걸친 카운티의 경우) 은 감정평가사가 검토해야 한다.

15.7.1 캘리포니아를 넘어: 관할권 팩과 시장지역 모듈

캘리포니아는 첫 번째 관할권이지 한계가 아니다. `earthUI` 는 국제적으로 사용되고 있으며, 기후 사전정보 지식은 **관할권 팩**으로 재구성되고 있다. 관할권 팩이란 프로젝트 시스템의 국가 및 행정 수준 코드에 결부된, 자체 완결형이며 데이터 형태를 띤 사실 표 (지대 정의, 행정 단위 배정, 특성 사전정보) 로서, 하나의 팩이 미국의 주, 독일의 란트, 또는 작은 나라 하나를 통째로 기술할 수 있다. 이와 짝을 이루는 **시장지역 모듈** 층은 이름이 붙은 시장지역의 지역 어휘와 정교화된 사전정보를 담는다.

두 가지 모두 감정평가공학과 감정평가사가 현장에서 자신의 언어로 작성하도록, 그리고 현실적으로는 LLM 의 도움을 받아 작성하도록 설계되었다. 세상에 Prolog 프로그래머는 많지 않으므로, 현장 산출물은 의도적으로 프로그램이 아니라 표로 되어 있다. 팩 형식은 완전한 실제 예시 (캘리포니아 자체) 를 갖춘 문서화된 스키마인데, 이는 Claude 같은 조력자가 시장에 대한 평이한 서술로부터 유효한 팩을 초안할 수 있게 하려는 의도가 명시적으로 담긴 것이다. `earthUI` 는 팩으로부터 바탕이 되는 Prolog 를 생성하며, 팩을 활성화하려면 반드시 차단형 검증기 (스키마, 적용 범위, 참조 무결성, 예행 실행 검사) 를 통과해야 한다. 현장에서 작성된 모든 팩은 의무적인 출처 정보 — 기록상의 작성자, 날짜, 그리고 인용된 근거 — 를 지니므로, 초안 작성 도구가 아니라 책임 있는 전문가가 그 지식을 뒷받침하며 보고서는 사용된 팩과 그 버전을 정확히 인용할 수 있다.

팁: 실무적인 방식은 이렇다. 비고를 확장하고, 잘게 나뉜 `pr_*` 플래그를 `derive/2` 규칙으로 몇 개의 파생 개수나 불리언으로 통합한 뒤, 통합에 쓰인 원 플래그는 `drop/1` 으로 버리고, 파생된 열만 후보 예측변수로 남기는 것이다. 이렇게 하면 예측변수 목록이 짧아지고 다중공선성도 관리할 수 있는 수준으로 유지된다.

참고문헌

- 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>.

Milborrow, Stephen (n.d.[b]). *Variance Models in earth*. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.

glmnetUI: glmnet(Lasso 및 Elastic Net) 패키지를 위한 대화형 인터페이스 (한국어판)

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 검토했습니다. 영문판이正本입니다.

William Bert Craytor

제1권 제1호 · 2026년 여름호

발행일: 2026년 7월 1일

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

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

이 논문은 오픈 액세스 논문으로 [크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스\(CC BY 4.0\)](#)의 조건에 따라 배포됩니다.

초록. glmnetUI는 R의 glmnet 패키지를 위한 그래픽 사용자 인터페이스(graphical user interface)이며, glmnet은 Lasso, 릿지(ridge), 엘라스틱 넷(elastic net) 벌점(penalty)을 통해 정규화(regularization)된 일반화 선형모형(generalized linear model)을 적합한다. glmnetUI는 세 가지 용도 모드—범용 예측 모형화, 부동산 감정평가, 시장지역 분석(market-area analysis)—를 제공하며, 데이터 가져오기, 모형 구성, 교차검증(cross-validation) 적합, 계수 경로(coefficient path) 확인, 보고서 내려받기까지 사용자를 안내한다. 본 논문은 glmnetUI의 데이터 형식 요건, 모형화 작업 흐름, 출력 화면, 그리고 전체 기능 참조 사항을 기술한다.

키워드: Lasso, 릿지 회귀, 엘라스틱 넷, 정규화 회귀, glmnet, R, 그래픽 사용자 인터페이스, 부동산 감정평가

발행인 주(2026년 7월): 인쇄 직전에 glmnetUI는 earthUI(0.11.0 이후 버전)에 통합되었으며, 본 논문에서 설명하는 엘라스틱 넷 작업 흐름은 이제 MARS 및 GAM 작업 흐름과 함께 earthUI가 제공한다. glmnetUI는 최종 배포판(0.5.0) 형태로 <https://github.com/wcraytor/glmnetUI>에 계속 공개되어 있으며, 향후 개발은 earthUI에서 이어진다. 본 논문은 배포된 상태의 glmnetUI에 적용된다.

목차

1	서론	3
1.1	역사와 배경	3
1.1.1	Lasso	3
1.1.2	릿지 회귀와 엘라스틱 넷	4
1.1.3	glmnet 패키지	4
1.1.4	MARS(earth) 및 GAM(mgcv)과의 비교	4
1.1.5	권장 작업 흐름	5
1.2	glmnetUI란 무엇인가?	5
1.3	세 가지 용도 모드	6
1.4	부동산 전용 기능	6
1.5	시작하기	7
2	MLS 입력 데이터 요건	8
2.1	파일 형식과 구조	8
2.2	감정평가 모드에 필요한 열	8
2.3	특수 열 명명 규칙	9
2.4	데이터 품질과 완전성	9
2.5	대상 행의 위치	9

3	General Purpose 모드	10
3.1	개요	10
3.1.1	Skip First Row	10
3.1.2	용도별 설정	10
3.2	사이드바 작업 흐름	10
3.3	프로젝트	11
3.4	earthUI에서 가져오기 (선택 사항)	11
3.5	주 패널 탭	11
3.6	설정의 보존	12
3.7	다크 모드	12
3.8	로케일과 지역 설정	12
3.8.1	기본값 저장	13
4	Appraisal 모드	14
4.1	대상 행의 처리	14
4.2	기준일과 매매 경과일	14
4.3	특수 열 지정	14
4.4	RCA 보정 개요	14
5	Market Area Analysis 모드	15
5.1	Appraisal 모드와의 차이	15
5.2	시장 모드를 쓸 때	15
6	변수 선택	16
6.1	목표변수	16
6.2	설명변수 표	16
6.3	자료형 감지와 재지정	16
6.4	기대 부호와 적용	16
6.5	특수 열 유형 참조	17
7	모수 선택	18
7.1	Alpha(혼합 모수)	18
7.2	Lambda 선택	18
7.3	Family	18
7.4	Standardize	19
7.5	부호 제약	19
7.6	완화 Lasso	19
7.7	난수 시드	19
7.8	교호작용 행렬	19
7.9	주효과에서 차단(교호작용 전용)	20
7.10	고급 모수	20
7.11	설정 기본값	20
8	모형 적합	22
8.1	Fit 버튼	22
8.2	적합 중에 일어나는 일	22
9	결과 탭	23
9.1	Data Preview	23
9.2	Equation	23
9.3	Correlation	23
9.4	Summary	23

9.5	Coefficients	23
9.6	Variable Importance	24
9.7	Contributions	24
9.8	ANOVA	24
9.9	Diagnostics	24
9.10	Glmnet Output	25
9.11	Report	25
10	데이터 내려받기	26
10.1	출력 열	26
10.2	CQA 점수	26
11	RCA 계산과 내려받기	27
11.1	RCA 대화상자 열기	27
11.2	CQA 점수 보간	27
11.3	출력 열	27
12	거래사례비교표	28
12.1	비교 사례 선택 대화상자	28
12.2	통합문서 구성	28
12.3	시트 보호와 잔여 칸	28
13	보고서 내려받기	29
13.1	두 단계 Quarto 작업 흐름	29
13.2	보고서 형식	29
13.3	보고서 내용	29
14	earthUI와의 비교	31
14.1	주요 차이	31
14.2	공통 기능	31
14.3	어느 것을 언제 쓸 것인가	31
15	예제 데이터셋: Appraisal_1.csv	32
15.1	설명	32
15.2	열	32
15.3	권장 빠른 시작	32
16	시스템 요구사항과 문제 해결	34
16.1	지원 플랫폼	34
16.2	선택적 의존 요소	34
16.3	플랫폼별 참고사항	34
16.4	점진적 기능 축소	34
16.5	문제 해결	34

1 서론

1.1 역사와 배경

1.1.1 Lasso

Lasso(Least Absolute Shrinkage and Selection Operator)는 1996년 Robert Tibshirani가 제안하였다. 그 이름은 카우보이의 올가미(lasso)—구속하고 골라내는 밧줄—를 의도적으로 빗댄 말장

난이다. Tibshirani가 이 이름을 택한 것은, 이 방법이 문자 그대로 계수를 “올가미로 낚아채어” 절댓값을 제약하고 일부를 정확히 0으로 끌어당김으로써 어떤 변수가 모형에 남을지를 선택하기 때문이다.

Tibshirani의 착안은 최소제곱 목적함수에 L_1 벌점(penalty, 계수 절댓값의 합)을 더하는 것이었다. 계수를 0 쪽으로 축소하되 결코 0에 도달하지는 않는 릿지 회귀(ridge regression; Hoerl & Kennard, 1970)의 L_2 벌점과 달리, L_1 벌점은 많은 계수가 정확히 0이 되는 희소(sparse) 해를 만들어 낸다. 이 덕분에 Lasso는 정규화(regularization) 방법이면서 동시에 변수 선택 방법이기도 하다.

1.1.2 릿지 회귀와 엘라스틱 넷

릿지 회귀(ridge regression)는 1970년 Arthur Hoerl과 Robert Kennard가 제안하였다. 이 방법은 L_2 벌점(계수 제곱의 합)을 더하여 모든 계수를 비례적으로 축소하되 모든 변수를 모형에 남겨 둔다. 릿지는 설명변수들이 상관되어 있을 때(다중공선성) 효과적이지만 모형을 단순화하지는 않는다.

엘라스틱 넷(elastic net)은 2005년 Hui Zou와 Trevor Hastie가 제안한 것으로 두 벌점을 결합한다. 혼합 모수 α 가 그 배합을 조절하는데, $\alpha = 1$ 은 순수한 Lasso, $\alpha = 0$ 은 순수한 릿지이며, 그 사이의 값은 절충안을 준다. 엘라스틱 넷은 Lasso의 한계를 극복한다. 설명변수들이 서로 강하게 상관되어 있을 때 Lasso는 하나만 선택하고 나머지를 무시하는 경향이 있는 반면, 엘라스틱 넷은 상관된 설명변수들을 함께 묶어 준다.

1.1.3 glmnet 패키지

R 패키지 **glmnet**은 스탠퍼드 대학교의 Jerome Friedman, Trevor Hastie, Robert Tibshirani가 개발하였다. 이 패키지는 매우 효율적인 좌표하강(coordinate descent) 알고리즘을 사용하여 일반화 선형모형 (generalized linear model)에 대한 엘라스틱 넷 정규화를 구현하며 (Friedman, Hastie, and Tibshirani 2010), 이후 Cox 비례위험모형 (Simon et al. 2011)과 모든 일반화 선형모형 계열 (Tay, Narasimhan, and Hastie 2023)로 확장되었다. 이 패키지는 가우시안(선형), 이항(로지스틱), 포아송을 비롯한 여러 GLM 계열을 지원한다.

glmnet의 주요 기능은 다음과 같다.

- **정규화 경로(regularization path)**: 람다(lambda) 값의 격자 전체에 대해 한 번의 계산으로 모형을 적합하므로, 각 람다마다 따로 적합하는 것보다 훨씬 빠르다.
- **교차검증(cross-validation)**: `cv.glmnet()`이 k -겹 교차검증을 통해 최적의 정규화 강도를 자동으로 선택한다.
- **계수 한계(coefficient bounds)**: 개별 계수에 대한 상한과 하한 (부호 제약)을 지원한다.
- **완화 Lasso(relaxed lasso)**: 선택된 변수만으로 벌점을 줄이거나 없앤 상태에서 다시 적합하여 편향을 줄인다.

1.1.4 MARS(earth) 및 GAM(mgcv)과의 비교

earth 애플리케이션과 그 두 개의 후처리 모듈(Post Processing Module, **glmnet**과 **mgcv**)은 서로 다른 모형화 엔진을 사용한다. 다음 표는 주요 차이를 요약한 것이다.

항목	glmnet (엘라스틱 넷)	earth (MARS)	mgcv (GAM)
모형 유형	선형(정규화 포함)	조각별 선형(적응적 스플라인)	매끄러운 비선형(별점 스플라인)
비선형성	교호작용이나 기저 확장을 통해서만 가능	데이터 기반 매듭을 갖는 힌지 함수로 자동 처리	매끄러운 함수(박판, 3차 등)로 자동 처리
변수 선택	L_1 별점(Lasso)으로 내장	GCV 가지치기를 동반한 전진/후진 단계적 선택으로 내장	이중 별점 축소로 선택적 지원(select=TRUE)
교호작용	쌍별 교차곱(해석하기 어려움)	힌지 기반 곱(해석 가능, 3차까지)	텐서곱 평활(te(), ti())
해석 가능성	계수가 선형 가중치이므로 개별 설명이 쉬움	g-함수가 변수별 조각별 선형 부분효과를 보여 줌	매끄러운 부분효과 곡선으로 매우 직관적
적합한 상황	고차원 데이터, 변수 선택, 소표본	비선형성과 교호작용의 자동 탐지	매끄러운 비선형 관계, 유연한 모형화
RCA 보정	변수별 선형 보정	g-함수를 통한 조각별 선형 보정	부분효과를 통한 매끄러운 보정
과적합 제어	교차검증된 람다	GCV 기반 가지치기	REML/GCV 평활모수 추정

MARS(Multivariate Adaptive Regression Splines, 다변량 적응 회귀 스플라인)는 1991년 Jerome Friedman이 제안하였다. MARS는 힌지 함수 $\max(0, x - k)$ 와 그 매듭 위치를 데이터로부터 적응적으로 선택하여 조각별 선형모형을 구성한다. R 구현은 Stephen Milborrow의 **earth** 패키지이다. MARS는 사용자가 따로 지정하지 않아도 비선형 관계와 교호작용을 자동으로 찾아내는 데 탁월하다. 그 g-함수(변수별로 묶인 기저 항)는 매우 해석하기 쉬운 부분효과 곡선을 제공한다.

GAM(Generalized Additive Model, 일반화 가법모형)은 Hastie와 Tibshirani(1990)가 정식화하였다. R 구현인 Simon Wood의 **mgcv**는 REML 또는 GCV를 통한 자동 평활도 선택 기능을 갖춘 별점 회귀 스플라인을 제공한다. GAM은 시각화하고 설명하기에 직관적인 매끄러운 부분효과 곡선을 산출한다. **earthUI**가 매듭 위치를 **mgcvUI**로 내보내면, **earth**에서 도출된 매듭이 GAM 평활항의 출발점 역할을 하여 MARS의 적응적 매듭 배치와 GAM의 매끄러운 추정을 결합하게 된다.

1.1.5 권장 작업 흐름

해석 가능하고 방어 가능한 모형이 필요한 부동산 감정평가 및 이와 유사한 용도에서는 다음을 권장한다.

1. **earthUI**로 시작하여 데이터에서 중요한 변수, 비선형 관계, 교호작용을 발견한다. **earth**의 자동 기저 선택과 g-함수는 훌륭한 초기 통찰을 제공한다.
2. 더 매끄러운 부분효과를 원한다면 **mgcvUI**로 정교화한다. **earthUI**의 매듭을 **mgcvUI**로 가져오면 조각별 선형모형에서 매끄러운 모형으로 매끄럽게 넘어갈 수 있다.
3. 공격적인 변수 선택(Lasso)이 필요할 때, 설명변수 개수가 관측치 개수에 근접하거나 이를 넘어설 때, 또는 방어 가능성을 위해 정규화된 순수 선형모형을 원할 때 **glmnetUI**를 사용한다.

1.2 glmnetUI란 무엇인가?

glmnetUI는 R **glmnet** 패키지를 위한 그래픽 사용자 인터페이스이다. 로컬 Shiny 애플리케이션으로 실행되며—로그인도, 서버도, 계정도 없다. R에서 실행한 뒤 데이터셋(CSV 또는 Excel)을 가져오고, 모형을 구성하고, 대화형으로 적합한다.

현재 배포판 기준으로 glmnetUI는 더 이상 별도로 설치하는 패키지가 아니다. 이제 earthUI 패키지 내부에 엘라스틱 넷 후처리 모듈(Post Processing Module)로 포함되어 `earthUI::launch_glmnet()`으로 시작한다. 애플리케이션 자체—그리고 본 논문에서 설명하는 모든 내용—는 변함이 없으며, 설치 방법과 실행 명령만 달라졌다. 세 루틴(earth, glmnet, mgcv)은 하나의 프로젝트 트리, 하나의 설정 데이터베이스, 하나의 지원 인프라를 공유한다.

이 애플리케이션은 데이터 가져오기, 변수 구성, 모형 적합, 진단 그림, 변수 중요도, 모형 방정식, 그리고 Word, PDF, HTML 형식으로 내려받을 수 있는 보고서에 이르는 완결된 작업 흐름을 제공한다.

엘라스틱 넷 회귀는 두 가지 정규화 기법을 결합한다.

- **릿지 회귀**($\alpha = 0$): 모든 계수를 0 쪽으로 비례적으로 축소한다. 모든 변수가 유의미하다고 볼 때 적합하다.
- **Lasso 회귀**($\alpha = 1$): 계수를 정확히 0으로 축소할 수 있어 자동 변수 선택을 수행한다. 더 단순하고 해석하기 쉬운 모형을 만든다.
- **엘라스틱 넷**($0 < \alpha < 1$): 두 방식의 혼합이다. 설명변수들이 상관되어 있을 때 유용하다.

1.3 세 가지 용도 모드

모든 glmnetUI 프로젝트는 생성 시 선택하는 **용도(purpose)**를 갖는다 (제1절, *Project*). 활성 프로젝트의 용도에 따라 어떤 도구와 인터페이스 요소가 나타날지가 결정되므로, 다른 모드로 작업하려면 해당 용도를 가진 프로젝트를 새로 만들거나 열어야 한다. 세 가지 모드는 다음과 같다.

- **General**(범용) — 모든 유형의 모집단이나 데이터셋에 대한 엘라스틱 넷 회귀이다. 기본 모드이며, 도메인 특화 기능 없이 정규화 회귀의 전 과정을 제공한다.
- **For Appraisal**(감정평가용) — 부동산 감정평가에 맞춘 엘라스틱 넷 회귀이다. 대상 부동산 처리, 특수 열 지정, 잔차 제약 접근법 (Residual Constraint Approach, RCA) 등 단일 부동산 평가에 특화된 기능을 추가한다.
- **Market Area Analysis**(시장지역 분석) — 시장지역 연구에 맞춘 엘라스틱 넷 회귀이다. 정의된 시장 내 부동산 집단을 분석하는 기능을 추가하며, 대상 행(subject row) 제외를 선택적으로 지원한다.

세 모드 모두에서 핵심 모형화 엔진은 동일하다—언제나 엘라스틱 넷 (glmnet) 모형을 적합하는 것이다. 용도 설정은 어떤 추가 도구와 인터페이스 요소를 사용할 수 있는지를 제어할 뿐이다.

1.4 부동산 전용 기능

For Appraisal 또는 **Market Area Analysis**를 선택하면 glmnetUI는 부동산 분석을 위해 설계된 여러 기능을 활성화한다.

- **특수 열 지정(special column designations)** — 각 설명변수에 `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions`, `display_only`와 같은 특수 역할을 지정할 수 있다. 이 지정에 따라 적합과 출력 과정에서 해당 열이 어떻게 처리될지가 결정된다.
- **위도와 경도의 반올림** — 위도 또는 경도로 지정된 열은 과적합을 막기 위해 자동으로 소수점 셋째 자리까지 반올림된다.
- **Sale Age 열** — 어떤 열이 `contract_date`로 지정되고 기준일(Effective Date)이 주어지면 glmnetUI는 `sale_age` 열(매매일과 기준일 사이의 일수)을 계산하여 설명변수로 대체 투입한다.
- **RCA 계산(감정평가 모드 전용)** — 감정평가 모드에서는 모형 적합 후 glmnetUI가 잔차 제약 접근법(Residual Constraint Approach, RCA) 출력을 계산할 수 있으며, 이는 사례별 보정치, 순/총 보정 요약, 그리고 대상 부동산의 보정된 매매가격을 산출한다.

요령: *General* 용도 모드는 금융, 과학, 공학, 부동산 등 어떤 데이터셋에도 적용된다. 감정평가 모드와 시장 모드는 부동산 실무자를 위한 편의 기능을 더할 뿐이다.

1.5 시작하기

glmnetUI를 사용하려면 다음과 같이 한다.

1. R에서 **earthUI**를 설치한다: `install.packages("earthUI")` — glmnet 루틴이 함께 포함되어 있다. (개발 버전과 선택 구성요소에 대해서는 이번 호 earthUI 논문의 “Installing earthUI and vProlog” 절을 참조하라.)
2. 애플리케이션을 실행한다: R에서 `earthUI::launch_glmnet()`을 실행한다. 앱은 웹 브라우저의 7879 포트에서 열린다. 브라우저에서 `http://localhost:7879`로 직접 접속할 수도 있다. 앱은 마지막에 사용한 용도 모드를 기억하여 자동으로 복원한다.
3. 사이드바 제1절에서 프로젝트를 생성하거나 연다. 프로젝트는 용도와 작업 위치(국가, 주, 카운티, 시)를 설정하며 입력/출력 폴더를 갖는다. 프로젝트는 공유되는 `regProj` 루트 아래에 저장되며 형제 앱인 earthUI 및 mgcvUI와 공유된다. 다른 용도를 가진 프로젝트를 여는 것이 모드를 전환하는 방법이다.
4. 제2절에서 활성 프로젝트의 입력 폴더에 있는 CSV 또는 Excel 파일을 선택하여 데이터를 가져온다. 해당 폴더에 파일을 넣은 뒤 **Refresh**를 눌러 목록을 갱신한다.
5. earthUI에서 가져오기(선택 사항) — 제3절에서는 earthUI 결과 `.rds` 파일을 가져와 earth의 `hinz` 기저 함수를 glmnet의 정규화와 함께 사용할 수 있다. earthUI 결과가 없다면 이 단계는 건너뛰는다.
6. 변수를 구성한다 — 목표변수와 설명변수를 고르고, 자료형과 기대 부호를 설정하며, 필요한 특수 열 역할을 지정한다.
7. glmnet 모수를 설정한다 — `alpha`, `lambda`, `family`, 부호 제약, 교호작용 및 기타 옵션.
8. 모형을 적합한다 — “Fit Glmnet Model”을 클릭하고 주 패널에서 결과를 검토한다.
9. 내보내기 — 예측값을 Excel로 내려받거나, Quarto 보고서를 생성·변환하거나, (감정평가 모드에서는) RCA 보정치와 거래사례비교표 (Sales Comparison Grid)를 계산한다.

설정은 브라우저의 로컬 저장소에 자동으로 보존되며 동일한 입력 파일을 다시 불러올 때 복원된다.

2 MLS 입력 데이터 요건

부동산 감정평가와 시장분석 작업 흐름에서 입력 데이터는 일반적으로 다중 매물정보 서비스 (Multiple Listing Service, MLS) 내보내기 파일에서 온다. 이 장에서는 기대되는 파일 구조와 glmnetUI가 사용할 수 있는 열을 설명한다.

2.1 파일 형식과 구조

glmnetUI는 CSV와 Excel(.xlsx, .xls) 파일을 받는다. 가져오기 시 열 이름은 자동으로 snake_case로 변환된다—예를 들어 “Living SqFt”는 living_sqft로, “Contract Date”는 contract_date로, “Sale Price”는 sale_price로 바뀐다. 이 정규화 덕분에 작업 흐름 전반에서 열 참조가 일관되게 유지된다. 가져오기에 사용되는 CSV 구분자와 소수점 기호는 로케일 설정에 따라 결정된다(제3장 “Locale & Regional Settings” 참조).

데이터 파일은 부동산 한 건당 한 행, 속성 하나당 한 열로 구성된 평평한 표여야 한다. 파일의 첫 행에는 반드시 열 머리글이 있어야 한다.

2.2 감정평가 모드에 필요한 열

glmnetUI는 어떤 열 구성에서도 동작하지만, 완전한 감정평가 작업 흐름을 위해서는 다음 열들이 있는 편이 좋다.

열	특수 유형	용도
Sale Price	(목표변수)	모형의 반응변수
Contract Date	contract_date	sale_age(기준일로부터의 일수) 계산에 사용
Listing Date	listing_date	DOM 열이 없을 때 계약일과 함께 DOM 계산에 사용
Days on Market	dom	시장 노출 일수; 내보내기 파일에 표시
Concessions	concessions	매매 양허; 순매매가 = 매매가격 - 양허
Living Area (SF)	living_area	제곱피트당 잔차(residual_sf, cqa_sf) 사용 가능
Lot Size	lot_size	대지 규모 열
Site Dimensions	site_dimensions	대지 규모와 함께 묶임
Latitude	latitude	소수점 셋째 자리 반올림
Longitude	longitude	소수점 셋째 자리 반올림
Area ID	area	시장지역/근린 식별자
Actual Age	actual_age	부동산 실제 연령
Effective Age	effective_age	부동산 유효 연령
Address	display_only	내보내기에는 표시하되 모형에서는 제외

스프레드시트의 열 이름은 외국어여도 무방하다—R 프로그램이 특별하게 처리할 수 있도록 “특수” 이름만 영어로 되어 있을 뿐이다. 그 외에는 주어진 열 이름이 회귀모형, 그래프, 그리고 감정평가를 수행하는 경우 출력 보고서에 그대로 나타난다.

모든 열이 필수는 아니다. glmnetUI는 상황에 맞게 적응한다—어떤 열이 없으면 그에 대응하는 기능이 단순히 생략된다. 다만 부동산 가격 모형에서 수용 가능한 적합도를 얻으려면 다음 열들이 강력히 권장된다.

1. **Sale Age**(매매 경과일) — 계약 매매일과 감정평가 또는 분석의 기준일 사이의 일수이다. 여러 해에 걸친 매매 이력을 사용하는 경우, 특히 5년이 넘는 기간이라면 sale_age가 매매 가격 추정에서 중심적인 역할을 하는 경우가 많다.
2. **Living Area**(주거면적) — “Living Sqft”, “GLA”(총 주거면적) 등으로도 불린다. 이 역시 매매가격의 주된 결정요인이다.
3. **Total Bath Count**(총 욕실 수) — 전용, 1/4, 1/2, 3/4 욕실을 모두 합한 수이다. 예를 들어 전용 욕실 두 개와 반욕실 하나는 2.5가 된다.

4. **Garage Bays**(차고 칸 수) 또는 **Garage Area**(차고 면적) — 차고 주차 공간의 수 또는 차고 면적(제곱피트).
5. **Lot Size**(대지 규모) — 부동산의 토지 면적으로, 보통 제곱피트나 에이커 단위이다.
6. **Longitude, Latitude**, 그리고 가능하다면 **Area ID**. 이 위치 변수들은 모형이 지리적 가격 변동을 반영하도록 돕는다.

2.3 특수 열 명명 규칙

glmnetUI는 열 이름이 아니라 특수 유형 지정(**special type designation**)으로 열을 식별한다. MLS 내보내기 파일에서 열 이름은 원하는 대로 붙여도 되며, 중요한 것은 변수 구성표(제6장)에서 올바른 특수 유형을 지정하는 것이다.

예를 들어 사용자의 MLS는 주거면적을 “GLA”, “Living SqFt”, “liv_area”, “gross_living_area” 등으로 내보낼 수 있다. 가져오기(이때 snake_case로 바뀐다) 후에 Special 드롭다운에서 living_area로 지정하기만 하면 된다. 그러면 glmnetUI는 원래 이름과 무관하게 그 열을 제곱피트당 잔차 계산에 사용한다.

2.4 데이터 품질과 완전성

- **결측값(NA)**: 설명변수나 목표변수 열에 NA가 있는 행은 적합 전에 자동으로 제거된다.
- **날짜 열**: 흔한 날짜 형식과 일치하는 문자열 열은 가져오기 시 자동으로 해석된다. 네 자리 연도가 감지되지 않으면 두 자리 연도 형식이 우선 적용된다.
- **수치형 열**: 매매가격, 주거면적, 대지 규모, 양허 등과 같은 항목은 반드시 수치형이어야 한다. MLS가 통화 기호나 천 단위 구분기호가 포함된 가격을 내보낸다면(예: “\$350,000” 또는 “350.000,00”) 가져오기 전에 정리해야 할 수 있다.
- **요인형 열**: 텍스트, TRUE/FALSE, R 요인(factor) 열은 자동으로 범주형으로 취급된다—회귀분석은 이들을 요인으로만 다룰 수 있기 때문이다. 수치형 열(예: 수치 코드로 된 지역이나 양식 코드)은 **Factor** 체크박스를 선택할 때에만 요인이 된다. glmnetUI는 수치의 값 범위만 보고 범주형이라고 추론하지 않으므로 bath_count 같은 이산 수치형 설명변수는 연속형으로 남는다. Special 역할은 요인 여부와는 무관하다.

요령: 가져오기 후 설명변수 표의 NA 열을 검토하라. 결측값이 많은 열은 행이 통째로 버려지는 원인이 될 수 있다. NA가 많은 열은 제외하거나 가져오기 전에 데이터를 정리하는 것을 고려하라.

2.5 대상 행의 위치

Appraisal(감정평가) 모드에서는 1행이 반드시 대상 부동산이어야 한다. 나머지 행은 모두 비교 거래사례이다. 대상 행은 모형 적합에서 제외된다. 적합 후에도 모형은 대상 행에 대한 예측값을 생성한다.

Market Area Analysis(시장지역 분석) 모드에서는 대상을 1행에 두는 것이 선택 사항이다.

General(범용) 모드에는 특별한 행 처리가 없다—모든 행이 동등하게 취급된다.

3 General Purpose 모드

3.1 개요

General Purpose(범용) 모드는 glmnetUI를 실행했을 때의 기본값이다. 이 모드는 부동산에 국한되지 않고 어떤 데이터셋에 대해서도 완전한 엘라스틱 넷 회귀 작업 흐름을 제공한다. 과학 데이터, 재무 분석, 공학 연구, 또는 어떤 회귀 문제에도 glmnetUI를 사용할 수 있다.

General 모드에서는 부동산 전용 기능(특수 열, 매매 경과일, 좌표 반올림, RCA)이 인터페이스에서 생략된다. 사이드바는 변수 선택, 모수 구성, 모형 적합, 내보내기에 집중하도록 간소화된다.

3.1.1 Skip First Row

General 또는 Market 프로젝트가 열려 있으면 사이드바 맨 위에 **Skip first row**(첫 행 건너뛰기) 체크박스가 나타난다. 이를 선택하면 1행이 모형 적합에서 제외된다. 1행에 모형에 영향을 주어서는 안 되는 대상 관측치나 기준 관측치가 들어 있을 때 유용하다. Appraisal 모드에서는 1행(대상 부동산)이 항상 자동으로 제외된다.

3.1.2 용도별 설정

설정(설명변수 선택, 모형 모수, 교호작용)은 입력 파일과 용도 모드의 조합마다 따로 저장된다. General, Appraisal, Market 모드를 오가더라도 각 모드의 설정은 독립적으로 유지된다.

3.2 사이드바 작업 흐름

사이드바는 번호가 매겨진 접이식 절들로 구성되어, 프로젝트 선택부터 보고서 내보내기까지 사용자를 안내한다. 대부분의 절은 활성 프로젝트가 열려 있어야 나타나며, 몇몇 절은 프로젝트의 용도에 따라서만 나타난다. 출력 폴더는 활성 프로젝트로부터 도출되므로 별도의 출력 폴더 입력란은 없다.

1. Project — 프로젝트를 생성하거나 연다. 프로젝트는 용도와 작업 위치(국가, 주, 카운티, 시), 그리고 입력 (<os>_in/) 및 출력 폴더를 설정한다. 프로젝트는 공유되는 regProj 루트(기본값 ~/regProj, REGPROJ_ROOT 환경변수로 변경 가능) 아래에 저장되며 형제 앱인 earthUI 및 mgcvUI와 공유되므로 세 앱의 모형이 나란히 놓인다. 전환하려면 **Close Project**를 클릭한다.

2. Import Data — 활성 프로젝트의 입력 폴더에서 CSV 또는 Excel 파일을 선택한다. Finder/탐색기로 폴더에 파일을 넣은 뒤 **Refresh**를 눌러 목록을 갱신한다. 시트가 여러 개인 Excel 파일은 시트 선택기가 나타난다. 열 이름은 자동으로 snake_case로 변환된다.

3. Import from earthUI (optional) — earthUI 결과 .rds 파일을 가져와 earth의 힌지 기저 함수를 glmnet의 엘라스틱 넷 정규화와 함께 사용한다. 찾아보기 버튼은 제2절의 파일 선택기와 같은 모양으로 꾸며져 있다. 이 단계는 선택 사항이며—가져올 earthUI 결과가 없다면 건너뛴다.

4. Variable Configuration — 목표변수 선택기, Include, Factor, Force 체크박스과 Sign 드롭다운이 있는 설명변수 표. 계약일, 좌표 등을 지정하는 Special 열은 Appraisal 및 Market 모드에서만 나타난다. 자세한 내용은 제6장을 참조하라.

5. glmnet Call Parameters — 모형 구성 전체: alpha, lambda, family, standardize, 부호 제약, 완화 lasso, 난수 시드, 교호작용 행렬, 그리고 고급 모수(lambda.min.ratio, nlambda, CV 손실 지표, 수렴 임계값, 최대 반복 횟수, 절편). 전체 모수 참조는 제7장에 있다.

6. Fit Glmnet Model — 모형을 실행하는 버튼이다. 결과는 오른쪽 탭에 나타난다.

7. Download Output — 예측값, 잔차, CQA 점수, 변수별 기여도를 Excel 파일로 프로젝트의 출력 폴더에 내보낸다. 모형이 적합되고 나면 모든 용도 모드에서 사용할 수 있다.

8. Calculate RCA Adjustments & Download — 각 비교 사례에 대한 변수별 RCA 보정치를 계산하고 CQA를 통해 대상의 잔차를 보간한다. Appraisal 및 Market 모드에서 적합 이후에만 나타난다. 제11장을 참조하라.

9. Generate Sales Grid & Download — 총 보정 비율로 비교 사례를 순위화한 중간 단계의 거래사례비교표(Sales Comparison Grid)를 생성한다. Appraisal 모드에서 RCA 보정치가 계산된 뒤에만 나타난다. 제12장을 참조하라.

10. Generate Quarto Report — 모든 그림과 표를 포함한 자기완결적 Quarto(.qmd) 보고서 묶음을 프로젝트의 출력 폴더에 기록한다. 모형이 적합되고 나면 사용할 수 있다.

11. Convert Quarto Report — 생성된 .qmd 묶음을 HTML, Word, PDF로 렌더링한다. PDF에는 LaTeX 설치가 필요하며, 설치가 감지되지 않으면 PDF 옵션은 숨겨진다. 제13장을 참조하라.

Market 및 General 모드에서는 감정평가 전용 단계들이 숨겨져 건너뛰어지고 두 개의 보고서 절 번호가 그에 맞게 다시 매겨진다.

3.3 프로젝트

glmnetUI는 작업을 공유되는 regProj 루트 폴더(earthUI 및 mgcvUI가 사용하는 것과 동일한 트리) 아래의 프로젝트로 조직한다. 각 프로젝트는 입력 폴더(<os>_in)와 엔진별 출력 폴더(예: <os>_out_glmnet)를 갖는다.

프로젝트를 만들려면 제1절에서 + **New...**를 클릭하고 다음을 입력한다.

- **Purpose**(용도) — General, Appraisal, Market Area 중 하나.
- **Location**(위치) — 프로젝트의 행정 구역 체계. 선택한 국가에 따라 채워야 할 단계가 달라지므로 외국의 명명 관행이 자동으로 지원된다—미국 프로젝트는 주/카운티/시를, 프랑스 프로젝트는 지역권/도(département)/코뮌을, 일본 프로젝트는 도도부현/시구/지구를 사용하는 식이다.
- **Project Name**(프로젝트 이름) — 8자 이내의 짧은 이름 (영문자, 숫자, _, -).

앱은 입력한 이름을 그대로 쓰지 않는다. 위치 코드와 생성 일시를 앞에 붙여 고유한 시각 표시 폴더 이름을 <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name> 형식으로 만든다—예를 들어 burl이라는 이름의 미국 프로젝트는 us_ca_081_burlin_20260604-131500_burl이 된다. 8자 제한은 사용자가 입력하는 부분에만 적용된다. 시각 접두어는 새로 생성한 프로젝트에만 붙으며, 이미 디스크에 있는 프로젝트는 원래 이름을 유지한 채 그대로 완전히 사용할 수 있다.

3.4 earthUI에서 가져오기 (선택 사항)

사이드바 제3절 — **Import from earthUI** — 에서는 earthUI 결과 .rds 파일을 가져올 수 있다. 이를 통해 glmnetUI는 earth의 힌지 기저 함수(조각별 선형 항)를 glmnet의 엘라스틱 넷 정규화와 함께 사용할 수 있으며, earth의 자동 비선형성 탐지와 glmnet의 변수 선택 및 계수 제약이 결합된다.

earthUI 결과를 가져오면 Include 체크박스가 아니라 earth 모형의 설명변수가 glmnet 적합을 주도한다. glmnet이 earth의 기저 확장 위에서 적합하기 때문이다. 변수 구성표는 이를 명시적으로 보여 준다. **Include**와 **Type**은 잠기고, earth 모형이 사용하지 않은 설명변수의 행은 비활성화되며, 설명변수 집합이 earth 모형에서 왔음을 알리는 안내문이 표시된다. **Force**와 **Sign**은 earth의 설명변수에 대해 계속 편집할 수 있는데, 적합 시점에 earth 기저에 여전히 적용되기 때문이다.

찾아보기 버튼은 제2절의 파일 선택기와 같은 모양으로 꾸며져 있다. 이 단계는 전적으로 선택 사항이다—가져올 earthUI 결과 파일이 없다면 제3절을 건너뛰고 제4절로 진행하면 된다.

3.5 주 패널 탭

적합이 끝나면 주 패널에 다음 탭들이 나타난다.

탭	내용
Data Preview	가져온 데이터의 미리보기. 감정평가 모드에서는 대상 부동산 표와 비교 거래사례 표로 나뉜다.
Equation	적합된 모형 방정식을 LaTeX/MathJax로 표시하며, 0이 아닌 각 계수와 교호작용 항을 보여 준다.
Correlation	수치형 설명변수 상관관계수의 히트맵(적합 전에도 사용 가능).
Summary	주요 지표(R^2 , 조정 R^2 , GR^2 , CV R^2 , RMSE, MAE)와 감정평가 지표(COD, PRD, 중위 비율).
Coefficients	부호 위반 경고가 표시된, 0이 아닌 계수 표.
Variable Importance	설명변수 중요도($ b \times sd(x)$)의 막대그래프와 순위 표. 가능한 경우 plotly 대화형 호버 지원.
Contributions	기울기 표시가 있는 변수별 부분효과 그림. 수치형 설명변수는 산점도와 적합선, 요인형은 상자그림, 교호작용은 3차원 곡면/산점도/히트맵으로 표시.
ANOVA	분산 분해 표: 변수별 제곱합, 모형 제곱합 대비%, 계수.
Diagnostics	네 가지 그림: 계수 경로, CV 오차, 실제값 대 예측값, 잔차 대 적합값.
Glmnet Output	원시 모형 출력: 사용된 난수 시드, 모형 출력, 선택된 람다, $\lambda_{min}/1se$, 감마(완화 적합인 경우), 전체 계수 벡터.
Report	보고서 구성 및 내보내기 입력란.
RCA Adjustments	RCA 분석 결과(감정평가 모드에서 계산 후).

3.6 설정의 보존

glmnetUI는 사용자의 구성을 입력 파일 이름을 키로 하여 브라우저의 로컬 저장소에 자동으로 저장한다. 같은 파일을 다시 불러오면 목표변수 선택, 설명변수 체크박스, 자료형, 기대 부호, glmnet 모수, 교호작용 행렬 등 모든 설정이 복원된다. 마지막에 사용한 용도 모드도 전역적으로 보존되어 앱을 다시 실행할 때 복원된다. 세 가지 옵션이 있다.

- **Use last settings for input file** — 이 특정 파일에 대해 저장된 설정을 복원
- **Use default settings** — 저장해 둔 전역 기본값을 적용
- **glmnet defaults** — 공장 기본값으로 초기화($\alpha = 1$, CV $\lambda_{min}/1se$, 10겹, gaussian, standardize 켜)

Save current as default 버튼은 현재의 모든 설정을 전역 기본값으로 저장한다.

3.7 다크 모드

오른쪽 위 모서리의 달/해 아이콘을 클릭하면 밝은 테마와 어두운 테마를 전환할 수 있다. 테마 설정은 로컬 저장소에 저장되어 세션이 바뀌어도 유지된다. 모든 UI 요소(표, 그림, 카드, 버튼)는 CSS 변수로 선택된 테마에 맞추어 조정된다.

3.8 로케일과 지역 설정

glmnetUI는 국가 기반 로케일 체계를 통해 국제적인 숫자 및 CSV 형식 관행을 지원한다. 제목 표 시줄의 **Settings** 드롭다운은 지원되는 31개 국가에 대해 **Country**와 **Paper** 선택기를 제공한다. 각 프리셋은 다음을 설정한다.

- **CSV 구분자** — 미국/영국/일본은 쉼표(,), 쉼표를 소수점 기호로 쓰는 유럽 대부분은 세미콜론(;).
- **소수점 기호** — 마침표(.) 또는 쉼표(,).
- **천 단위 구분기호** — 쉼표(미국/영국/일본), 마침표(독일/이탈리아/스페인), 공백(핀란드/프랑스/폴란드/발트 3국/우크라이나/러시아), 아포스트로피(스위스).
- **용지 크기** — Letter(미국/캐나다/멕시코) 또는 A4(그 밖의 모든 지역).

3.8.1 기본값 저장

Save as my default를 클릭하면 로케일 설정이 전역으로 저장된다. 이 기본값은 어떤 데이터 파일을 불러오든 이후의 모든 세션에 적용된다.

요령: 그림 축과 기울기 표시의 숫자 형식은 로케일에 맞추어 자동으로 조정된다. 독일 로케일은 천 단위에 마침표(200.000)를, 핀란드는 공백(200 000)을, 스위스는 아포스트로피(200'000)를 사용한다. 통화 기호는 표시되지 않는다—*glmnetUI*는 통화에 중립적이다.

4 Appraisal 모드

용도로 **For Appraisal**을 선택하면 glmnetUI는 단일 부동산 평가에 맞추어 스스로를 구성한다. 제3장에서 설명한 모든 기능은 그대로 사용할 수 있으며, 이 장에서는 감정평가에 특화된 추가 사항만 다룬다.

4.1 대상 행의 처리

감정평가 모드에서는 데이터셋의 **1행이 대상 부동산**이고 나머지 행은 모두 비교 거래사례이다. 입력 파일은 그에 맞게 구성되어야 한다 (제2장 참조). 대상의 매매가격은 비워 두거나 아무 값이나 넣어도 된다—glmnetUI가 적합 시 자동으로 NA로 취급한다.

가져오기 후 Data Preview 탭은 두 부분으로 나뉜다. **Subject Property**(1행)와 **Comparable Sales**(2행 이하)이다. 1행은 언제나 모형 적합에서 제외된다. 적합 후에도 모형은 대상 행에 대한 예측값을 생성한다.

4.2 기준일과 매매 경과일

감정평가 모드와 시장 모드에서는 변수 구성 절에 **Effective Date**(기준일) 입력란이 나타난다 (기본값은 오늘 날짜). Special 열 드롭다운에서 어떤 열을 **contract_date**로 지정하면 glmnetUI는 **sale_age** 열—각 매매의 계약일과 기준일 사이의 정수 일수—을 계산한다. 이 열은 설명변수로 추가된다.

기준일이 바뀌면 **sale_age**는 자동으로 다시 계산된다.

4.3 특수 열 지정

감정평가 모드와 시장 모드에서는 변수 구성표의 각 설명변수마다 **Special** 드롭다운이 나타난다. 특수 유형과 그 효과의 전체 참조는 제6장에 있다.

4.4 RCA 보정 개요

Calculate RCA Adjustments & Download 버튼(사이드바 제8절, 적합 후 Appraisal 및 Market 모드에서 표시)은 대상 대비 각 비교 사례의 시장 도출 보정치를 계산한다. RCA 작업 흐름 전체는 제11장에서 설명한다.

요령: 대상에 부여하는 **CQA** 점수는 잔차 분포 가운데 얼마만큼을 대상에 귀속시킬지를 결정한다. 5.00점은 대상을 비교 사례들의 중앙값에 놓는다.

5 Market Area Analysis 모드

용도로 **Market Area Analysis**를 선택하면 glmnetUI는 감정평가 모드와 동일한 부동산 전용 기능(특수 열, 매매 경과일, 좌표 반올림)을 제공하되, 단일 대상의 가치를 평가하기보다 부동산 집단을 분석하는 데 초점을 맞춘다.

5.1 Appraisal 모드와의 차이

- 대상 행 처리가 유연하다 — 시장 모드에서는 1행이 자동으로 대상 부동산으로 취급되지 않는다.
- 거래사례비교표가 없다 — “Calculate RCA Adjustments & Download” 단계(제8절)는 시장 모드에서도 사용할 수 있지만, Sales Grid 단계(제9절)는 감정평가 모드 전용이다.

5.2 시장 모드를 쓸 때

Market Area Analysis 모드는 다음과 같은 경우에 적합하다.

- 가치 결정요인을 이해하기 위해 근린이나 시장지역에 대한 회귀모형을 구축할 때
- 면적, 연령, 대지 규모, 위치 같은 변수가 부동산 집단 전반의 매매가격에 어떻게 영향을 미치는지 분석할 때
- 시장 상황 분석이나 근린 확정에 대한 근거를 준비할 때
- 특수 열 기능(매매 경과일, 좌표 반올림)은 원하지만 RCA 보정 작업 흐름은 필요하지 않은 데이터셋을 다룰 때

요령: 시장 모드는 특수 열 기능은 원하지만 RCA 보정 작업 흐름은 필요하지 않은 일반적인 부동산 회귀분석에도 유용하다.

6 변수 선택

사이드바 제4절 — **Variable Configuration** — 은 어떤 열을 모형에 참여시키고 그것을 어떻게 다룰지 정하는 곳이다.

6.1 목표변수

제4절 맨 위의 **Target (response) variable** 드롭다운에는 데이터셋의 모든 열이 나열된다. 한 열을 반응변수(예: 매매가격)로 선택한다. 목표변수 열은 자동으로 설명변수 목록에서 제외된다.

6.2 설명변수 표

목표변수 선택기 아래에는 남은 모든 열이 다음 항목과 함께 표로 나열된다.

열	설명
Variable	열 이름
Type	자료형 드롭다운: numeric, integer, character, Date, POSIXct
Include	체크박스 — 이 열을 모형의 설명변수로 포함
Factor	체크박스 — 변수를 범주형 요인으로 취급(모형 행렬에 가변수 열 생성)
Force	체크박스 — 모형에 강제 투입(벌점 계수 = 0, lasso가 걸코 제거하지 않음)
Special	드롭다운(감정평가/시장 모드 전용) — 아래의 특수 열 유형 참조를 보라
Sign	기대되는 계수 부호: 양, 음, 또는 무관
NAs	결측값의 개수

6.3 자료형 감지와 재지정

glmnetUI는 가져오기 시 자료형을 자동으로 감지한다. 수치형, 정수형, 날짜형 열이 인식된다. 날짜처럼 보이는 문자형 열(흔한 날짜 형식 패턴)은 **Date**로 분류된다.

Type 드롭다운을 바꾸어 어떤 감지 결과든 재지정할 수 있다. 자료형을 바꾸면 해당 열이 모형 행렬에 어떻게 부호화되는지가 달라진다.

텍스트, TRUE/FALSE, R 요인 열은 자동으로 범주형으로 취급된다. 수치형 열은 **Factor** 체크박스를 선택할 때에만 요인이 된다—glmnetUI는 수치의 값 범위만 보고 범주형이라고 추론하지 않으므로, 욕실 수 같은 이산 수치형 설명변수는 사용자가 달리 지정하지 않는 한 연속형으로 남는다. Factor 설정은 localStorage를 통해 유지된다.

6.4 기대 부호와 적용

각 설명변수에 대해 **Sign** 드롭다운은 그 계수가 양이 될 것으로 기대하는지, 음이 될 것으로 기대하는지, 아니면 어느 쪽이든 무방한지를 지정한다. 적합 후에는 다음과 같이 처리된다.

- 기대와 다른 부호를 가진 계수는 **Coefficients** 표에 표시된다
- 부호 경고가 Summary 탭에 나타난다

glmnet 모수(제7장)에서 **Enforce Sign Constraints** 체크박스를 활성화하면 모형이 기대 부호에 부합하는 계수를 산출하도록 강제된다. 이는 glmnet의 upper.limits와 lower.limits로 구현된다.

6.5 특수 열 유형 참조

감정평가 모드와 시장 모드에서 **Special** 드롭다운은 다음 옵션을 제공한다.

가중치:

- **weight** — 관측치 가중치 열(모든 모드에서 사용 가능; 하나만 허용; 가중치가 0인 행은 적합에서 제외)

날짜 및 시간 유형:

- **contract_date** — 기준일로부터 **sale_age**를 자동 계산하도록 함
- **listing_date** — 시장 노출 일수 계산의 대체 근거로 사용
- **dom** — 시장 노출 일수 열을 식별
- **sale_age** — 기존의 매매 경과일 열을 지정하므로, glmnetUI가 **contract_date**로부터 계산하는 대신 임의의 열을 매매 경과일로 쓸 수 있다

금액 유형:

- **concessions** — 매매 양허(매도인 신용, 매수인 유인책 등)를 식별

거래 유형:

- **sale_type** — 매매/거래 유형 열을 식별

규모 및 위치 유형:

- **latitude** — 값이 자동으로 소수점 셋째 자리까지 반올림됨
- **longitude** — 위도와 동일한 반올림 처리
- **area** — 시장지역 또는 근린 식별자(예: **area_id**, **neighborhood**, **market_area**)
- **living_area** — 제곱피트당 잔차 계산 (**residual_sf**, **cqa_sf**)을 가능하게 함
- **lot_size** — 대지 규모 열
- **site_dimensions** — 대지 규모와 함께 묶임(예: “75x120”)

연령 유형:

- **actual_age** — 부동산 실제 연령 열
- **effective_age** — 부동산 유효 연령

표시 유형:

- **display_only** — 해당 열은 Excel 내보내기에는 포함되지만 모형 적합에서는 완전히 제외된다. 주소 항목, MLS 번호, 필지 ID, 기타 참조 자료에 사용한다. 여러 열에 이 지정을 줄 수 있다.

요령: *display_only*로 지정된 열은 데이터셋에 남아 *Excel* 내보내기에 나타나지만 설명변수 표에서는 완전히 제외된다. *ID* 열, 주소, 기타 참조 항목에 사용하라.

7 모수 선택

사이드바 제5절 — **glmnet Call Parameters** — 은 엘라스틱 넷 모형의 모든 구성 옵션에 접근할 수 있게 한다. 각 모수에는 설명 툴팁이 달린 파란색 도움말 아이콘(?)이 있다.

7.1 Alpha(혼합 모수)

alpha 모수는 정규화의 유형을 조절한다.

Alpha 값	유형	동작
0	릿지	모든 계수를 비례적으로 축소하 되 어느 것도 0으로 만들지 않음
1	Lasso	계수를 정확히 0으로 만들 수 있 음(변수 선택)
0에서 1 사이	엘라스틱 넷	릿지와 lasso의 혼합

alpha를 선택하는 방법은 두 가지이다.

- **Fixed** — 슬라이더로 단일 alpha 값을 설정(기본값: 1.0)
- **Grid Search** — 여러 alpha 값을 시험하여 CV 오차가 가장 낮은 값을 선택한다. 범위와 값의 개수를 설정할 수 있다.

7.2 Lambda 선택

lambda는 정규화의 강도를 조절한다.

방법	설명
Cross-validation	권장. k겹 CV를 사용하여 많은 lambda 값을 시험한다.
Manual	타당한 이유가 있다면 특정 lambda 값을 직접 입력한다.

교차검증을 사용할 때는 두 가지 lambda 선택지가 있다.

- **lambda.1se**(기본값) — 최솟값으로부터 1 표준오차 이내에 있는 가장 큰 lambda이다. 더 단순하고 더 강하게 정규화된 모형을 만든다. 감정평가 업무에 권장된다.
- **lambda.min** — 교차검증 오차를 최소화하는 lambda이다. 적합도는 가장 좋지만 모형이 더 복잡해질 수 있다.

Number of CV Folds 모수는 교차검증을 위해 데이터를 어떻게 나눌지를 조절한다(기본값: 10겹).

7.3 Family

반응변수의 분포 계열을 선택한다.

Family	사용 상황
gaussian	연속형 반응(예: 매매가격). 가장 흔하다.
binomial	이항 결과(예: 매도됨/매도되지 않음).
poisson	계수(count) 자료(예: 매매 건수).

7.4 Standardize

선택하면(기본값) 적합 전에 모든 설명변수가 평균 0, 표준편차 1이 되도록 척도화된다. 이로써 원래 척도와 무관하게 별점이 모든 설명변수를 동등하게 다루게 된다. 계수는 원래 척도로 되돌려 반환된다. 대개 켜 둔 채로 두는 것이 좋다.

7.5 부호 제약

Enforce Sign Constraints가 활성화되면 계수는 변수 표(제6장)에서 설정한 기대 부호에 부합하도록 제약된다.

- “positive” 부호는 계수를 ≥ 0 이 되도록 강제한다
- “negative” 부호는 계수를 ≤ 0 이 되도록 강제한다
- “either”로 설정된 변수는 제약을 받지 않는다

이는 `glmnet()`의 `upper.limits` 및 `lower.limits` 모수로 구현된다.

7.6 완화 Lasso

일반적인 lasso는 변수 선택과 계수 추정에 동일한 벌점을 쓰기 때문에 중요한 계수를 0 쪽으로 과도하게 축소할 수 있다. **완화 lasso**(relaxed lasso)는 이 두 단계를 분리한다.

1. 먼저 lasso로 변수를 선택한다(벌점이 어떤 계수가 살아남을지 결정한다)
2. 그다음 살아남은 변수들을 더 작은 벌점 또는 무벌점으로 다시 적합하여 편향이 덜한 추정치를 얻는다

Gamma 슬라이더는 완화의 정도를 조절한다.

Gamma	효과
0	완전 완화(OLS 재적합, 축소 전혀 없음)
1	완화 없음(일반 glmnet과 동일)
그 사이	완화 적합과 벌점 적합을 혼합

교차검증을 쓰면 최적 성능을 위해 감마가 자동으로 선택된다.

7.7 난수 시드

무작위 정수가 미리 채워진 **Random seed** 텍스트 입력란은 교차검증 겹 배정을 재현 가능하게 만든다—적합 전에 `set.seed()`가 호출되며, 사용된 시드는 Glmnet Output 탭과 적합 상태 메시지에 표시된다. 매 적합 후에는 (R의 난수 생성기가 아니라 시스템 시각을 사용하여) 새 시드가 자동 생성된다. 현재 파일에 대해 마지막으로 사용한 시드 5개는 클릭 가능한 링크로 표시되어 쉽게 다시 불러올 수 있으며, 새 파일을 불러오면 시드 이력은 초기화된다.

7.8 교호작용 행렬

대화형 상삼각 행렬을 통해 어떤 변수 쌍의 교호작용을 허용할지 조절할 수 있다. 각 칸에는 체크 박스가 있으며, 선택되어 있으면 교호작용 항 `x1:x2`가 모형 행렬에 포함된다.

- **Allow All** — 모든 교호작용 쌍을 선택
- **Clear All** — 모든 쌍의 선택을 해제(새 파일의 기본값은 전부 해제 상태)
- **변수 이름을 클릭** — 해당 변수의 모든 교호작용을 토글

교호작용 선택은 입력 파일별로 브라우저 localStorage에 저장되어 같은 파일을 다시 불러올 때 복원된다.

교호작용이 하나라도 활성화되면 내보낸 보고서에 **Enabled Interactions** 행렬(양 축에 설명변수, `x1:x2` 곱 항이 포함된 자리에 체크 표시)이 가로 방향으로 실려 earthUI의 Allowed

Interactions 행렬과 짝을 이룬다. 다만 활성화된 항이라도 lasso가 그 계수를 0으로 축소했을 수 있다는 점에 유의하라.

요령: 설명변수가 많을 때 lasso는 불필요한 교호작용 항을 자동으로 0으로 만든다. "Allow All"로 시작해 모형이 선택하도록 두는 것도 합리적인 접근이다.

경고

glmnet 교호작용을 활성화하면 범정이나 감사 상황에서 해석하고 설명하기 어려운 보정치가 산출될 수 있다. 교호작용은 신중하게 사용하고, 그 결과로 나온 보정치가 합리적인지 검증하라.

7.9 주효과에서 차단(교호작용 전용)

교호작용 행렬에서 변수 이름을 오른쪽 클릭하면 그 변수가 주효과로 모형에 들어가지 못하도록 차단할 수 있다. 변수 이름 뒤에 굵은 "1" 표시가 나타난다. 차단되면 다음과 같다.

- 해당 변수는 주효과 항에서 제거된다
- 해당 변수는 교호작용 항($x_1:x_2$)에는 여전히 나타날 수 있다
- 다시 오른쪽 클릭하면 차단이 해제된다
- 이 상태는 입력 파일별로 localStorage를 통해 유지된다

전형적인 사용 사례는 다음과 같다. 부동산 규모에 비례해야 하는 시점 보정을 모형화할 때 sale_age를 주효과에서 차단하고 living_area와의 교호작용에서만 허용한다. 그러면 평평한 sale_age 계수 대신 sale_age:living_area 교호작용 항이 만들어져—시간에 따른 가격 상승률이 주택 규모에 따라 달라진다.

7.10 고급 모수

glmnet Call Parameters 맨 아래의 접이식 "Advanced" 절에는 추가 설정이 있다. 이들은 합리적인 기본값을 사용하지만, 범정이나 감사를 위해 모든 모형 설정을 문서화할 수 있도록 노출되어 있다.

- **Lambda Min Ratio** — 정규화 경로에서 가장 작은 lambda와 가장 큰 lambda의 비율. 기본값: 0.00001. 소규모 데이터셋 (관측치 50개 미만)에서는 더 큰 값(예: 0.01)이 성능을 개선할 수 있다.
- **Number of Lambda Values** — 경로상의 lambda 값 개수. 기본값: 100.
- **CV Loss Metric** — 교차검증의 손실함수: MSE (gaussian의 기본값), MAE(이상치에 강건), Deviance.
- **Convergence Threshold** — 좌표하강의 수렴 임계값. 기본값: $1e-07$.
- **Max Iterations** — 좌표하강의 최대 반복 횟수. 기본값: 100,000. 수렴 경고가 나타나면 늘린다.
- **Fit Intercept** — 절편(상수) 항 포함 여부. 기본값: 켜. 절편은 감정평가 RCA에서 "기준 값(basis value)"에 해당한다.

7.11 설정 기본값

파일을 불러올 때 모수를 어떻게 초기화할지를 세 가지 옵션이 제어한다.

- **Use last settings for input file** — 이 파일에 대해 지난번에 사용한 설정을 그대로 복원
- **Use default settings** — 저장해 둔 전역 기본값을 적용
- **glmnet defaults** — 공장 설정으로 초기화($\alpha = 1$, CV lambda.1se, 10겹, gaussian, standardize 켜)

Save current as default를 클릭하면 현재의 모든 설정이 이후 파일들에 적용될 전역 기본값으로 저장된다.

8 모형 적합

8.1 Fit 버튼

사이드바 제6절에는 **Fit Glmnet Model** 버튼이 있다. 이를 클릭하면 현재 구성으로 모형이 실행된다.

8.2 적합 중에 일어나는 일

Fit을 클릭하면 glmnetUI는 다음을 수행한다.

1. 데이터를 준비한다 — NA가 있는 행을 제거하고, (지정되어 있다면) 가중치를 적용하며, (감정평가 모드에서는) 대상 행을 제외하고, 요인 변수를 가변수 열로 부호화한다.
2. 모형 행렬을 구성한다 — 교호작용 행렬에서 허용된 교호작용 항을 포함하여 설계행렬을 만든다.
3. 제약을 적용한다 — 부호 적용이 활성화되어 있으면 계수에 `upper.limits`와 `lower.limits`를 설정한다.
4. 교차검증을 수행한다 — CV가 선택되어 있으면 `cv.glmnet()`을 호출하여 최적 `lambda`를 찾는다. 완화 lasso가 활성화되어 있으면 `cv.glmnet(..., relax = TRUE)`를 호출한다.
5. 최종 모형을 적합한다 — 적합된 모형, 계수, 예측값, 잔차를 저장한다.
6. 모든 결과 탭을 갱신한다 — Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA, Diagnostics가 채워진다.

Fit 버튼 아래의 상태 메시지는 관측치 수, 제외된 행 수, 선택된 `lambda`를 보여 준다.

9 결과 탭

9.1 Data Preview

가져온 데이터를 대화형 DataTable로 보여 준다. 감정평가/시장 모드에서는 미리보기가 두 표로 나뉜다. **Subject Property**(1행)와 **Comparable Sales**(2행 이하)이다.

각 칸은 한 줄로 유지되고 열 너비에 맞추어 잘리므로, 폭이 넓은 자유 서술 항목(예: 매물 설명)이 행을 아래로 늘어뜨리지 않는다. 아무 칸이나 더블클릭하면 잘리지 않은 전체 내용을 보여 주는 팝업이 열린다.

9.2 Equation

MathJax를 통해 LaTeX로 렌더링된 적합 모형 방정식을 표시한다. 0이 아닌 모든 계수를 적절한 수학 서식으로 보여 준다.

- 양의 계수는 + 부호로 표시
- 음의 계수는 - 부호로 표시
- 교호작용 항은 $x_1 \times x_2$ 로 표시
- 숫자는 쉼표와 적절한 소수 자릿수로 서식 지정

9.3 Correlation

모든 수치형 설명변수와 반응변수 사이의 피어슨 상관계수를 보여 주는 히트맵 행렬이다. 데이터를 가져온 직후 바로 사용할 수 있다—모형 적합이 필요하지 않다.

- 색상 그레데이션: 파랑(-1) → 흰색(0) → 빨강(+1)
- 각 칸에 상관계수 값이 인쇄됨
- 변수 개수에 따라 글자와 축 크기가 조정됨
- 적합 전에 다중공선성을 확인하는 데 유용함

9.4 Summary

모형 적합 통계량이 카드 형태로 표시된다.

- R^2 — 설명된 분산의 비율
- 조정 R^2 — 설명변수 개수를 반영해 조정한 값
- GR^2 — 선택된 lambda에서 glmnet의 이탈도 비율로부터 얻은 일반화 R^2
- $CV R^2$ — 교차검증 R^2 (CV를 사용한 경우)
- **RMSE** — 평균제곱근오차
- **MAE** — 평균절대오차

감정평가/시장 모드에서는 다음 카드가 추가로 표시된다.

- **Median Ratio**(중위 비율) — (예측값 / 실제값)의 중앙값
- **COD** — 산포계수(Coefficient of Dispersion, 낮을수록 좋음)
- **PRD** — 가격관련편차(Price-Related Differential)

훈련 R^2 가 $CV R^2$ 를 0.1 넘게 초과하면 **과적합 경고**가 나타난다.

카드 아래에는 편의를 위해 Coefficients 탭과 동일한, 부호 경고가 표시된 계수 표가 반복해서 실린다.

9.5 Coefficients

0이 아닌 모든 계수를 다음과 함께 보여 주는 표이다.

- 변수 이름
- 계수 값

- 기대 부호(변수 구성에서 지정한 값)
- 계수의 방향이 기대와 맞지 않을 경우의 부호 경고 표시

9.6 Variable Importance

표준화된 계수 크기, 즉 각 설명변수에 대한 $|\beta| \times \text{sd}(x)$ 이며, 요인 변수의 경우 가변수 열들에 걸쳐 합산된다. 두 가지 표시 방식이 있다.

- **대화형 plotly**(plotly 패키지가 설치된 경우): 정확한 중요도, 계수, 상대 백분율을 보여주는 호버 툴팁이 달린 가로 막대그래프.
- **정적 ggplot2**(대체 방식): 심표 서식의 축 이름표가 달린 가로 막대그래프.

그래프 아래의 DataTable은 Variable, Importance, Coefficient, Relative %를 보여 준다.

9.7 Contributions

각 설명변수가 모형의 예측에 기여하는 정도를 보여 주는 변수별 부분효과 그림이다. 드롭다운에서 변수를 선택한다.

수치형 설명변수의 경우: 변수의 x값 대 그 기여도 ($\beta \times x$)의 산점도이며, 빨간 선분과 기울기 표시가 함께 나타난다. 기울기 표시는 단위당 한계효과(예: +1,234.56/unit)를 보여 주며, 위도/경도처럼 값의 범위가 좁은 변수에는 적응적 단위(예: +0.12/0.001)를 사용한다. 부제에는 절편(기준) 값이 표시된다.

요인 변수의 경우: 요인 수준별 기여도 값의 분포를 보여 주는 상자그림이다.

두 변수 교호작용의 경우: 세 가지 시각화가 위아래로 배치된다 — 결합 기여도의 3차원 곡면 그림(가능하면 대화형 plotly, 그렇지 않으면 정적 persp()), 기여도에 따라 점의 색을 달리한 산점도, 그리고 두 변수의 범위 전반에 걸친 평균 기여도의 히트맵이다.

9.8 ANOVA

각 설명변수에 대해 다음을 보여 주는 분산 분해 표이다.

- **SS** — 제곱합: $\sum (\text{contribution}_j - \overline{\text{contribution}_j})^2$
- **% of Model SS** — 모형 총제곱합에서 차지하는 비율
- **Coefficient** — 해당 설명변수의 주된 계수

절편 행과 TOTAL MODEL 행이 포함된다. 교호작용 항(“:”을 포함하는 항)은 별도의 그룹으로 표시된다.

9.9 Diagnostics

큰 글자(기본 크기 16pt), 15개의 축 눈금, 심표 서식의 이름표(지수 표기 없음)를 갖춘 네 가지 진단 그림이다.

1. **계수 경로(Coefficient Path)** — $\log(\lambda)$ 가 커짐에 따라 모든 계수가 어떻게 변하는지 보여 준다. 선은 변수별로 색이 구분된다. 정규화 경로와 변수 선택을 시각적으로 파악하는 데 도움이 된다.
2. **CV 오차(CV Error)** — $\log(\lambda)$ 의 함수로 나타낸 교차검증 오차이다. 점은 평균 CV 오차를, 오차막대는 ± 1 표준오차를 나타낸다. 파선 수직선은 $\lambda_{\min}$ (파랑)과 λ_{1se} (초록)를 표시한다.
3. **실제값 대 예측값(Actual vs Predicted)** — 관측값 대 예측값의 산점도이다. 점들은 45도 파선 기준선 주변에 모여야 한다. 산포가 넓을수록 예측 오차가 크다는 뜻이다.
4. **잔차 대 적합값(Residuals vs Fitted)** — 잔차 대 적합값의 산점도이다. 0선(파선) 주위에 무작위로 흩어져야 한다. 어떤 패턴이 보인다면 모형 설정 오류(예: 비선형성, 이분산성)를 시사한다.

각 그림에는 150 DPI로 저장할 수 있는 **Download PNG** 버튼이 있다.

9.10 Glmnet Output

원시 모형 출력으로, 사용된 난수 시드, 모형 출력, 선택된 λ , $\lambda_{\min}/\lambda_{1se}$, 감마 (완화 lasso가 활성화된 경우), 전체 계수 벡터를 보여 준다.

9.11 Report

보고서 항목(감정평가사 이름, 부동산 주소, 보고서 일자, 문서 번호)을 구성하고 이 탭에서 곧바로 Word(.docx) 또는 PDF로 내보낸다. Quarto 기반 보고서 작업 흐름 전체(HTML, Word, PDF)는 제13장을 참조하라.

10 데이터 내려받기

적합 후 **Download Output (Excel)** 버튼(사이드바 제7절, 모든 용도 모드에서 사용 가능)은 예측값과 진단 정보가 담긴 Excel 파일을 프로젝트의 출력 폴더로 내보낸다. 이 절의 제목은 감정평가 모드와 시장 모드에서는 “Download Estimated Sale Prices & Residuals”로, General 모드에서는 “Download Estimated Target Variable(s) & Residuals”로 표시된다.

10.1 출력 열

열	설명
est_<target>	모형 예측값(예: est_sale_price)
residual	실제값 - 예측값
cqa	상태-품질-선호도(Condition-Quality-Appeal) 점수(0-10 척도)
residual_sf	잔차 / 주거면적(living_area가 지정된 경우)
cqa_sf	residual_sf 순위로부터 계산한 CQA
<var>_contribution	예측값에 대한 변수별 기여도
basis	절편 값(모든 행에서 동일)

성공적으로 완료되면 내려받기 버튼에 흰색 체크 표시가 나타난다.

10.2 CQA 점수

CQA는 각 행의 잔차를 다른 모든 행과 견주어 0-10 척도로 순위화한다.

- **높은 CQA(~9-10):** 예측보다 훨씬 비싸게 팔림 — 상태/품질/선호도가 우수할 가능성이 높다
- **낮은 CQA(~0-1):** 예측보다 훨씬 싸게 팔림 — 상태가 열등하거나 급매일 가능성이 높다
- **CQA ~5:** 중앙값 부근

감정평가/시장 모드에서는 living_area 열이 지정되어 있으면 행이 residual_sf 내림차순으로 정렬된다.

11 RCA 계산과 내려받기

RCA(잔차 제약 접근법, Residual Constraint Approach) 작업 흐름은 모형 적합 후 **Appraisal** 및 **Market Area Analysis** 모드에서 사용할 수 있다. 후속 단계인 거래사례비교표(제12장)는 감정 평가 모드 전용이다.

11.1 RCA 대화상자 열기

사이드바 제8절의 **Calculate RCA Adjustments & Download** 버튼을 클릭한다. 그다음 선택한다.

- **CQA** 또는 **CQA per SF** 점수 유형. “CQA per SF”를 선택하면 대상의 잔차가 주거면적으로 척도화된다.
- 대상의 CQA 점수를 입력한다(0.00–9.99, 기본값 5.00)

11.2 CQA 점수 보간

1. 비교 사례들의 CQA 점수와 잔차가 정렬된다
2. 선형 보간으로 입력한 CQA 값을 잔차로 대응시킨다
3. 대상 가치 = 모형 예측값 + 보간된 잔차

11.3 출력 열

열	설명
subject_value	모형 예측값 + 보간된 잔차
<var>_adjustment	대상 기여도 – 비교 사례 기여도
residual_adjustment	대상 잔차 – 비교 사례 잔차
net_adjustments	모든 보정치의 합
gross_adjustments	보정치 절댓값의 합
adjusted_sale_price	비교 사례 매매가격 + 순보정치

계산이 성공적으로 끝나면 버튼에 흰색 체크 표시가 나타난다.

12 거래사례비교표

거래사례비교표(Sales Comparison Grid)는 RCA 보정치를 계산한 뒤(제8단계) **Appraisal** 모드에서만 사용할 수 있다. 이 기능은 감정평가서에 넣기에 적합한 서식화된 Excel 통합문서를 생성한다.

12.1 비교 사례 선택 대화상자

사이드바 제9절의 **Generate Sales Grid & Download** 버튼을 클릭하면 RCA 출력에 있는 모든 비교 거래사례를 나열하는 모달 대화상자가 열린다. 비교 사례는 두 그룹으로 나뉜다.

- **권장 사례**(미리 선택됨): 총 보정치가 매매가격의 25% 미만인 사례로, 매매 경과일 순(가장 최근 순)으로 정렬된다.
- **기타 사례**(기본적으로 선택 해제): 나머지 사례로, 매매 경과일 순으로 정렬된다.

최대 30개의 사례를 선택한 뒤 **Generate Sales Grid**를 클릭하면 통합문서가 만들어진다.

12.2 통합문서 구성

생성된 통합문서는 최대 10개의 시트를 가지며, 각 시트에는 비교 사례 3건이 나란히 배치된다.

- **머리글 행**: 대상 및 비교 사례의 주소, 매매가격, 매매일, 시장 노출 일수
- **묶음 행**: 위치(위도, 경도, 지역), 대지 규모 (대지 면적, 대지 치수), 연령(실제 연령, 유효 연령)을 그룹 머리글과 하위 항목으로 구성
- **모형 변수 행**: 0이 아닌 모형 계수마다 한 행씩, 대상 기여도, 비교 사례 기여도, 보정 비율을 표시
- **잔여 특성 행**: 모형이 포착하지 못한 특성(예: 상태, 품질, 조망)에 대해 감정평가사가 입력할 수 있는 빈 잠금 해제 칸
- **요약 행**: 순보정%, 총보정%, 그리고 실시간 Excel 수식이 들어간 보정 매매가격

12.3 시트 보호와 잔여 칸

각 시트는 실수로 편집되는 것을 막기 위해 암호로 보호되지만, 잔여 특성 값 칸은 명시적으로 잠금이 해제되어 있다. 이로써 감정평가사는 수식 기반 보정 계산을 그대로 유지하면서 모형에 없는 특성의 값을 입력할 수 있다.

13 보고서 내려받기

13.1 두 단계 Quarto 작업 흐름

사이드바에서의 보고서 내보내기는 두 단계의 Quarto 작업 흐름이다.

1. **Generate Quarto Report**(사이드바 제10절) — 자기완결적 Quarto 묶음(.qmd 원본과 그림, report_data.rds, reference.docx)을 프로젝트의 glmnet 출력 폴더에 기록한다.
2. **Convert Quarto Report**(사이드바 제11절) — .qmd 파일을 선택한 출력 형식으로 렌더링한다. 경로는 가장 최근에 생성된 묶음이 기본값이지만, 어떤 .qmd든 찾아볼 수 있다.

13.2 보고서 형식

변환 단계에서는 세 가지 형식을 쓸 수 있다.

- **HTML** — 자기완결적 HTML 문서. pandoc 외에 추가 소프트웨어 없이 모든 플랫폼에서 작동한다. 권장 기본값이다.
- **Word (.docx)** — 서식화된 문서. 편집과 배포에 적합하며 모든 플랫폼에서 작동한다.
- **PDF** — LaTeX을 사용하는 Quarto로 생성된다. 용지 크기는 로케일 설정(Letter 또는 A4)을 따른다. **LaTeX** 설치가 필요하다(시스템 요구사항 부록 참조). LaTeX이 감지되지 않으면 형식 체크박스에서 PDF 옵션이 자동으로 숨겨진다.

보고서는 가능한 경우 Quarto로 생성되며, 그렇지 않으면 rmarkdown으로 대체된다. Report 탭은 이와 별도로 Word 내보내기와 PDF 내보내기 버튼을 직접 제공한다.

13.3 보고서 내용

보고서에는 모든 탭의 내용이 포함된다.

- 데이터셋 설명(파일 이름, 관측치 수, 반응변수, 설명변수, 용도)
- 모형 사양(alpha, lambda, family, standardize, 완화 여부, 0이 아닌 계수)
- 결과 요약(R^2 , 조정 R^2 , 일반화 R^2 , CV R^2 , RMSE, MAE)
- 모형 방정식(LaTeX으로 렌더링)
- 계수 표
- 변수 중요도(막대그래프와 표)
- Contributions 탭과 동일한 변수별 기여도 그림:
 - 수치형 설명변수: 선형 적합선이 있는 산점도
 - 요인 변수: 수준별 상자그림
 - 교호작용: 기여도로 색을 구분한 산점도, 평균 기여도의 히트맵, 정적 3차원 곡면 스냅샷(persp)
- Enabled Interactions 행렬(교호작용이 하나라도 활성화된 경우) — 양 축에 설명변수를 두고 $x1:x2$ 곱 항이 포함된 자리에 체크 표시를 하며, 가로 방향으로 실린다
- 상관 행렬 히트맵
- ANOVA 분해 표
- 진단 그림:
 - 계수 경로
 - 교차검증 오차
 - 실제값 대 예측값
 - 잔차 대 적합값
 - 정규 Q-Q 그림
- 모형 출력

모든 축 이름표와 색상 범례는 심표 서식의 숫자를 사용한다(지수 표기 없음).

보고서가 성공적으로 생성되면 보고서 버튼에 흰색 체크 표시가 나타나며, 보고서를 렌더링하는 동안에는 경과 시간이 표시된다.

14 earthUI와의 비교

earth 애플리케이션과 그 glmnet 후처리 모듈은 회귀 모형화를 위한 하나의 도구 모음이다. 두 도구는 동일한 데이터 형식, 특수 열 유형, RCA 작업 흐름, 예제 데이터셋을 공유하지만 모형화 엔진은 서로 다르다.

14.1 주요 차이

항목	glmnetUI	earthUI
방법	엘라스틱 넷(glmnet)	MARS/Earth(earth)
변수 선택	lasso 별점을 통한 자동 선택	전진/후진 가지치기를 통한 자동 선택
계수 관계	선형 효과만	조각별 선형(힌지 함수)
교호작용	교호작용 행렬을 통한 쌍별 교호작용	3차까지
계수 부호	제약을 통해 강제 가능	직접 강제 불가
완화 적합	완화 lasso 사용 가능	해당 없음
Alpha 조정	격자 탐색 사용 가능	해당 없음(혼합 모수 없음)
거래사례비교표	사용 가능(제9단계)	사용 가능(제8단계)
g-함수 그림	해당 없음	묶음 향 시각화
보고서 형식	HTML, Word, PDF	HTML, Word, PDF

14.2 공통 기능

두 도구 모두 다음을 제공한다.

- snake_case 열 변환을 동반한 CSV 및 Excel 가져오기
- CSV 구분자, 소수점 기호, 용지 크기를 포함한 31개국 로케일 지원
- 특수 열 지정(contract_date, living_area, area, latitude, longitude 등)
- 기준일로부터의 sale_age 자동 계산
- 다크/라이트 모드 전환
- localStorage를 통한 설정 보존
- 상관 히트맵, 변수 중요도, 기여도, ANOVA, 진단
- CQA 점수 산정과 RCA 보정 작업 흐름
- 비교 사례 선택과 서식화된 Excel 출력을 갖춘 거래사례비교표
- 동일한 예제 데이터셋(Appraisal_1.csv)과 호환

14.3 어느 것을 언제 쓸 것인가

- 자동 변수 선택을 갖춘 단순하고 해석하기 쉬운 선형모형을 원한다면 **glmnetUI**를 사용하라. 엘라스틱 넷은 설명변수가 많고 모형이 가장 중요한 것들을 골라 주기를 바랄 때 특히 적합하다.
- 비선형 관계를 예상하거나(예: 주거면적이 가격에 미치는 효과가 규모에 따라 달라지는 경우) 임계효과를 포착할 힌지 함수가 필요하다면 **earthUI**를 사용하라. earth 모형은 더 높은 차수의 교호작용(2차, 3차)도 포착할 수 있다. earthUI는 또한 glmnetUI에는 대응물이 없는 실험적인 Prolog/DCG 기반 매물 설명 추출 기능을 제공한다.
- 같은 데이터에 대해 선형모형과 비선형모형을 비교하려면 둘 다 사용하라. 두 모형의 R^2 값이 비슷하다면 해석 가능성 측면에서 더 단순한 glmnet 모형이 선호될 수 있다.

15 예제 데이터셋: Appraisal_1.csv

15.1 설명

예제 데이터셋은 glmnetUI에 함께 들어 있으며 earthUI와도 공유된다. 다음으로 위치를 확인할 수 있다.

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

이 파일에는 모의 MLS 내보내기 자료로 만든 주거용 매매 1,502건(그리고 1행의 대상 부동산 1건)이 들어 있다. 데이터는 규모, 연령, 위치가 다양한 여러 지역 시장의 단독주택 매매를 나타낸다.

이는 실제 데이터가 아니지만 북부 캘리포니아의 현실적인 근린을 바탕으로 한 것이다. 모든 식별 정보는 변경되거나 삭제되었다.

15.2 열

열	자료형	특수 유형	설명
weight	numeric	weight	관측치 가중치(0 = 적합에서 제외)
id	numeric	display_only	내부 레코드 ID
property_id	numeric	display_only	MLS 부동산 식별자
listing_id	character	display_only	MLS 매물 번호
parcel_number	character	display_only	카운티 과세평가관 필지번호(APN)
street_address	character	display_only	부동산 주소
city_name	character	display_only	시
postal_code	character	display_only	우편번호
county_name	character	display_only	카운티
contract_date	Date	contract_date	매매 계약일(sale_age를 계산)
sale_age	numeric	—	계약일부터 기준일까지의 일수
sale_price	numeric	(목표변수)	매매가격 — 반응변수
living_sqft	numeric	living_area	총 주거면적(제곱피트)
beds_total	integer	—	침실 수
baths_total	numeric	—	총 욕실 수(예: 2.5 = 전용 2 + 반욕실 1)
lot_size	numeric	lot_size	대지 면적(제곱피트)
area_id	integer	area	MLS 지역 식별자
age	numeric	actual_age	부동산 연령(년)
latitude	numeric	latitude	위도(모형에서는 소수점 셋째 자리로 반올림)
longitude	numeric	longitude	경도(모형에서는 소수점 셋째 자리로 반올림)
garage_spaces	integer	—	차고 칸 수
days_on_market	integer	dom	시장 노출 일수
listing_date	Date	listing_date	매물 등록일
sale_concessions	numeric	concessions	매도인 양허

15.3 권장 빠른 시작

1. glmnet 루틴을 실행한다: `earthUI::launch_glmnet()`
2. 용도를 **For Appraisal**로 하여 새 프로젝트를 만든다(제1절)
3. Appraisal_1.csv를 프로젝트의 입력 폴더에 넣고 제2절에서 선택한다(필요하면 **Refresh**를 클릭)
4. sale_price를 목표변수로 선택한다
5. 위 표에 나온 대로 특수 유형을 지정한다

6. 설명변수를 포함시킨다: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id`(요인으로 지정), `age`, `latitude`, `longitude`, `garage_spaces`
7. `alpha = 1`(Lasso), `lambda.1se`를 쓰는 CV를 그대로 둔다
8. **Fit Glmnet Model**을 클릭한다
9. Summary, Coefficients, Diagnostics 탭을 검토한다
10. 출력을 내려받고(제7단계) CQA 순위를 검토한다
11. CQA 점수를 ~5.00으로 하여 RCA 보정치를 계산한다 (제8단계)
12. 권장 비교 사례로 거래사례비교표를 생성한다(제9단계)

16 시스템 요구사항과 문제 해결

16.1 지원 플랫폼

glmnetUI는 **macOS**, **Windows**, **Linux**에서 실행된다. RStudio Desktop(2023.06 이상)을 적극 권장하는데—보고서에 필요한 **pandoc**이 함께 들어 있기 때문이다.

16.2 선택적 의존 요소

- **LaTeX**(TinyTeX, MiKTeX 또는 MacTeX): PDF 보고서 생성에만 필요하다. 감지되지 않으면 PDF 옵션이 자동으로 숨겨진다. 설치 방법: `tinytex::install_tinytex()`
- **sysfonts + showtext**: 그림용 Roboto Condensed 글꼴. 사용할 수 없거나 오프라인이면 시스템 산세리프로 자동 대체된다.
- **earth**($\geq 5.3.0$): earthUI 가져오기 파이프라인에만 필요하다.

16.3 플랫폼별 참고사항

macOS: 문제가 가장 적다. PDF를 위해 TinyTeX을 설치하라: `tinytex::install_tinytex()`

Windows: RStudio와 잘 작동한다. 기업용/보안이 강화된 컴퓨터는 임시 디렉터리 제한이 있을 수 있는데—앱은 콘솔에 경고를 표시하되 설정 보존 없이 계속 실행된다.

Linux: 시스템 라이브러리가 필요할 수 있다: `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

16.4 점진적 기능 축소

없는 구성요소	동작
LaTeX 없음	PDF 옵션이 숨겨짐. HTML과 Word는 계속 사용 가능.
인터넷 없음 / 글꼴 실패	시스템 산세리프 사용. 콘솔에 메시지 기록.
SQLite 없음 / 읽기 전용 파일시스템	설정이 보존되지 않음. 앱은 정상 실행.
임시 디렉터리에 쓸 수 없음	보고서 생성이 명확한 오류와 함께 실패. TMPDIR을 쓰기 가능한 위치로 설정하라.

16.5 문제 해결

“**PDF option not available**”: R에서 `tinytex::install_tinytex()`을 실행한 뒤 앱을 다시 시작하라.

“**Settings will not persist**”: 사용자 데이터 디렉터리의 권한을 확인하라(macOS: `~/Library/Application Support/R/glmnetUI/`, Linux: `~/.local/share/R/glmnetUI/`, Windows: `%APPDATA%/R/data/glmnetUI/`).

7879 포트가 이미 사용 중임: `lsof -ti:7879 | xargs kill`(macOS/Linux) 또는 `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell)를 실행하라.

참고문헌

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(일반화 가법 모형) 패키지를 위한 대화형 인터페이스 (한국어판)

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 검토했습니다. 영문판이正本입니다.

William Bert Craytor

제 1 권 제 1 호 · 2026 년 여름

발행일: 2026 년 7 월 1 일

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

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

이 논문은 오픈 액세스로 공개되며 [크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스 \(CC BY 4.0\)](#)의 조건에 따라 배포됩니다.

초록. mgcvUI 는 R mgcv 패키지를 위한 그래픽 사용자 인터페이스 (graphical user interface) 이다. mgcv 는 자동 평활도 선택 (smoothness selection) 을 갖춘 벌점 회귀 스플라인 (penalized regression spline) 으로 일반화 가법 모형 (generalized additive model, GAM) 을 적합한다. mgcvUI 는 범용 예측 모델링, 부동산 감정평가, 시장지역 분석 (market-area analysis) 의 세 가지 목적 모드를 제공하며, 데이터 가져오기, 평활항 (smooth term) 설정, 모형 적합, 진단 그림 및 효과 그림 확인, 보고서 내려받기에 이르는 과정을 사용자에게 안내한다. 본고는 mgcvUI 의 데이터 형식 요건, 모델링 작업 흐름, 출력 화면, 그리고 전체 기능 참조를 문서화한다.

키워드: 일반화 가법 모형, GAM, 벌점 스플라인, 평활화, mgcv, R, 그래픽 사용자 인터페이스, 부동산 감정평가

발행인 주 (2026 년 7 월): 인쇄 직전에 mgcvUI 는 earthUI(0.11.0 이후 버전) 에 통합되었으며, 본고에서 설명하는 GAM 작업 흐름은 이제 MARS 및 엘라스틱 넷 작업 흐름과 함께 earthUI 에서 제공된다. mgcvUI 는 최종 배포판 (0.3.0) 형태로 <https://github.com/wcraytor/mgcvUI>에 계속 공개되어 있으며, 향후 개발은 earthUI 에서 이어진다. 본고는 배포된 상태의 mgcvUI 를 다룬다.

목차

1	서론	4
1.1	역사와 배경	4
1.1.1	일반화 가법 모형	4
1.1.2	mgcv 패키지	4
1.1.3	MARS(earth) 및 엘라스틱 넷 (glmnet) 과의 비교	5
1.1.4	earthUI-mgcvUI 파이프라인	5
1.1.5	권장 작업 흐름	5
1.2	mgcvUI 란 무엇인가?	6
1.3	세 가지 목적 모드	6
1.4	부동산 전용 기능	6
1.5	시작하기	7
2	MLS 입력 데이터 요건	8
2.1	파일 형식과 구조	8
2.2	감정평가 모드에 필요한 열	8
2.3	특수 열 명명 규칙	9
2.4	데이터 품질과 완전성	9
2.5	대상 행의 위치	9

3	범용 모드	10
3.1	개요	10
3.2	사이드바 작업 흐름	10
3.3	earthUI 에서 가져오기 (선택 사항)	11
3.4	메인 패널 탭	11
3.5	설정의 유지	12
3.6	다크 모드	12
3.7	로케일과 지역 설정	12
4	감정평가 모드	14
4.1	대상 행의 처리	14
4.2	기준일과 매매 경과일	14
4.3	특수 열 지정	14
4.4	RCA 조정 개요	14
5	시장지역 분석 모드	15
5.1	감정평가 모드와의 차이	15
5.2	시장 모드를 사용하는 경우	15
6	변수 선택	16
6.1	목표변수	16
6.2	반응변수 변환	16
6.3	예측변수 표	16
6.4	데이터 유형 탐지와 재지정	17
6.5	특수 열 유형 참조	17
7	모수 선택	18
7.1	모수 프리셋	18
7.2	Family(분포족)	18
7.3	Method(방법)	18
7.4	Gamma(평활 승수)	18
7.5	교차검증	19
7.6	Select(수축)	19
7.7	기본 기저	19
7.8	기본 k(기저 차원)	19
7.9	텐서 유형	19
7.10	허용 상호작용	20
7.11	요인별 평활 상호작용	20
7.12	고급 모수	20
8	모형 적합	21
8.1	Fit 버튼	21
8.2	적합 중에 일어나는 일	21
8.3	백그라운드 적합과 진행 상황	21
9	결과 탭	22
9.1	Data(데이터)	22
9.2	Equation(식)	22
9.3	Summary(요약)	22
9.4	Variable Importance(변수 중요도)	22
9.5	기여도 그림	23
9.5.1	일변량 평활항 — $s(x)$	23

9.5.2	요인별 평활항 — $s(x, \text{by} = \text{factor})$	23
9.5.3	상호작용항 — $ti(x_1, x_2)$ 및 $te(x_1, x_2)$	23
9.5.4	모수항 — 선형항과 요인항	23
9.6	Correlation(상관)	23
9.7	Diagnostics(진단)	23
9.8	RCA Adjustments(RCA 조정)	23
9.9	Anova(분산분석)	24
9.10	Mgecv Output(mgecv 출력)	24
9.11	Sign Check(부호 점검)	24
9.12	Concurvity(공곡선성)	24
9.13	Basis Check(기저 점검)	24
10	데이터 내려받기	25
10.1	출력 열	25
10.2	CQA 점수	25
11	RCA 계산과 내려받기	26
11.1	RCA 대화상자 열기	26
11.2	CQA 점수 보간	26
11.3	출력 열	26
12	거래사례비교 그리드	27
12.1	비교사례 선택 대화상자	27
12.2	통합문서 구성	27
12.3	시트 보호와 잔차 칸	27
13	보고서 내려받기	28
13.1	1 단계: Quarto 보고서 생성	28
13.2	2 단계: Quarto 보고서 변환	28
13.3	보고서 내용	28
14	모형 함수 내보내기	29
14.1	R 함수	29
14.2	Python 함수	29
14.3	C++ 헤더	29
14.4	SQLite 데이터베이스	29
15	earthUI 및 glmnetUI 와의 비교	30
15.1	주요 차이	30
15.2	공통 기능	30
15.3	권장 작업 흐름	30
16	시연용 데이터셋: Appraisal_1.csv	31
16.1	설명	31
16.2	열	31
16.3	빠른 시작 예시	32
17	시스템 요건과 문제 해결	33
17.1	지원 플랫폼	33
17.2	R 버전	33
17.3	필수 R 패키지	33
17.4	선택적 의존성	33
17.5	플랫폼별 유의 사항	34

17.5.1 macOS	34
17.5.2 Windows	34
17.5.3 Linux	34
17.6 점진적 기능 축소	34
17.7 문제 해결	35
17.7.1 “PDF option is not available”	35
17.7.2 “Settings will not persist”	35
17.7.3 “Diagnostics plot unavailable”	35
17.7.4 “Report error: number of items to replace...”	35
17.7.5 포트 7880 이 이미 사용 중	35
17.7.6 Linux 에서 패키지 설치 실패	35

1 서론

1.1 역사와 배경

1.1.1 일반화 가법 모형

일반화 가법 모형 (generalized additive model, GAM) 은 Trevor Hastie 와 Robert Tibshirani 가 1990 년 저서에서 정식화한 모형이다. GAM 은 각 선형항 $\beta_j x_j$ 를 평활 함수 (smooth function) $f_j(x_j)$ 로 대체함으로써 선형 모형을 확장하며, 모형은 다음과 같은 형태가 된다:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

각 f_j 는 벌점 회귀 스플라인 (penalized regression spline) 을 이용해 데이터로부터 추정된다. 이러한 평활 함수 덕분에 모형은 사용자가 함수 형태를 미리 지정하지 않아도 비선형 관계를 포착할 수 있다. 벌점은 각 평활항의 굴곡 정도 (wiggliness) 를 제어하여, 실제 패턴을 따라갈 만한 유연성은 남겨 두면서 과적합을 방지한다.

1.1.2 mgcv 패키지

R 패키지 mgcv(Mixed GAM Computation Vehicle) 는 베스 대학교의 Simon Wood 가 개발했다 (Wood 2017). 2001 년에 처음 공개된 이후 현재는 R 에서 GAM 을 구현하는 표준으로 자리 잡았으며, 모든 R 설치본에 함께 배포된다. 그 기반 방법론은 안정적인 다중 평활 모수 추정 (Wood 2004), 박판 회귀 스플라인 (thin-plate regression spline) (Wood 2003), 빠르고 안정적인 REML 추정 (Wood 2011), 일반 평활 모형의 선택 (Wood, Pya, and Säfken 2016) 등을 다룬 일련의 논문 에 기술되어 있다. 주요 기능은 다음과 같다:

- 자동 평활도 선택 (smoothness selection): 각 항의 평활 모수는 모형 적합 과정의 일부로 추정되며, REML(제한 최대가능도), GCV(일반화 교차검증), ML(최대가능도) 중 하나를 사용한다.
- 다양한 기저 유형: 박판 회귀 스플라인 (tp), 삼차 회귀 스플라인 (cr), P-스플라인 (ps), B-스플라인 (bs) 이 있으며 각각 성질이 다르다.
- 텐서곱 평활: `te()` 와 `ti()` 는 차원의 저주 없이 변수 간 평활 상호작용을 구현한다.
- 변수 선택: `select = TRUE` 옵션은 평활항을 0 으로 수축시킬 수 있는 추가 벌점을 더하여 사실상 변수 선택을 수행한다.
- 공곡선성 진단 (concurvity diagnostics): 평활항에서 다중공선성에 대응하는 현상을 측정한다.

1.1.3 MARS(earth) 및 엘라스틱 넷 (glmnet) 과의 비교

earth 응용 프로그램과 두 개의 후처리 모듈 (glmnet 과 mgcv) 은 서로 다른 모델링 엔진을 사용한다. 다음 표는 주요 차이를 요약한 것이다:

특성	mgcv(GAM)	earth(MARS)	glmnet(엘라스틱 넷)
모형 유형	평활 비선형 (별점 스플라인)	구간별 선형 (적응 스플라인)	선형 (정칙화 적응)
비선형성	평활 함수 (박판, 삼차 등) 를 통해 자동 처리	데이터 기반 매듭점을 갖는 힌지 함수로 자동 처리	상호작용이나 기저 확장을 통해서만 가능
변수 선택	이중 별점 수축 (select=TRUE) 을 통한 선택적 기능	GCV 가지치기를 동반한 전진/후진 단계적 선택으로 내장	L_1 별점 (라쏘) 을 통해 내장
상호작용	텐서곱 평활 (te(), ti())	힌지 기반 곱 (해석 가능, 최대 3 차)	쌍별 교차곱 (해석이 어려움)
해석 가능성	평활한 부분효과 곡선으로 매우 직관적	g-함수가 변수별 구간별 선형 부분효과를 보여 줌	계수가 선형 가중치이므로 개별 설명이 쉬움
적합한 용도	평활한 비선형 관계, 유연한 모델링	비선형성과 상호작용의 자동 탐지	고차원 데이터, 변수 선택, 소표본
RCA 조정	부분효과를 통한 평활 조정	g-함수를 통한 구간별 선형 조정	변수별 선형 조정
과적합 제어	REML/GCV 평활 모수 추정	GCV 기반 가지치기	교차검증된 람다

MARS(Multivariate Adaptive Regression Splines, 다변량 적응 회귀 스플라인) 는 1991 년 Jerome Friedman 이 제안했다. 이 방법은 힌지 함수와 그 매듭점 위치를 데이터로부터 적응적으로 선택하여 구간별 선형 모형을 구축한다. R 구현은 Stephen Milborrow 의 **earth** 패키지이다.

엘라스틱 넷 (elastic net) 회귀는 Friedman, Hastie, Tibshirani (2010) 의 **glmnet** 패키지로 구현되어 있으며, 라쏘 (L_1) 와 릿지 (L_2) 별점을 결합하여 변수 선택과 계수 수축을 동시에 수행한다.

1.1.4 earthUI-mgcvUI 파이프라인

earthUI 의 자동 매듭점 탐색과 mgcvUI 의 평활 추정을 결합하면 강력한 작업 흐름이 만들어진다. earthUI 가 .rds 결과 파일을 내보내고 mgcvUI 가 이를 가져오면 다음과 같이 진행된다:

1. **earth** 가 매듭점을 찾아낸다 — MARS 는 각 변수에 대해 데이터에 기반한 힌지 위치 (변화점) 를 찾는다.
2. **mgcvUI** 가 이를 기준으로 사용한다 — earth 의 매듭점은 GAM 에서 삼차 회귀 스플라인 (cr 기저) 의 바탕이 되며, 데이터의 실제 구조를 반영하는 매듭점 위치로 평활향을 유리하게 출발시킨다.
3. **mgcv** 가 적합을 정교화한다 — 별점 스플라인 추정은 earth 의 구간별 선형 적합을 매끄럽게 다듬어, earth 가 발견한 전반적 형태를 유지하면서 평활한 부분효과 곡선을 산출한다.

이 파이프라인은 탐색적 모델링 (earth) 과 확증적 모델링 (GAM) 을 이어 주며, 두 접근법의 장점을 결합한다.

1.1.5 권장 작업 흐름

해석 가능하고 방어 가능한 모형이 필요한 부동산 감정평가 및 이와 유사한 용도에서는 다음을 권장한다:

1. **earthUI** 로 시작하여 데이터에서 중요한 변수, 비선형 관계, 상호작용을 발견한다. earth의 자동 기저 선택과 g-함수는 훌륭한 초기 통찰을 제공한다.
2. 부분효과를 더 매끄럽게 만들고 싶다면 **mgcvUI** 로 정교화한다. earthUI의 매듭점을 mgcvUI로 가져오면 구간별 선형 모형에서 평활 모형으로 자연스럽게 이행할 수 있다.
3. 공격적인 변수 선택 (라쏘)이 필요할 때, 예측변수의 개수가 관측치 수에 근접하거나 이를 초과할 때, 또는 방어 가능성을 위해 정칙화된 순수 선형 모형을 원할 때에는 **glmnetUI**를 사용한다.

1.2 mgcvUI란 무엇인가?

mgcvUI는 R mgcv 패키지를 위한 그래픽 사용자 인터페이스이다. 로컬 Shiny 응용 프로그램으로 실행되며 로그인, 서버, 계정이 필요 없다. R에서 실행한 뒤 데이터셋 (CSV 또는 Excel)을 가져오고, 모형을 설정한 다음 대화형으로 GAM을 적합하면 된다.

현재 배포판 기준으로 mgcvUI는 더 이상 별도로 설치하는 패키지가 아니다. mgcvUI는 earthUI 패키지 안에 GAM 후처리 모듈로 포함되어 배포되며 `earthUI::launch_mgcv()`로 시작한다. 응용 프로그램 자체와 본고에 설명된 모든 내용은 달라지지 않았고, 설치 및 실행 명령만 달라졌다. 세 루틴 (earth, glmnet, mgcv)은 하나의 프로젝트 트리, 하나의 설정 데이터베이스, 하나의 지원 인프라 사본을 공유한다.

이 응용 프로그램은 데이터 가져오기, 평활항 설정을 포함한 변수 구성, 백그라운드 처리 방식의 모형 적합, 진단 그림, 평활한 부분효과 곡선, 그리고 Word, PDF, HTML 형식으로 내려받을 수 있는 보고서에 이르는 완결된 작업 흐름을 제공한다.

일반화 가법 모형은 각 예측변수에 대해 선형 관계 βx 를 평활 함수 $f(x)$ 로 대체한다. 이는 다음을 의미한다:

- **함수 형태를 지정할 필요가 없다** — 각 관계의 형태는 데이터가 결정한다.
- **평활한 부분효과** — 각 예측변수의 기여는 직선이 아니라 평활한 곡선이므로, 변수가 서로 다른 값에서 반응변수에 어떤 영향을 주는지 쉽게 확인할 수 있다.
- **자동 평활도 선택** — mgcv는 REML 또는 GCV를 통해 각 곡선이 얼마나 굴곡져야 하는지를 결정하여, 적합도와 과적합 사이의 균형을 맞춘다.

1.3 세 가지 목적 모드

모든 프로젝트에는 목적 (purpose)이 있으며, 이는 사이드바 1 절 (**Project**)에서 프로젝트를 생성할 때 선택되고 이후에는 고정된다. 활성 프로젝트의 목적에 따라 어떤 도구와 인터페이스 요소가 나타나는지가 결정된다. 다른 모드로 작업하려면 해당 목적을 가진 프로젝트를 새로 만들거나 열어야 한다. 세 가지 모드는 다음과 같다:

- **General(범용)** — 모든 유형의 모집단 또는 데이터셋에 대한 GAM 회귀분석이다. 기본 모드이며, 분야별 추가 기능 없이 GAM 작업 흐름 전체를 제공한다.
- **For Appraisal(감정평가용)** — 부동산 감정평가에 맞춘 GAM 회귀분석이다. 대상 부동산 처리, 특수 열 지정, 잔차 제약 접근법 (Residual Constraint Approach, RCA) 등 단일 부동산 평가에 특화된 기능을 추가한다.
- **Market Area Analysis(시장지역 분석)** — 시장지역 연구에 맞춘 GAM 회귀분석이다. 정의된 시장 내 부동산 집단을 분석하기 위한 기능을 추가하며, 대상 행 처리는 선택 사항이다.

세 모드 모두에서 핵심 모델링 엔진은 동일하다. 언제나 `mgcv::gam()`을 통해 GAM을 적합한다. 목적 설정은 어떤 부가 도구와 인터페이스 요소를 사용할 수 있는지만 제어한다.

1.4 부동산 전용 기능

For Appraisal 또는 **Market Area Analysis**를 선택하면 mgcvUI는 부동산 분석을 위해 설계된 여러 기능을 활성화한다:

- **특수 열 지정** — 각 예측변수에는 `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions`, `display_only` 같은 특수 역할을 지정할 수 있다. 이 지정은 적합 및 출력 과정에서 해당 열을 어떻게 다룰지를 결정한다.
- **Sale Age 열** — 어떤 열이 `contract_date` 로 지정되고 기준일 (Effective Date) 이 제공 되면 `mgcvUI` 는 `sale_age` 열 (매매일과 기준일 사이의 일수) 을 계산하여 예측변수로 대체한다.
- **RCA 계산 (감정평가 모드 전용)** — 감정평가 모드에서는 모형 적합 후 `mgcvUI` 가 잔차 제약 접근법 (RCA) 출력을 계산할 수 있다. 이는 비교사례별 조정치, 순/총 조정 요약, 그리고 대상 부동산의 조정된 매매가격을 산출한다.

도움말: *General* 목적 모드는 금융, 과학, 공학, 부동산 등 어떤 데이터셋에도 사용할 수 있다. 감정평가 모드와 시장 모드는 부동산 전문가를 위한 편의 기능을 더할 뿐이다.

1.5 시작하기

`mgcvUI` 를 사용하려면:

1. R 에서 **earthUI** 를 설치한다: `install.packages("earthUI")` — `mgcv` 루틴이 함께 포함되어 있다. (개발 버전과 선택적 구성 요소에 관해서는 이번 호 `earthUI` 논문의 “Installing earthUI and vProlog” 절을 참조하라.)
2. **응용 프로그램을 실행한다:** R 에서 `earthUI::launch_mgcv()` 를 실행한다. 앱은 웹 브라우저에서 포트 7880 으로 열린다.
3. 사이드바 1 절에서 프로젝트를 만들거나 연다. 프로젝트는 목적 (General, For Appraisal, Market Area Analysis) 과 작업 위치 (국가, 주, 카운티, 도시) 및 입출력 폴더를 설정한다. 프로젝트는 공유 `regProj` 루트 아래에 저장되며 형제 앱인 `earthUI`, `glmnetUI` 와 공유된다. 다른 목적을 가진 프로젝트를 여는 것이 모드를 전환하는 방법이다.
4. 2 절에서 활성 프로젝트의 입력 폴더에 있는 CSV 또는 Excel 파일을 선택해 데이터를 가져온다. 해당 폴더에 파일을 추가한 뒤 **Refresh** 를 클릭하면 목록에 표시된다.
5. **earthUI** 에서 가져오기 (선택 사항) — 3 절에서는 `earthUI` 결과 `.rds` 파일을 가져와 `earth` 가 발견한 매듭점으로 GAM 을 초기화할 수 있다. `earthUI` 결과가 없다면 이 단계는 건너뛴다.
6. **변수를 구성한다** — 목표변수와 예측변수를 선택하고, 데이터 유형을 설정하며, 특수 열 역할을 지정하고, 변수를 요인 (factor) 또는 선형으로 표시한다.
7. **mgcv** 모수를 설정한다 — `family`, `method`, `gamma`, 기저 유형, `k`, 텐서 유형, 상호작용 및 기타 옵션을 지정한다.
8. **모형을 적합한다** — “Fit Mgcgv GAM Model” 을 클릭하고 메인 패널에서 결과를 검토한다.
9. **내보낸다** — 예측값을 Excel 로 내려받거나, Quarto 보고서를 생성하고 변환하거나, (감정평가 모드에서) RCA 조정치와 매매사례 그리드를 계산한다.

설정은 SQLite 데이터베이스에 자동으로 저장되며, 동일한 입력 파일을 다시 불러올 때 복원된다.

2 MLS 입력 데이터 요건

부동산 감정평가 및 시장분석 작업 흐름에서 입력 데이터는 보통 다중 매물 정보 서비스 (Multiple Listing Service, MLS) 내보내기 파일에서 나온다. 이 장에서는 예상되는 파일 구조와 mgcvUI 가 사용할 수 있는 열을 설명한다.

2.1 파일 형식과 구조

mgcvUI 는 CSV 와 Excel(.xlsx, .xls) 파일을 받아들인다. 가져오기 시 열 이름은 자동으로 snake_case 로 변환된다. 예를 들어 “Living SqFt” 는 living_sqft 로, “Contract Date” 는 contract_date 로, “Sale Price” 는 sale_price 로 바뀐다. 이 정규화는 작업 흐름 전반에서 열 참조를 일관되게 유지해 준다. 가져오기에 사용되는 CSV 구분자와 소수점 기호는 로케일 설정에 따라 결정된다 (3 장 “Locale & Regional Settings” 참조).

데이터 파일은 부동산 한 건이 한 행, 속성 하나가 한 열인 평면형 표여야 한다. 파일의 첫 행에는 반드시 열 머리글이 있어야 한다.

2.2 감정평가 모드에 필요한 열

mgcvUI 는 어떤 열 조합에서도 작동하지만, 완전한 감정평가 작업 흐름을 위해 다음 열이 있으면 유리하다:

열	특수 유형	용도
Sale Price	(목표변수)	모형의 반응변수
Contract Date	contract_date	sale_age(기준일로부터의 일수) 계산에 사용
Listing Date	listing_date	DOM 열이 없을 때 계약일과 함께 DOM 계산에 사용
Days on Market	dom	시장 노출 일수, 내보내기에 표시됨
Concessions	concessions	매매 양여 조건, 순매매가 = 매매가격 - 양여액
Living Area (SF)	living_area	단위면적당 잔차 (residual_sf, cqa_sf) 계산을 가능하게 함
Lot Size	lot_size	대지 규모 열
Site Dimensions	site_dimensions	대지 규모와 함께 묶임
Latitude	latitude	소수점 3 자리로 반올림
Longitude	longitude	소수점 3 자리로 반올림
Area ID	area	시장지역/근린 식별자
Actual Age	actual_age	부동산의 실제 연수
Effective Age	effective_age	부동산의 유효 연수
Address	display_only	내보내기에는 표시되고 모형에서는 제외됨

스프레드시트의 열 이름은 외국어여도 무방하다. “특수” 이름이 영어인 이유는 R 프로그램이 해당 열을 특별히 처리할 수 있도록 하기 위해서이다. 그 밖의 경우에는 사용자가 부여한 열 이름이 회귀 모형, 그래프, 그리고 감정평가를 수행할 때는 출력 보고서에 그대로 표시된다.

모든 열이 필수는 아니다. mgcvUI 는 상황에 맞추어 작동하며, 어떤 열이 없으면 해당 기능만 생략된다. 다만 부동산 가격 모형에서는 다음 열들을 적극 권장한다:

1. **Sale Age** — 계약 매매일과 기준일 사이의 일수. 여러 해에 걸친 매매 이력을 사용할 경우 sale_age 가 핵심 역할을 하는 경우가 많다.
2. **Living Area** — “Living Sqft”, “GLA”(gross living area, 총 거주면적) 등의 이름으로도 불린다.
3. **Total Bath Count** — 전체, 1/4, 1/2, 3/4 욕실을 모두 합한 개수.
4. **Garage Bays** 또는 **Garage Area** — 차고 주차 대수 또는 면적.
5. **Lot Size** — 부동산의 토지 면적.
6. 지리적 가격 변동을 반영하기 위한 **Longitude**, **Latitude**, 그리고 가능하다면 **Area ID**.

2.3 특수 열 명명 규칙

mgcvUI 는 열 이름이 아니라 특수 유형 지정으로 열을 식별한다. MLS 내보내기 파일에서 열 이름은 원하는 대로 지어도 된다. 중요한 것은 변수 구성 표 (6 장) 에서 올바른 특수 유형을 지정하는 일이다.

2.4 데이터 품질과 완전성

- **결측값 (NA):** 예측변수나 목표변수 열에 NA 가 있는 행은 적합 전에 자동으로 제거된다. mgcvUI 는 모형에 사용되는 열만 골라 데이터를 부분 선택하므로, 사용하지 않는 열의 NA 는 행 삭제를 일으키지 않는다.
- **수치형 열:** 매매가격, 거주면적, 대지 규모 등과 같은 항목은 반드시 수치형이어야 한다.
- **요인형 열:** 텍스트, TRUE/FALSE, R 요인 (factor) 열은 자동으로 범주형으로 처리되며, 모형에는 요인으로만 들어갈 수 있다. 수치형 열 (예: 수치로 된 지역 코드나 양식 코드) 은 **Factor** 체크박스를 선택했을 때에만 요인이 된다. mgcvUI 는 수치의 값 범위만 보고 범주형이라고 추론하지 않으므로, `bath_count` 같은 이산 수치형 예측변수는 연속형 (평활 처리) 으로 유지된다. 특수 역할 지정은 요인 여부 결정과 무관하다.

도움말: 가져오기 후 예측변수 표의 NA 열을 살펴보자. 결측값이 많은 열은 행 삭제를 유발할 수 있다. NA 가 50% 이상인 열은 빨간색으로 표시되고 자동으로 제외된다.

2.5 대상 행의 위치

Appraisal 모드에서는 1 행이 반드시 대상 부동산이어야 한다. 나머지 행은 모두 비교 매매사례이다. 대상 행은 모형 적합에서 제외된다. 적합 후에도 모형은 대상 행에 대한 예측값을 생성한다.

Market Area Analysis 모드에서는 대상을 1 행에 두는 것이 선택 사항이다.

General 모드에는 특별한 행 처리가 없으며 모든 행이 동등하게 취급된다.

3 범용 모드

3.1 개요

범용 (General Purpose) 모드는 mgcvUI 를 실행했을 때의 기본 모드이다. 이 모드는 부동산에 국한되지 않고 어떤 데이터셋에 대해서도 완전한 GAM 작업 흐름을 제공한다. 과학 데이터, 재무 분석, 공학 연구를 비롯해 평활한 비선형 관계가 예상되는 모든 회귀 문제에 mgcvUI 를 사용할 수 있다.

General 모드에서는 부동산 전용 기능 (특수 열, 매매 경과일, RCA) 이 인터페이스에서 생략된다. 사이드바는 변수 선택, 모수 설정, 모형 적합, 내보내기에 집중하도록 간소화된다.

3.2 사이드바 작업 흐름

사이드바는 번호가 매겨진 접이식 절들로 구성되어 프로젝트 선택부터 보고서 내보내기까지를 안내한다. 대부분의 절은 활성 프로젝트가 열려 있어야 나타나며, 몇몇 절은 프로젝트의 목적에 따라 달라진다. 출력 폴더는 활성 프로젝트에서 유도되며 별도의 출력 폴더 입력란은 없다.

1. Project(프로젝트) — 프로젝트를 만들거나 연다. 프로젝트는 목적과 작업 위치 (국가, 주, 카운티, 도시), 그리고 입력 폴더 (<os>_in/) 와 출력 폴더를 설정한다. 프로젝트는 공유 regProj 루트 (기본값 ~/regProj, REGPROJ_ROOT 환경변수나 Settings 드롭다운으로 변경 가능) 아래에 저장되고 형제 앱인 earthUI, glmnetUI 와 공유되므로, 세 앱의 모형이 나란히 놓이게 된다. 전환하려면 **Close Project** 를 클릭한다.

2. Import Data(데이터 가져오기) — 활성 프로젝트의 입력 폴더에서 CSV 또는 Excel 파일을 선택한다. Finder/탐색기로 폴더에 파일을 추가한 뒤 **Refresh** 를 클릭하면 목록에 표시된다. 시트가 여러 개인 Excel 파일에는 시트 선택기가 나타난다. 열 이름은 자동으로 snake_case 로 변환된다.

3. Import from earthUI (optional)(earthUI 에서 가져오기, 선택 사항) — earthUI 결과 .rds 파일을 가져와 earth 가 발견한 매듭점 위치 (및 변수 선택 상태) 로 GAM 을 초기화한다. 이 단계는 선택 사항이며, 가져올 earthUI 결과가 없다면 건너뛰는다.

4. Variable Configuration(변수 구성) — 목표변수 선택기, 반응 변환 (none, log, log10), 그리고 Include, Factor, Linear 체크박스가 있는 예측변수 표로 구성된다. Special 열은 Appraisal 과 Market 모드에서만 나타난다. 자세한 내용은 6 장을 참조하라.

5. MgcV Call Parameters(mgcV 호출 모수) — 모수 프리셋, family, method, gamma, 교차검증, select, 기저 유형, k, 텐서 유형, 상호작용 행렬, 고급 모수 등 모든 모형 설정을 다룬다. 전체 모수 참조는 7 장에 있다.

6. Fit MgcV GAM Model(mgcV GAM 모형 적합) — 모형을 실행하는 버튼이다. 결과는 오른쪽 탭에 표시된다.

7. Download Output(출력 내려받기) — 예측값, 잔차, CQA 점수, 변수별 기여도를 Excel 파일로 프로젝트의 출력 폴더에 내보낸다. 모형을 적합한 뒤에는 모든 목적 모드에서 사용할 수 있다.

8. Calculate RCA Adjustments & Download(RCA 조정치 계산 및 내려받기) — 각 비교사례에 대한 변수별 RCA 조정치를 계산하고 CQA 를 통해 대상의 잔차를 보간한다. Appraisal 과 Market 모드에서 적합 후에만 나타난다.

9. Generate Sales Grid & Download(매매사례 그리드 생성 및 내려받기) — 총조정률을 기준으로 비교사례를 추천하면서 거래사례비교 그리드를 생성한다. Appraisal 모드에서만 나타난다.

10. Generate Quarto Report(Quarto 보고서 생성) — 모든 그림과 표를 포함한 자립형 Quarto(.qmd) 보고서 묶음을 프로젝트의 출력 폴더에 작성한다. 모형을 적합한 뒤 사용할 수 있다.

11. Convert Quarto Report(Quarto 보고서 변환) — 생성된 .qmd 묶음을 HTML, Word, PDF 로 렌더링한다. PDF 는 LaTeX 설치가 필요하며, 설치가 감지되지 않으면 PDF 옵션이 숨겨진다. 13 장을 참조하라.

다른 모드에서는 목적별 단계가 생략되고 나머지 단계의 번호가 다시 매겨진다. Market 모드에서는 매매사례 그리드 단계가 없고, General 모드에서는 RCA 와 매매사례 그리드 단계가 모두 없다.

3.3 earthUI 에서 가져오기 (선택 사항)

사이드바 3 절 **Import from earthUI**에서는 earthUI 결과 .rds 파일을 가져올 수 있다. 이를 통해 earth-mgcv 파이프라인이 가능해진다. 즉 earth 가 데이터로부터 찾아낸 매듭점 위치가 GAM 평활항의 기준점이 된다.

earthUI 결과를 가져오면 mgcvUI 는 다음을 수행한다:

- 목표변수, 결정계수 (R-squared), 항의 개수, 변수별 매듭점 위치를 보여 주는 요약을 표시한다.
- “Earth Pipeline” 모수 프리셋 (삼차 회귀 스플라인 기저, $\gamma = 1.4$) 을 자동으로 선택한다.
- 목표변수를 earth 의 목표변수와 일치시킨다.
- earth 의 예측변수에 대해 Include 를 미리 선택한다.
- earth 의 범주형 변수에 대해 Factor 를, earth 의 선형 예측변수에 대해 Linear 를 미리 선택한다.
- earth 가 탐지한 상호작용을 미리 채운다 (상호작용 행렬에서 노란색으로 강조되며 수동 변경이 잠긴다).
- earth 의 매듭점을 CSV 파일로 내려받을 수 있는 **Export Knots CSV** 버튼을 제공한다.

그러면 변수 구성 표에는 변수 선택 (Include, Linear, Factor) 과 스플라인 매듭점이 earth 모형에서 초기값으로 설정되었다는 안내가 나타난다. glmnet 적합이 earth 의 기저에 묶여 있는 glmnetUI 와 달리, mgcvUI 는 현재 선택 상태를 바탕으로 GAM 을 적합한다. earth 는 출발점만 제공할 뿐이며 모든 제어 항목은 계속 편집할 수 있다.

Clear 버튼은 가져온 earth 데이터를 제거하고 독립 실행 모드로 되돌린다.

3.4 메인 패널 탭

데이터를 가져오면 메인 패널에 다음 탭들이 나타난다 (모형에 의존하는 탭은 적합 후에 채워진다):

탭	내용
Data	가져온 데이터의 대화형 DataTable. 감정평가 및 시장 모드에서는 대상 부동산 표와 비교 매매사례 표로 분리된다.
Equation	MathJax 로 렌더링된 전체 GAM 식과 평활 함수 정의 표.
Summary	주요 지표 (R^2 , $CV R^2$, 설명된 이탈도, AIC, n, 평활항 개수) 와 평활항/모수항 표.
Variable Importance	F-통계량 (평활항) 과 $ t $ 값 (모수항) 의 가로 막대그래프 및 순위 표.
Contribution	신뢰 구간 띠와 earth 매듭점 표식을 포함한 대화형 plotly 평활 부분효과 곡선.
Correlation	수치형 예측변수 간 상관 히트맵 (적합 전에도 사용 가능).
Diagnostics	잔차 대 적합값, Q-Q 그림, 히스토그램, 반응변수 대 적합값 (<i>gratia</i> 사용). 실측값 대 예측값 산점도.
RCA Adjustments	RCA 분석 히스토그램 (감정평가/시장 모드에서 계산 후).
Anova	모수항과 평활항을 함께 담은 ANOVA 표.
Mgcv Output	원시 텍스트: 소요 시간, 식, 모형 요약, family, 공곡선성.
Sign Check	earth 와 GAM 의 방향 일치 여부 점검 (earth 를 가져온 경우).
Concurvity	색상으로 구분된 전체 및 쌍별 공곡선성 행렬.
Basis Check	<code>gam.check()</code> 의 기저 차원 적정성 검증.

3.5 설정의 유지

mgcvUI 는 입력 파일 이름을 키로 하여 설정을 SQLite 데이터베이스에 자동 저장한다. 같은 파일을 다시 불러오면 목표변수 선택, 예측변수 체크박스, 데이터 유형, mgcv 모수, 상호작용 행렬 등 모든 설정이 복원된다. 목적은 전역 설정이 아니라 활성 프로젝트의 속성이며, 마지막으로 사용한 입력 파일은 프로젝트별로 기억되어 프로젝트를 다시 열 때 자동으로 불러와진다.

3.6 다크 모드

오른쪽 위 모서리의 달/해 아이콘을 클릭하면 Nord Light 와 Nord Dark 테마를 전환할 수 있다. 테마 설정은 localStorage 에 저장되어 세션이 바뀌어도 유지된다. 모든 UI 요소가 선택한 테마에 맞추어 바뀐다.

3.7 로케일과 지역 설정

mgcvUI 는 국가 기반 로케일 체계를 통해 국제적인 숫자 및 CSV 형식 관례를 지원한다. 제목 표 시줄의 **Settings** 드롭다운에는 30 개국 이상을 지원하는 **Country** 와 **Paper** 선택기가 있다. 각 프리셋은 다음을 설정한다:

- **CSV 구분자** — 미국/영국/일본은 쉼표 (,), 유럽 대부분은 세미콜론 (;).
- **소수점 기호** — 마침표 (.) 또는 쉼표 (,).
- **천 단위 구분자** — 쉼표 (미국/영국/일본), 마침표 (독일/이탈리아/스페인), 공백 (핀란드/프랑스/폴란드/발트 3 국), 또는 아포스트로피 (스위스).
- **용지 크기** — Letter(미국/캐나다/멕시코) 또는 A4(그 외 지역).

Save 를 클릭하면 로케일 설정이 전역으로 저장된다. Import Data 절에도 관례가 다른 일회성 가져오기를 위한 파일별 **Locale** 선택기가 있다. 같은 Settings 드롭다운에는 **regProj** 루트 폴더 입력란이 있으며, 이는 mgcvUI, earthUI, glmnetUI 가 함께 사용하는 공유 프로젝트 트리의 위치를 바꾼다.

도움말: 그림 축과 기울기 라벨의 숫자 형식은 로케일에 맞추어 자동으로 조정된다. 독일 로케일은 천 단위에 마침표 (200.000), 핀란드는 공백 (200 000), 스위스는 아포스트로피 (200'000) 를 사용한다. 통화 기호는 표시되지 않으며, *mgcvUI* 는 통화에 중립적이다.

4 감정평가 모드

활성 프로젝트의 목적이 **For Appraisal** 일 때 mgcvUI 는 단일 부동산 평가에 맞추어 스스로를 구성한다. 3 장에서 설명한 모든 기능은 그대로 사용할 수 있으며, 이 장에서는 감정평가 전용으로 추가되는 부분만 다룬다.

4.1 대상 행의 처리

감정평가 모드에서는 데이터셋의 1 행이 대상 부동산이고 나머지 행은 모두 비교 매매사례이다. 입력 파일은 그에 맞게 구성되어야 한다 (2 장 참조). 대상의 매매가격은 비워 두거나 아무 값이나 넣어도 된다. mgcvUI 가 적합 과정에서 이를 자동으로 NA 로 처리하기 때문이다.

적합 후 모형은 대상 행에 대한 예측값을 생성하며, 이것이 RCA 조정 작업 흐름의 출발점이 된다.

4.2 기준일과 매매 경과일

감정평가 및 시장 모드에서는 변수 구성 절에 **Effective Date**(기준일) 입력란이 나타난다 (기본값은 오늘 날짜). Special 드롭다운에서 어떤 열을 `contract_date` 로 지정하면 mgcvUI 는 `sale_age` 열, 즉 각 매매의 계약일과 기준일 사이의 정수 일수를 계산한다. 이 열은 예측변수로 추가된다.

기준일이 바뀌면 `sale_age` 는 자동으로 다시 계산된다.

4.3 특수 열 지정

감정평가 및 시장 모드에서는 변수 구성 표의 각 예측변수마다 **Special** 드롭다운이 나타난다. 특수 유형과 그 효과에 관한 전체 참조는 6 장에 있다.

4.4 RCA 조정 개요

Calculate RCA Adjustments & Download 버튼 (감정평가 및 시장 모드에서 적합 후 표시됨) 은 대상 대비 각 비교사례에 대해 시장에서 도출된 조정치를 계산한다. 전체 RCA 작업 흐름은 11 장에서 설명한다.

도움말: 대상에 부여하는 *CQA* 점수는 잔차 분포 중 얼마만큼을 대상에 귀속시킬지를 좌우한다. 점수 5.00 은 대상을 비교사례들의 중앙값에 위치시킨다.

5 시장지역 분석 모드

활성 프로젝트의 목적이 **Market Area Analysis** 일 때 mgcvUI 는 감정평가 모드와 동일한 부동산 전용 기능 (특수 열, 매매 경과일) 을 제공하지만, 단일 대상의 평가보다는 부동산 집단의 분석에 초점을 둔다.

5.1 감정평가 모드와의 차이

- 대상 행 처리가 유연하다 — 시장 모드에서는 1 행이 자동으로 대상 부동산으로 취급되지 않는다.
- 매매사례 그리드가 없다 — “Generate Sales Grid & Download” 단계를 사용할 수 없다.

5.2 시장 모드를 사용하는 경우

시장지역 분석 모드는 다음과 같은 경우에 적합하다:

- 비선형적인 가치 결정 요인을 파악하기 위해 근린이나 시장지역에 대한 GAM 모델을 구축할 때
- 면적, 연수, 대지 규모, 위치 같은 변수가 부동산 집단 전반의 매매가격에 어떤 영향을 주는지 분석할 때
- 시장 상황 분석이나 근린 확정을 뒷받침할 근거를 준비할 때
- 특수 열 기능 (매매 경과일) 은 사용하되 매매사례 그리드 작업 흐름은 필요하지 않은 데이터셋을 다룰 때

도움말: 시장 모드는 특수 열 기능은 쓰고 싶지만 단일 부동산 평가 작업 흐름은 필요 없는 일반적인 부동산 회귀분석에도 유용하다.

6 변수 선택

사이드바 4 절 **Variable Configuration** 은 어떤 열을 모형에 참여시킬 것인지, 그리고 그 열을 어떻게 다룰 것인지를 정하는 곳이다.

6.1 목표변수

4 절 상단의 **Target (response) variable** 드롭다운에는 데이터셋의 모든 수치형 열이 나열된다. 그중 하나를 반응변수 (예: 매매가격) 로 선택한다. 목표변수로 지정된 열은 예측변수 목록에서 자동으로 제외된다.

6.2 반응변수 변환

목표변수 선택기 아래의 **Response Transform** 드롭다운은 세 가지 선택지를 제공한다:

- **None (raw)** — 반응변수를 원래대로 사용한다. 예측변수와 반응변수의 관계가 원래 척도에서 대체로 가법적일 때 적절하다.
- **Log (natural)** — 적합 전에 자연로그 변환을 적용한다. 역변환 시 조정치가 절대액 (\$) 이 아니라 비율 (%) 이 된다. 1 만 달러의 조정이 가격 수준에 따라 전혀 다른 의미를 갖는 매매가격에 특히 유용하다.
- **Log10** — 밑이 10 인 로그 변환을 적용한다.

도움말: 로그 변환은 시점 조정을 절대액 (\$) 대신 비율 (%) 로 만든다. 가격대가 넓게 분포한 부동산 모형에서는 로그 변환이 더 안정적인 잔차를 낳는 경우가 많다.

로그 변환을 선택하면 반응변수의 ≤ 0 값은 자동으로 걸러진다. 모든 출력 (예측값, 기여도, 잔차) 은 자동으로 원래 척도로 역변환된다.

6.3 예측변수 표

목표변수 선택기 아래에는 나머지 모든 열을 나열하는 스크롤 가능한 표가 있다. 체크박스 열은 공간 절약을 위해 세로로 회전된 머리글을 사용한다. 열 순서는 다음과 같다:

열	설명
Variable	열 이름 (고정폭 글꼴, 길면 말줄임표로 잘림). 테두리가 있는 칸.
Type	데이터 유형 드롭다운: numeric, integer, character, factor, logical, Date, POSIXct. 테두리가 있는 드롭다운.
Include	체크박스 — 이 열을 모형의 예측변수로 포함한다. 세로 머리글.
Factor	체크박스 — 범주형 요인으로 취급한다 (수준마다 별도의 계수를 만든다). 세로 머리글.
Linear	체크박스 — 선형 관계를 강제한다 (평활 없이 βx 만). 세로 머리글.
Special	드롭다운 (감정평가/시장 모드 전용) — 아래의 특수 열 유형 참조를 보라. 테두리가 있는 드롭다운.
NAs	결측값의 개수와 비율. 색상으로 구분됨: 빨강 ($\geq 50\%$), 주황 ($\geq 20\%$), 회색 (> 0). 테두리가 있는 칸.

Factor 대 **평활** 대 **Linear**: 기본적으로 포함된 수치형 변수에는 평활항 $f(x)$ 가 부여된다. **Factor** 를 선택하면 범주형 항 (수준마다 계수 하나) 이 만들어진다. **Linear** 를 선택하면 평활 대신 단순 선형항 βx 가 만들어진다. Factor 와 Linear 를 모두 표시한 변수는 Factor 로 취급된다.

6.4 데이터 유형 탐지와 재지정

mgcvUI 는 가져오기 시 데이터 유형을 자동으로 탐지한다. 수치형, 정수형, 논리형, 요인형, 날짜형 열이 인식된다. **Type** 드롭다운을 바꾸면 탐지 결과를 임의로 재지정할 수 있다. 유형을 바꾸면 해당 열이 모형에서 다루어지는 방식도 달라진다.

6.5 특수 열 유형 참조

감정평가 및 시장 모드에서 **Special** 드롭다운은 다음 선택지를 제공한다:

가중:

- **weight** — 관측치 가중치 열 (하나만 지정 가능. 가중치가 0 인 행은 적합에서 제외됨)

날짜 및 시간 유형:

- **contract_date** — 기준일로부터 **sale_age** 를 자동으로 계산하도록 한다
- **sale_age** — 이미 계산되어 있는 매매 경과일 열 (기준일로부터의 일수) 을 식별하며, 그대로 사용된다
- **listing_date** — 시장 노출 일수를 계산할 때의 대체 수단으로 사용된다
- **dom** — 시장 노출 일수 (Days on Market) 열을 식별한다

레코드 유형:

- **sale_type** — 대상과 비교 매매사례를 구분하는 레코드 유형 열을 식별한다

금액 유형:

- **concessions** — 매매 양여 조건 (매도인 크레딧, 매수인 유인책 등) 을 식별한다

규모 및 위치 유형:

- **latitude** — 값이 자동으로 소수점 3 자리로 반올림된다
- **longitude** — 위도와 동일한 반올림 처리
- **area** — 시장지역 또는 근린 식별자
- **living_area** — 단위면적당 잔차 계산 (**residual_sf**, **cqa_sf**) 을 가능하게 한다
- **lot_size** — 대지 규모 열
- **site_dimensions** — 대지 규모와 함께 묶인다

연수 유형:

- **actual_age** — 부동산의 실제 연수 열
- **effective_age** — 부동산의 유효 연수

표시 유형:

- **display_only** — 해당 열은 Excel 내보내기에는 포함되지만 모형 적합에서는 완전히 제외된다. 주소 항목, MLS 번호, 필지 ID 등의 참조 데이터에 사용한다.

도움말: *display_only* 로 지정된 열은 데이터셋에 그대로 남아 *Excel* 내보내기에 나타나지만 예측변수 표에서는 완전히 제외된다. *ID* 열, 주소, 기타 참조 항목에 사용하라.

7 모수 선택

사이드바 5 절 **Mgcv Call Parameters** 는 GAM 의 모든 설정 항목에 접근할 수 있게 해 준다. 각 모수에는 설명 풍선말이 달린 파란색 도움말 아이콘 (?) 이 있다.

7.1 모수 프리셋

상단의 드롭다운에는 흔한 작업 흐름에 맞는 합리적 기본값을 설정해 주는 두 가지 프리셋이 있다:

프리셋	기저	설정
Standalone (탐색)	tp (박판)	k = 자동 (10), gamma = 1.2, select = TRUE
Earth Pipeline (정교화)	cr (삼차 회귀)	k = 자동 (10), gamma = 1.4, select = TRUE

earthUI 매듭점을 가져오면 **Earth Pipeline** 프리셋이 자동으로 선택된다. earth 매듭점을 통합하려면 정확한 매듭점 위치를 지정할 수 있는 삼차 회귀 스플라인 (**cr**) 기저가 필요하다.

7.2 Family(분포족)

반응변수의 분포족을 선택한다:

Family	용도
gaussian	연속형 반응변수 (예: 매매가격). 가장 흔하다.
Gamma	분산이 평균에 비례하는 양의 연속형 반응변수.
poisson	계수 자료 (예: 매매 건수).
binomial	이항 결과.
inverse.gaussian	오른쪽으로 크게 치우친 양의 자료.

7.3 Method(방법)

평활 모수 추정 방법이다:

Method	설명
REML	제한 최대가능도 (기본값, 권장). 과적합에 가장 강건하다.
GCV.Cp	일반화 교차검증. 다소 더 복잡한 모형을 고르는 경향이 있다.
ML	최대가능도. REML 과 비슷하지만 평활이 부족해질 수 있다.
P-REML	피어슨 REML 변형.
P-ML	피어슨 ML 변형.
GACV.Cp	일반화 근사 교차검증.

7.4 Gamma(평활 승수)

gamma 모수는 평활 기준에서 유효자유도에 곱해져 더 매끄러운 (덜 굴곡진) 적합을 유도한다. 기본값은 1.2 이다.

값	효과
1.0	표준 평활

값	효과
1.2-1.4	기본보다 약간 더 매끄러움 (권장)
2.0 이상	훨씬 더 매끄러움 — 잡음이 많은 자료나 소표본에 적합

gamma 값이 클수록 복잡도에 더 무거운 벌점을 부과하여 과적합을 막아 준다. Earth Pipeline 프리셋은 earth 의 매듭점을 정교화할 때 추가 평활을 위해 1.4 를 사용한다.

7.5 교차검증

체크하면 (기본값) mgcvUI 는 적합 후 10 겹 교차검증을 실행하여 CV 결정계수를 계산한다. 이는 표본 외 예측력에 대한 정직한 추정치를 제공한다.

7.6 Select(수축)

체크하면 (기본값) 각 평활항에 그 항을 0 으로 수축시킬 수 있는 추가 벌점이 더해진다. 이로써 자동 변수 선택이 가능해지며, 중요하지 않은 평활항은 사실상 모형에서 제거된다. gam() 의 select = TRUE 로 구현된다.

7.7 기본 기저

평활항에 사용할 스플라인 기저 함수 유형이다:

기저	전체 명칭	설명
tp	박판 회귀 스플라인	기본값. 등방적이며 매듭점 배치가 필요 없다. 범용으로 무난한 선택이다.
cr	삼차 회귀 스플라인	매듭점을 명시적으로 배치할 수 있다. earth 매듭점 통합에 필수적이다.
ps	P-스플라인	차분 벌점을 갖는 등간격 B-스플라인 기저. 계산 효율이 좋다.
bs	B-스플라인	유연한 B-스플라인 기저.

7.8 기본 k(기저 차원)

기저 차원 k 는 각 평활항이 가질 수 있는 최대 굴곡 정도를 제어한다. 0(기본값) 은 “자동” 을 뜻하며, mgcvUI 는 상황에 따라 $k = 10$ 또는 earth 매듭점의 개수를 사용한다.

- 작은 $k(3-5)$: 거의 직선에 가까운 매우 매끄러운 곡선.
- 중간 $k(8-15)$: 적당한 유연성 (대개 이 정도면 충분하다).
- 큰 $k(20 이상)$: 매우 유연함 — 강한 정칙화가 없으면 과적합 위험이 있다.

도움말: Summary 탭의 유효자유도 (EDF) 는 각 평활항이 실제로 사용하는 자유도가 얼마인지를 알려 준다. EDF 가 $k - 1$ 에 가깝다면 기저가 너무 작을 수 있으므로 k 를 늘려라. EDF 가 k 보다 훨씬 작다면 평활항에 벌점이 잘 작동하고 있는 것이다.

7.9 텐서 유형

연속형 변수 간 상호작용에는 두 가지 텐서곱 유형을 사용할 수 있다:

유형	함수	설명
ti	ti(x1, x2)	텐서 상호작용 — 상호작용 효과만 모형화하고 주효과는 일변량 $s()$ 항이 따로 담당한다. 해석 가능성 면에서 권장된다.
te	te(x1, x2)	텐서곱 — 주효과를 포함한 결합 효과 전체를 모형화한다. RCA 조정을 위해 분해하기가 더 어렵다.

7.10 허용 상호작용

접이식 **Allowed Interactions** 절에는 포함된 모든 평활 (비선형, 비요인) 예측변수에 대한 상삼각 체크박스 행렬이 표시된다. 각 체크박스는 해당 변수 쌍 사이의 텐서곱 평활을 활성화한다.

- **Allow All / Clear All** — 일괄 전환 버튼.
- 변수 이름 클릭 — 그 변수에 관한 모든 상호작용을 전환한다.
- 변수 이름 오른쪽 클릭 — “Block from main effect” — 해당 변수를 상호작용 전용 (일변량 평활 없음) 으로 표시한다.

earthUI 매듭점을 가져오면 earth 가 탐지한 상호작용이 미리 선택되고 잠긴다 (노란 배경으로 강조됨). earth 모형이 degree = 1(상호작용 없음) 을 사용했다면 안내 메시지가 표시된다.

주의

텐서 상호작용을 활성화하면 모형 복잡도가 크게 증가한다. 해당 분야의 지식이나 earth 가 탐지한 근거가 뒷받침될 때에만 상호작용을 사용하라. 그 결과로 나오는 조정치가 합리적이고 설명 가능한지 반드시 확인하라.

7.11 요인별 평활 상호작용

요인 변수와 평활 변수가 모두 포함되면 별도의 **Factor-by-Smooth Interactions** 행렬이 나타난다. 각 체크박스는 $s(x, by = factor)$ 항, 즉 요인 수준마다 별개의 평활 곡선을 만든다.

7.12 고급 모수

접이식 “Advanced” 절에서는 추가 설정을 노출한다:

- **Optimizer** — 평활 모수 추정 알고리즘: `outer/newton`(기본값), `outer/bfgs`, 또는 `efs`(확장 Fellner-Schall).
- **Scale** — 척도 모수. 0(기본값) 은 데이터로부터 추정함을 뜻한다. 척도가 고정된 모형에서는 알려진 값을 지정한다.
- **Discrete covariate method** — 대용량 데이터셋을 위한 빠른 이산화 적합 (`discrete = TRUE`).
- **Threads** — 병렬 스레드 수 (1-32). `discrete = TRUE` 일 때에만 효과가 있다.

8 모형 적합

8.1 Fit 버튼

사이드바 6 절에는 **Fit Mgcgv GAM Model** 버튼이 있다. 이를 클릭하면 현재 설정으로 모형이 실행된다.

8.2 적합 중에 일어나는 일

Fit 을 클릭하면 mgcvUI 는 다음을 수행한다:

1. 설정을 검증한다 — 예측변수가 하나 이상 있고 완전한 행이 10 개 이상인지 확인한다. 완전한 행이 50% 미만이면 경고하고 NA 를 가장 많이 유발하는 열들을 알려 준다.
2. 데이터를 준비한다 — 모형에 필요한 열만 남기고, NA 가 있는 행을 제거하며, (지정된 경우) 가중치를 적용하고, (감정평가 모드에서) 대상 행을 제외하며, 반응변수 변환을 적용한다.
3. GAM 식을 구성한다 — 수치형 예측변수에는 평활항 $s(x, bs="cr", k=10)$ 을, 범주형에는 요인항을, Linear 로 표시된 변수에는 선형항을, 선택된 상호작용에는 텐서곱을 만든다. earth 매듭점은 cr 기저 항의 명시적 매듭점 위치로 삽입된다.
4. 예외 상황을 처리한다 — 상수 변수를 제거하고, 이항 예측변수를 요인으로 변환하며, k 를 고유값 개수에서 1 을 뺀 값으로 제한하고, 매듭점이 더 필요할 때에는 데이터 분위수로 earth 매듭점을 보강한다.
5. GAM 을 적합한다 — 구성된 식, family, method, gamma, select 및 기타 모수와 함께 `mgcv::gam()` 을 호출한다.
6. 교차검증을 실행한다 (활성화된 경우) — 표본 외 결정계수를 계산하는 10 겹 CV.
7. 모든 결과 탭을 갱신한다 — Summary, Equation, Contribution 그림, Diagnostics 를 비롯한 모든 탭이 채워진다.

8.3 백그라운드 적합과 진행 상황

`callr` 패키지를 사용할 수 있으면 모형 적합이 백그라운드 프로세스에서 실행되어 앱이 계속 반응한다. 어두운 테마의 터미널 형태 진행 오버레이에는 다음이 표시된다:

- 경과 초/분을 보여 주는 타이머.
- 색상으로 구분된 추적 로그: 상태는 파랑, CV 겹 진행은 노랑, 오류는 빨강, 결과는 초록.
- 필요할 때 적합 프로세스를 중단하는 **Abort** 버튼.
- 완료 후 나타나는 **Close** 버튼.

`callr` 를 사용할 수 없으면 적합이 동기적으로 실행된다 (적합이 끝날 때까지 앱이 멈춘다).

적합에 성공하면 Fit 버튼에 흰색 체크 표시가 나타나고 상태 표시줄에 “R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X” 가 표시된다.

9 결과 탭

9.1 Data(데이터)

가져온 데이터를 가로 스크롤이 가능하고 한 페이지에 15 행씩 표시되는 대화형 DataTable 로 보여 준다. 감정평가 및 시장 모드에서는 탭이 **Subject Property** 표 (1 행) 와 **Comparable Sales** 표 (나머지 행) 로 나뉘며, General 모드에서는 하나의 표만 표시된다.

각 칸은 한 줄로 유지되고 열 너비에 맞추어 잘리므로, 부동산 비고란처럼 긴 자유 서술 항목 이 행을 아래로 늘어뜨리지 않는다. 한 번 클릭하면 행이 선택되고, 어떤 칸이든 두 번 클릭하면 잘리지 않은 전체 내용을 보여 주는 팝업이 열린다.

9.2 Equation(식)

적합된 GAM 모형을 두 부분으로 나누어 표시한다:

모형 식 (MathJax 로 렌더링됨):

- 좌변에는 해당하는 변환 (log, log10) 과 연결함수를 적용한 반응변수가 표시된다.
- 우변에는 절편값, f_i (변수) 형태의 평활항, 수준 개수와 함께 변수별로 묶인 요인항, 그리고 계수값을 포함한 선형항이 표시된다.
- family 와 method 는 식 위에 표시된다.

평활 함수 정의 (Smooth Function Definitions) 표:

- 항의 식 문자열 (예: `s(living_sqft, bs="cr", k=8)`)
- 변수 이름
- 유형 (s, te, ti 또는 linear)
- 기저 함수 유형
- k 값
- earth 매듭점 위치 (가져온 경우)

9.3 Summary(요약)

상단에 여섯 개의 지표 카드가 표시된다:

- **R-squared** — 설명된 이탈도의 비율
- **CV R-squared** — 교차검증 결정계수 (CV 를 활성화한 경우)
- **Dev. Explained %** — 설명된 이탈도의 백분율
- **AIC** — 아카이케 정보기준
- **n** — 사용된 관측치 수
- **Smooths** — 평활항의 개수

카드 아래에는 다음이 표시된다:

- **평활항 표** — 항, EDF(유효자유도), Ref.df, F 통계량, p-값
- **모수항 표** — 항, 추정값, 표준오차, t-값, p-값

9.4 Variable Importance(변수 중요도)

각 모형 항의 상대적 중요도를 보여 주는 가로 막대그래프이다:

- **평활항**: F-통계량 순위 (파란 막대)
- **모수항**: $|t|$ 값 순위 (초록 막대)

그래프 아래의 DataTable 에는 항, 유형, EDF, 통계량, p-값이 통계량 내림차순으로 정렬되어 표시된다.

9.5 기여도 그림

예측값에 기여하는 모든 모형 항에는 대화형 **plotly** 기여도 그림이 있으며, 테두리가 있는 카드 안에 표시된다. 이 탭은 모든 유형의 항에 대한 그림을 보여 준다:

9.5.1 일변량 평활항 — $s(x)$

- 95% 신뢰 구간 띠를 두른 파란 선
- 부분 잔차를 보여 주는 회색 산점 (로그 변환하지 않은 모형에서)
- earth 매듭점 위치의 빨간 파선 수직선 (earth 를 가져온 경우)
- 마우스오버 푼션말: 변수값, 기여도 (\$), 95% 신뢰구간, 변화율 (단위당 기울기)

로그 변환 모형에서는 y 축이 $\bar{y} \times (e^{f(x)} - 1)$ 공식을 통해 달러 기여도로 역변환되어, 곡선을 곧바로 금액 단위로 해석할 수 있다.

위도/경도 변수에서는 기울기가 1 도당이 아니라 0.001 도당으로 표시되는데, 지리적으로 그 편이 더 의미 있는 증분이기 때문이다.

9.5.2 요인별 평활항 — $s(x, \text{by} = \text{factor})$

요인 수준마다 색이 다른 선 하나씩을 한 그림에 겹쳐 표시한다. 예측변수의 효과가 집단 (예: 서로 다른 근린) 에 따라 어떻게 달라지는지 확인하는 데 유용하다.

9.5.3 상호작용항 — $ti(x_1, x_2)$ 및 $te(x_1, x_2)$

2 차원 기여도 곡면을 보여 주는 히트맵이다. 색 척도는 빨강 (음) → 흰색 (0) → 파랑 (양) 이다. 마우스를 올리면 두 변수의 값과 기여량이 함께 표시된다.

9.5.4 모수항 — 선형항과 요인항

- 수치형 선형항: 적합선을 겹쳐 그린 데이터 산점도
- 요인항: 요인 수준별 평균 기여도를 보여 주는 막대그래프

9.6 Correlation(상관)

모든 수치형 예측변수 간의 피어슨 상관을 보여 주는 히트맵 행렬이다. 데이터를 가져온 직후 바로 사용할 수 있으며 모형 적합이 필요 없다.

- 색 그라데이션: 파랑 (-1) → 흰색 (0) → 빨강 (+1)
- 각 칸에 상관계수 값이 표시됨
- 적합 전에 공곡선성 (다중공선성의 평활 대응 개념) 을 확인하는 데 유용함

9.7 Diagnostics(진단)

두 가지 진단 시각화를 제공한다:

진단 패널 (`gratia::appraise` 사용): 잔차 대 적합값, Q-Q 그림, 잔차 히스토그램, 반응변수 대 적합값의 네 가지 그림.

실측값 대 예측값 산점도: 45 도 기준 파선과 함께 관측값 대 예측값을 보여 준다. 로그 변환 모형에서는 값이 원래 척도로 역변환된다.

9.8 RCA Adjustments(RCA 조정)

RCA 계산 후에 표시되는 세 개의 히스토그램이다 (감정평가/시장 모드 전용):

- **Residual Adjustment %** — 매매가격 대비 잔차 조정의 백분율 분포
- **Net Adjustment %** — 순조정치의 분포
- **Gross Adjustment %** — 총조정치의 분포

각 히스토그램은 부제에 평균, 중앙값, 표준편차를 표시하고 평균과 중앙값 위치에 파선 기준선을 그린다.

9.9 Anova(분산분석)

GAM 요약의 모수항과 평활항을 합친 통합 ANOVA 표이다. 항, 유형 (모수항 또는 평활항), 그리고 관련된 모든 통계량을 보여 준다.

9.10 MgcV Output(mgcV 출력)

다음은 보여 주는 원시 텍스트 출력이다:

- 적합 소요 시간
- GAM 식
- 모형 출력 (print)
- 전체 `summary(model)` 출력
- family 와 연결함수
- 전체 공곡선성 측도

9.11 Sign Check(부호 점검)

earthUI 매듭점을 가져온 경우, 이 탭은 earth 모형에서의 각 변수 방향과 GAM 에서의 방향을 비교한다:

- **earth_direction**: 증가, 감소 또는 혼합 (한지 함수의 부호에 근거)
- **gam_direction**: 증가, 감소 또는 혼합 (100 개 격자점에서 평활항의 도함수로 추정)
- **consistent**: TRUE/FALSE
- **warning**: 불일치가 있을 경우 그 내용

상태 표시줄에는 “All directions consistent” 또는 불일치가 있는 변수의 개수가 표시된다.

9.12 Concurvity(공곡선성)

공곡선성은 다중공선성의 평활 대응 개념이다. 이 탭은 다음을 보여 준다:

- 전체 공곡선성 표: 행은 worst/observed/estimate 측도, 열은 각 평활항이다. 색으로 구분된다: 흰색 (<0.5), 노랑 (0.5–0.8), 빨강 (>0.8).
- 쌍별 공곡선성 표: 동일한 색 구분을 적용한 최악의 경우 쌍별 행렬.

“worst” 행의 값이 약 0.8 을 넘으면 해당 평활항이 다른 평활항들로 잘 근사될 수 있다는 뜻이며, 중복 가능성을 시사한다.

9.13 Basis Check(기저 점검)

`mgcv::gam.check()` 의 출력으로, 각 평활항에 대한 기저 차원 적정성 검정이다. 어떤 평활항의 k 가 너무 작으면 이 검정이 낮은 p-값으로 이를 표시하며, k 를 늘릴 것을 시사한다.

10 데이터 내려받기

적합 후 **Download Output (Excel)** 버튼 (사이드바 7 절) 은 예측값과 진단 결과를 담은 Excel 파일을 활성 프로젝트의 mgcv 출력 폴더로 내보낸다.

10.1 출력 열

열	설명
basis	역변환된 절편값 (모든 행에서 동일)
<var>_contribution	평활항별 예측 기여도. 로그 모형에서는 금액 단위로 비례 배분된다.
est_<target>	모형 예측값 (예: est_sale_price). 해당하는 경우 로그 척도에서 역변환됨
residual	실측값 - 예측값
calc_residual	검증용: 실측값 - 기여도 합계 (역변환됨)
cqa	상태-품질-선호도 (Condition-Quality-Appeal, CQA) 점수 (0-10 척도)
residual_sf	잔차 / 거주면적 (living_area 가 지정된 경우)
cqa_sf	residual_sf 에 의한 순위로 계산한 CQA

감정평가 모드에서 1 행 (대상) 의 residual/cqa/cqa_sf 는 NA 로 설정된다. living_area 열이 지정된 경우 행은 residual_sf 내림차순으로 정렬된다.

성공적으로 완료되면 내려받기 버튼에 흰색 체크 표시가 나타난다.

10.2 CQA 점수

CQA 는 각 행의 잔차를 다른 모든 행과 견주어 0-10 척도로 순위를 매긴다:

- **높은 CQA(약 9-10):** 예측보다 훨씬 높은 가격에 매매됨 — 상태/품질/선호도가 우수할 가능성이 크다
- **낮은 CQA(약 0-1):** 예측보다 훨씬 낮은 가격에 매매됨 — 상태가 열악하거나 급매일 가능성이 크다
- **CQA 약 5:** 중앙값 부근

11 RCA 계산과 내려받기

RCA(잔차 제약 접근법, Residual Constraint Approach) 작업 흐름은 **Appraisal** 과 **Market Area Analysis** 모드에서 모형을 적합한 뒤 사용할 수 있다. 대상 부동산은 반드시 1 행에 있어야 한다.

11.1 RCA 대화상자 열기

사이드바 8 절의 **Calculate RCA Adjustments & Download** 버튼을 클릭한다. 다음 항목이 있는 모달 대화상자가 나타난다:

- **Score type:** CQA 또는 CQA per SF(living_area 가 지정된 경우). 단위면적당 옵션은 대상의 잔차를 거주면적으로 조정한다.
- **Subject CQA Score:** 0.00 에서 10.00 사이의 값을 받는다. 처음에는 빈칸이며, earthUI 모형을 가져온 경우 earthUI 에서 사용한 점수 유형과 값이 공유 프로젝트 설정을 통해 그대로 이어진다.

Generate 를 클릭하면 계산과 내려받기가 이루어진다. Excel 파일은 활성 프로젝트의 mgcv 출력 폴더 (<os>_out_mgcv) 에 기록된다.

11.2 CQA 점수 보간

1. 비교사례들의 CQA 점수와 잔차를 정렬한다
2. 선형 보간 (approx()) 으로 입력한 CQA 값을 잔차에 대응시킨다
3. 대상값 = 모형 예측값 + 보간된 잔차
4. 가중치가 0 인 행은 대상과 동일한 잔차를 가진 것처럼 취급된다

11.3 출력 열

열	설명
subject_value	모형 예측값 + 보간된 잔차
subject_cqa	사용자가 입력한 CQA 점수
<var>_adjustment	대상 기여도 - 비교사례 기여도 (금액 단위)
residual_adjustment	대상 잔차 - 비교사례 잔차
net_adjustments	모든 조정치의 합
gross_adjustments	조정치 절댓값의 합
residual_adj_pct	매매가격 대비 잔차 조정%
net_adj_pct	매매가격 대비 순조정%
gross_adj_pct	매매가격 대비 총조정%
adjusted_sale_price	실제 매매가격 + 순조정치

계산이 성공적으로 끝나면 버튼에 흰색 체크 표시가 나타난다.

12 거래사례비교 그리드

거래사례비교 그리드 (Sales Comparison Grid) 는 **Appraisal** 모드에서만, RCA 조정치를 계산한 뒤에 사용할 수 있다. 이 기능은 감정평가 보고서에 넣기 적합한 서식이 적용된 Excel 통합문서를 생성한다.

12.1 비교사례 선택 대화상자

Generate Sales Grid & Download 버튼을 클릭하면 모든 비교 매매사례를 나열하는 모달 대화상자가 열린다. 비교사례는 두 무리로 나뉜다:

- 추천 비교사례 (미리 선택됨): 총조정액이 매매가격의 $< 25\%$ 이며 매매 경과일 순 (최근 순) 으로 정렬된다. 최대 30 건.
- 기타 비교사례 (기본적으로 선택 해제): 나머지 사례로, 총조정률 순으로 정렬된다.

비교사례를 선택한 뒤 **Generate Sales Grid** 를 클릭하면 통합문서가 만들어진다.

필수 특수 유형: `contract_date` 와 `living_area` 가 반드시 지정되어야 한다. 권장 특수 유형: `latitude`, `longitude`, `lot_size` — 이들이 없어도 그리드는 작동하지만 일부 항목이 비게 된다.

12.2 통합문서 구성

생성된 통합문서에는 여러 시트가 있으며, 각 시트에는 비교사례 3 건이 나란히 배치된다:

- 머리글 행: 대상과 비교사례의 주소, 매매가격, 매매일, 시장 노출 일수
- 묶음 행: 위치 (위도, 경도, 지역), 대지 규모 (대지 면적, 대지 치수), 연수 (실제 연수, 유효 연수) 를 묶음 머리글과 하위 항목으로 표시
- 모형 변수 행: 평활항마다 한 행씩, 대상 기여도, 비교사례 기여도, 조정액을 보여 준다
- 잔차 특성 행: 모형이 포착하지 못한 특성 (예: 상태, 품질, 조망) 에 대해 감정평가사가 입력할 수 있는 빈 잠금 해제 칸
- 요약 행: 순조정%, 총조정%, 조정된 매매가격을 실제 Excel 수식으로 표시

12.3 시트 보호와 잔차 칸

각 시트는 실수로 편집되는 것을 막기 위해 보호되지만, 잔차 특성의 값 칸만은 명시적으로 잠금이 해제되어 있다. 이로써 감정평가사는 수식 기반의 조정 계산을 유지한 채 모형에 없는 특성의 값을 입력할 수 있다.

13 보고서 내려받기

보고서 생성은 이전 배포판의 단일 단계 보고서를 대체하는 2 단계 Quarto 작업 흐름이다.

13.1 1 단계: Quarto 보고서 생성

Generate Quarto Report 버튼 (모형을 적합한 뒤 사용 가능) 은 자립형 .qmd 묶음을 활성 프로젝트의 mgcv 출력 폴더에 기록한다. 이 묶음에는 Quarto 원본, 미리 렌더링된 그림 파일 (PNG 와 PDF), 모든 표와 모형 메타데이터를 담은 `report_data.rds` 페이로드, 그리고 Word 출력을 위한 `reference.docx` 서식 템플릿이 들어 있다. 묶음이 자립형이므로 렌더링 전에 .qmd 원본을 편집하거나 다른 Quarto 문서와 결합할 수 있다.

그림은 묶음을 생성할 때 미리 렌더링된다. 개별 그림 생성이 실패하면 자리표시 이미지가 기록되고 묶음의 나머지는 정상적으로 생성된다.

13.2 2 단계: Quarto 보고서 변환

Convert Quarto Report 절은 `quarto::quarto_render()` 를 통해 .qmd 파일을 하나 이상의 출력 형식으로 렌더링한다. 기본값은 가장 최근에 생성된 묶음이지만, **Browse** 버튼으로 임의의 .qmd 파일을 선택할 수도 있다. 세 가지 형식을 사용할 수 있다:

- **HTML** — 자립형 HTML 문서. 추가 소프트웨어 없이 모든 플랫폼에서 작동한다. 권장 기본값이다.
- **Word (.docx)** — 함께 제공되는 `reference.docx` 서식 템플릿으로 렌더링된다. 편집과 배포에 적합하다.
- **PDF** — **LaTeX** 설치가 필요하다 (아래 시스템 요건 참조). LaTeX 이 감지되지 않으면 형식 체크박스에서 PDF 옵션이 자동으로 숨겨진다.

이전 방식의 단일 단계 렌더러인 `render_gam_report()` 와 `export_gam_docx()` 는 프로그래밍 방식으로 계속 사용할 수 있다.

13.3 보고서 내용

생성되는 보고서에는 다음이 포함된다:

- **데이터셋과 모형:** 데이터 파일, 표본 크기, family, method, 적합 정보
- **결과 요약:** R^2 , CV R^2 , 설명된 이탈도, AIC 및 관련 적합 지표
- **모형 식:** 평활항은 f_i 표기로, 선형항은 계수를 명시한 모형 식과 평활 함수 정의. 긴 식은 PDF 에서 여러 줄로 줄바꿈된다.
- **요약 표:** 평활항 (EDF, F, p-값) 과 모수항 (추정값, 표준오차, t, p-값)
- **활성화된 상호작용:** 모형에 포함된 텐서 및 요인별 평활 상호작용 항의 행렬 (상호작용이 있을 때만 표시)
- **변수 중요도:** F-통계량 (평활항) 또는 $|t|$ (모수항) 순위의 막대그래프
- **평활 효과** (모든 항 유형):
 - 일변량 평활 그림 (대화형 그림의 정적 ggplot2 버전)
 - 요인별 평활 다중선 그림
 - 상호작용 히트맵 (t_i 및 t_e 항)
 - 모수항 그림 (수치형은 산점도, 요인형은 막대그래프)
- **예측변수 상관 히트맵**
- **진단:** 잔차 진단과 실측값 대 예측값 산점도
- **mgcv 요약:** 적합된 모형의 전체 텍스트 출력
- **표기법 부록:** 보고서에서 사용한 가치 기여 함수 표기법을 설명하는 번호 없는 마무리 절

생성이 성공하면 Generate Quarto Report 버튼에 흰색 체크 표시가 나타난다.

14 모형 함수 내보내기

mgcvUI 는 적합된 GAM 의 평활 함수를 여러 언어의 독립 코드로 내보낼 수 있다. 각 평활항은 200 개 격자점에서 평가되어 선형 보간을 사용하는 조회 표 (lookup table) 로 내보내지므로, mgcv 에 의존하지 않고 R 밖에서도 모형을 사용할 수 있다.

현재 배포판에서 함수 내보내는 것은 프로그래밍 방식으로 이루어진다. 내보내기 보조 함수들은 (fit_gam() 이 생성한) 적합 결과 객체를 받아 아래에 설명한 파일들을 기록한다:

- generate_r_code(), generate_python_code(), generate_cpp_code() — 언어별 소스 파일
- export_functions_sqlite() — SQLite 조회 표 데이터베이스
- export_functions_zip() — 네 가지 형식을 임의로 조합해 zip 압축 파일로 묶는다

14.1 R 함수

파일: gam_functions.R

- 반응변수, R^2 , CV R^2 , 식을 담은 머리말 주석.
- INTERCEPT 상수.
- 평활 변수마다 미리 계산된 x/y 벡터를 갖는 approxfun() 하나씩.
- 선형항 계수를 이름 있는 상수로 제공.
- 사용법: 절편, 각 g_var(value), 각 coef_var * value 를 합산한다.

14.2 Python 함수

파일: gam_functions.py

- numpy 가져오기.
- INTERCEPT 상수.
- 변수별: _x_var 및 _fx_var 배열과 np.interp() 를 사용하는 g_var(x) 함수.
- 선형항 계수를 COEF_* 상수로 제공.

14.3 C++ 헤더

파일: gam_functions.hpp

- #pragma once 를 사용하는 헤더 전용 파일.
- gam_functions 네임스페이스.
- std::lower_bound 를 사용하는 interp() 보조 함수.
- 변수별: 정적 상수 배열을 갖는 인라인 함수.
- 선형항 계수를 constexpr double 로 제공.

14.4 SQLite 데이터베이스

파일: gam_functions.sqlite

- smooth_grids 테이블: (variable, x, fx) 열 — 200 점 조회 표.
- model_info 테이블: response, r_squared, cv_rsqr, dev_explained, aic, n_obs, family, method, formula.
- linear_terms 테이블: variable, coefficient.
- intercept 테이블: 절편값.

15 earthUI 및 glmnetUI 와의 비교

earth 응용 프로그램과 그 후처리 모듈들은 회귀 모델링을 위한 하나의 도구 모음이다. 이들은 동일한 데이터 형식, 특수 열 유형, RCA 작업 흐름, 시연용 데이터셋을 공유하지만 모델링 엔진은 서로 다르다.

15.1 주요 차이

특성	mgcvUI	earthUI	glmnetUI
방법	GAM (mgcv)	MARS/Earth (earth)	엘라스틱 넷 (glmnet)
관계	평활 비선형	구간별 선형	선형
변수 선택	<code>select=TRUE</code> 수축	단계적 가지치기	라쏘 별점
상호작용	텐서곱 (te/ti)	힌지 곱 (2-3 차)	쌍별 곱
부분효과	신뢰구간을 갖는 평활 곡선	g-함수 (구간별)	선형 기울기
earth 파이프라인	earth 에서 매듭점을 가져옴	매듭점의 출처	힌지 기저를 가져옴
코드 내보내기	R, Python, C++, SQLite	R, Python, C++, SQLite	—
보고서 형식	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

earthUI 에만 있는 차별점 하나를 언급할 만하다. earthUI 에는 자유 서술 형태의 MLS 비교란에서 구조화된 부동산 속성을 추출하는 실험적 Prolog/DCG 기능이 포함되어 있다.

15.2 공통 기능

세 도구 모두 다음을 제공한다:

- 공유 `regProj` 프로젝트 트리 (기본값 `~/regProj`) — 프로젝트, 입력 파일, 방법별 출력 형제 앱들 사이에서 공유된다
- `snake_case` 열 변환을 동반한 CSV 및 Excel 가져오기
- CSV 구분자, 소수점 기호, 용지 크기를 포함한 30 개국 이상의 로케일 지원
- 특수 열 지정 (`contract_date`, `living_area`, `area`, `latitude`, `longitude` 등)
- 기준일로부터의 `sale_age` 자동 계산
- 다크/라이트 모드 전환 (Nord 테마)
- 설정 유지 (SQLite 및/또는 `localStorage`)
- 상관 히트맵, 변수 중요도, 진단
- CQA 점수 산정과 RCA 조정 작업 흐름
- 비교사례 선택과 서식이 적용된 Excel 출력을 갖춘 거래사례비교 그리드
- 동일한 시연용 데이터셋 (`Appraisal_1.csv`) 호환

15.3 권장 작업 흐름

- 중요한 변수, 비선형 관계, 상호작용을 발견하려면 **earthUI** 로 시작하라. earth 의 자동 모형 구축이 가장 좋은 출발점을 제공한다.
- 부분효과를 더 매끄럽게 만들고 싶다면 **mgcvUI** 로 정교화하라. earthUI 의 매듭점을 가져오면 매끄럽게 이행할 수 있다.
- 공격적인 변수 선택이나 계수 부호 제약이 필요한 순수 선형 모형이 필요하다면 **glmnetUI** 를 사용하라.
- 비선형 모형의 추가적인 복잡도가 적당한지 판단하려면 동일한 데이터에 세 도구를 모두 적용해 비교하라.

16 시연용 데이터셋: Appraisal_1.csv

16.1 설명

시연용 데이터셋은 earthUI, glmnetUI 와 공유된다. 이 파일은 mgcvUI 소스의 demo_mls/ 폴더에 있으며, earthUI 가 설치되어 있다면 다음과 같이 찾을 수 있다:

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

이 파일에는 모의 MLS 내보내기에서 만든 약 1,500 건의 주거용 매매 기록이 들어 있으며, 1 행이 대상 부동산이고 나머지 행이 비교 매매사례이다. 데이터는 부동산의 규모, 연수, 위치가 다양하게 분포한 다중 지역 시장의 단독주택 매매를 나타낸다.

이는 실제 데이터가 아니며, 북부 캘리포니아의 실제와 유사한 근린을 바탕으로 만든 것이다. 모든 식별 정보는 변경되거나 삭제되었다.

16.2 열

열	유형	특수 유형	설명
sale_type	character	sale_type	대상과 비교 매매사례를 구분하는 레코드 유형 코드
weight	numeric	weight	관측치 가중치 (0 = 적합에서 제외)
id	numeric	display_only	내부 레코드 ID
property_id	numeric	display_only	MLS 부동산 식별자
listing_id	character	display_only	MLS 매물 번호
parcel_number	character	display_only	카운티 과세평가관의 필지 번호 (APN)
street_address	character	display_only	부동산 주소
city_name	character	display_only	도시
postal_code	character	display_only	우편번호
county_name	character	display_only	카운티
contract_date	Date	contract_date	매매 계약일 (sale_age 를 계산함)
sale_age	numeric	—	계약일부터 기준일까지의 일수 (미리 계산됨)
coe_date	Date	display_only	에스크로 종결일
listing_status	character	display_only	매물 상태 (예: "Sold")
sale_price	numeric	(목표변수)	매매가격 — 반응변수
rent	numeric	—	월 임대료 (다중 목표변수 모형용)
list_price	numeric	display_only	호가
original_list_price	numeric	display_only	최초 호가
living_sqft	numeric	living_area	총 거주면적 (제곱피트)
ppsf	numeric	display_only	제곱피트당 가격
beds_total	integer	—	침실 수
baths_total	numeric	—	총 욕실 수 (예: 2.5 = 전체 2 + 반 1)
lot_size	numeric	lot_size	대지 면적 (제곱피트)
area_id	integer	area	MLS 지역 식별자
area_text	character	display_only	지역명
age	numeric	actual_age	부동산의 연수 (년)
year_built	integer	display_only	건축 연도
latitude	numeric	latitude	위도 (모형에서는 소수점 3 자리로 반올림)
longitude	numeric	longitude	경도 (모형에서는 소수점 3 자리로 반올림)
latitude6	numeric	display_only	전체 정밀도 위도
longitude6	numeric	display_only	전체 정밀도 경도
garage_spaces	integer	—	차고 주차 대수
fp_count	integer	—	벽난로 개수
no_of_stories	numeric	—	층수
style	character	—	건축 양식
view	character	—	조망 유형 (예: "Neighborhood", "Hills")
days_on_market	integer	dom	시장 노출 일수
listing_date	Date	listing_date	매물 등록일

열	유형	특수 유형	설명
sale_concessions	numeric	concessions	매도인 양여 조건

16.3 빠른 시작 예시

1. mgcv 루틴을 실행한다: `earthUI::launch_mgcv()`
2. 목적을 **For Appraisal** 로 하여 새 프로젝트를 만든다 (1 절)
3. `Appraisal_1.csv` 를 프로젝트의 입력 폴더에 넣고 2 절에서 선택한다 (필요하면 **Refresh** 를 클릭한다)
4. `sale_price` 를 목표변수로 선택한다
5. 비율 조정을 원한다면 Response Transform 을 **Log (natural)** 로 설정한다 (선택 사항)
6. 위의 표에 나온 대로 특수 유형을 지정한다
7. 예측변수를 포함시킨다: `sale_age, living_sqft, baths_total, lot_size, area_id`(요인으로), `age, latitude, longitude, garage_spaces`
8. 기본값을 유지한다: `REML, gamma = 1.2, tp 기저, k = 자동, select = TRUE`
9. **Fit Mgcvm GAM Model** 을 클릭한다
10. Summary, Contribution(평활 그림), Diagnostics 탭을 검토한다
11. 출력을 내려받고 (7 단계) CQA 순위를 살펴본다
12. CQA 점수 ~5.00 으로 RCA 조정치를 계산한다 (8 단계)
13. 추천 비교사례로 거래사례비교 그리드를 생성한다 (9 단계)

17 시스템 조건과 문제 해결

17.1 지원 플랫폼

mgcvUI 는 **macOS, Windows, Linux** 에서 실행된다. 이 응용 프로그램은 RStudio 를 사용하는 macOS 에서 개발되고 주로 시험되었다. 플랫폼별 유의 사항은 아래에 정리했다.

17.2 R 버전

R $\geq 4.1.0$ 이 필요하다. RStudio Desktop(2023.06 이상) 을 적극 권장한다. RStudio 에는 보고서 렌더링에 필요한 **Quarto** 와 **pandoc** 이 함께 들어 있으며 앱을 실행하기에 편리한 환경을 제공한다.

17.3 필수 R 패키지

다음 패키지들은 의존성으로 자동 설치된다:

패키지	용도
shiny	웹 응용 프로그램 프레임워크
mgcv	GAM 모형 적합
valengrCore	공유 감정평가 코어 (특수 역할, earthUI 설정 이어받기)
ggplot2, plotly	정적 및 대화형 기여도 그림
gratia	평활항 추출과 진단
DT	대화형 데이터 표
quarto, rmarkdown	Quarto 보고서 생성과 렌더링
readxl	Excel 파일 가져오기
shinyFiles	파일 및 폴더 탐색기
RSQLite, DBI	설정 및 프로젝트 데이터베이스 (SQLite)
jsonlite	변수 상태 직렬화

17.4 선택적 의존성

구성 요소	필요한 경우
LaTeX (TinyTeX, MiKTeX 또는 MacTeX)	PDF 보고서 생성 시에만 필요하다. LaTeX 설치가 감지되지 않으면 형식 드롭다운에서 PDF 옵션이 자동으로 숨겨진다. 설치 방법: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	ggplot2 그림에 쓰이는 Roboto Condensed 글꼴용. 사용할 수 없거나 오프라인이면 앱이 자동으로 시스템 산세리프 글꼴로 대체한다.
earth ($\geq 5.3.0$)	earthUI 가져오기 파이프라인에만 필요하다. 단독 사용에는 필요하지 않다.
callr	백그라운드 모형 적합 (앱의 반응성을 유지한다). 없으면 적합이 동기적으로 실행된다.
writexl, openxlsx	Excel 내려받기 (모형 출력, RCA, 매매사례 그리드). 없으면 앱이 안내한다.
bslib	Bootstrap 5 테마 적용 (Nord Light/Dark).
officer	이전 방식의 프로그래밍 Word 내보내기 (<code>export_gam_docx()</code>).

17.5 플랫폼별 유의 사항

17.5.1 macOS

문제가 가장 적다. 시스템 라이브러리에는 Homebrew 를 권장한다. PDF 보고서가 필요하다면:

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

RStudio Desktop 과 잘 작동한다. 주요 유의 사항은 다음과 같다:

- 기업용/보안이 강화된 컴퓨터: 임시 디렉터리나 사용자 데이터 디렉터리가 제한되어 있으면 설정 유지와 보고서 생성이 실패할 수 있다. 앱은 R 콘솔에 경고를 표시하지만 설정 유지 없이 계속 작동한다.
- **LaTeX**: R 에서 TinyTeX 를 설치하거나 (위 macOS 와 동일한 명령) miktex.org 에서 MiKTeX 를 설치한다.
- 긴 파일 경로: Windows 에는 260 자 경로 제한이 있다. 작업 디렉터리 경로를 짧게 유지하라.

17.5.3 Linux

배포판에 따라 편차가 가장 크다. Ubuntu/Debian 사용자는 R 패키지를 컴파일 하기 전에 시스템 라이브러리가 필요할 수 있다:

```
sudo apt install libcurl4-openssl-dev libssl-dev \
libxml2-dev libsqlite3-dev libfontconfig1-dev
```

PDF 보고서를 위해서는 `tinytex::install_tinytex()` 를 실행하거나 시스템 패키지 관리자로 `texlive-full` 을 설치한다.

디스플레이가 없는 헤드리스 서버에서는 plotly 대화형 그림이 보고서에서 정적 대체 이미지로 렌더링된다. Shiny 앱 자체는 웹 브라우저 연결이 필요하다.

17.6 점진적 기능 축소

mgcvUI 는 선택적 구성 요소가 없어도 작동하도록 설계되었다:

누락된 구성 요소	동작
LaTeX 없음	보고서 형식 드롭다운에서 PDF 옵션이 숨겨진다. HTML 과 Word 는 계속 사용할 수 있다.
인터넷 없음 / 글꼴 실패	Roboto Condensed 글꼴이 시스템 산세리프로 대체된다. 콘솔에 메시지가 기록된다.
RSQLite 없음 / 읽기 전용 파일 시스템	세션 사이에 설정이 유지되지 않는다. 콘솔에 경고가 표시된다. 그 밖에는 앱이 정상 작동한다.
gratia::appraise 실패	Diagnostics 탭과 보고서에 자리표시 메시지가 표시된다. 대안으로 콘솔에서 <code>mgcv::gam.check()</code> 를 사용하라.
임시 디렉터리에 쓸 수 없음	보고서 생성이 명확한 오류 메시지와 함께 실패한다. <code>TMPDIR</code> 환경변수를 쓰기 가능한 위치로 지정하여 해결한다.

17.7 문제 해결

17.7.1 “PDF option is not available”

LaTeX 설치가 감지되지 않은 경우이다. R 에서 다음을 실행한다:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

그런 다음 앱을 다시 시작하면 PDF 옵션이 나타난다.

17.7.2 “Settings will not persist”

SQLite 데이터베이스 디렉터리를 만들거나 여기에 쓸 수 없는 경우이다. 다음 경로의 권한을 확인하라:

- macOS/Linux: ~/Library/Application Support/R/mgcvUI/ 또는 ~/.local/share/R/mgcvUI/
- Windows: %APPDATA%/R/data/mgcvUI/

17.7.3 “Diagnostics plot unavailable”

gratia::appraise() 함수는 복잡한 모형 (텐서 상호작용이 많거나 비표준 family 인 경우) 에서 이따금 실패한다. 이는 상위 패키지의 알려진 문제이다. 우회 방법: R 콘솔에서 mgcv::gam.check(model) 을 실행해 진단 그림을 직접 확인하라.

17.7.4 “Report error: number of items to replace...”

이는 gratia 패키지와 관련하여 이전 버전에서 알려졌던 문제이다. 최신 버전의 mgcvUI 가 설치되어 있는지 확인하라. 현재 버전은 이 오류를 포착하여 진단 절에 자리표시를 넣은 보고서를 생성한다.

17.7.5 포트 7880 이 이미 사용 중

이전 세션 때문에 포트 7880 이 점유되어 앱이 시작되지 않는다면 다음을 실행한다:

```
# macOS/Linux:
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 Linux 에서 패키지 설치 실패

시스템 라이브러리가 없는 경우이다. 위 Linux 절에 나열된 개발용 헤더를 설치한 뒤 install.packages() 를 다시 시도하라.

참고문헌

- 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).

- Wood, S. N. (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).

기초 개념

개관

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor, 편집장 • 2026년 8월 3일

판본 1.0.0

DOI: [10.67330/82vssv81](https://doi.org/10.67330/82vssv81)

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

《기초 개념》(Foundations)은 현대 감정평가 방법론의 바탕을 이루는 개념들에 대한 정밀하고 인용 가능한 서술을 발행한다. 각 항목은 실무 감정평가사가 마주치는 용어를 — 때로는 다른 이름으로, 때로는 그것인 줄 알아채지 못한 채 마주치는 용어를 — 골라 그 형식적 맥락 속에 자리매김한다. 그 개념이 어디에서 왔는지, 기술적으로 무엇을 뜻하는지, 실무의 어느 지점에 나타나는지, 그리고 왜 방어 가능성(defensibility)에 중요한지를 밝힌다.

장간호는 헤도닉(hedonic) 체계의 여섯 가지 핵심 개념을 확립하며, 각 항목이 앞선 항목의 어휘 위에 쌓이도록 배열되어 있다. 이 순서는 《이질재》(Heterogeneous Good, 애초에 감정평가 방법론을 필요하게 만드는 부동산의 속성)에서 출발하여, 《특성 공간》(Characteristics Space)과 《헤도닉 가격함수》(Hedonic Price Function, 부동산이 대표하는 속성 묶음을 조직하는 형식적 장치)를 거쳐, 《잠재가격》(Implicit Price)과 《헤도닉 회귀분석》(Hedonic Regression, 시장 증거로부터 각 속성의 한계기여분을 복원하는 방법)에 이르고, 마지막으로 《잠재변수》(Latent Variable, 모든 헤도닉 모형에서 복원되지 않은 잔차 신호가 실제로 무엇을 나타내는가)에 도달한다.

캘리포니아주 퍼시픽(Pacifica) 근린의 여덟 개 비교사례라는 하나의 관통 예제가 여섯 항목 전체를 꿰뚫으며 본 호의 교육적 실마리를 이룬다.

왜 이 여섯 개념인가, 왜 이 순서인가

여섯 항목의 선택은 임의적이지 않다. 각 개념은 현대 헤도닉 방법론의 하층을 지탱하는 구조재이고, 각각은 일상적인 감정평가 실무에서 실제로 마주치게 되며, 각각은 오해되거나 이름 붙여지지 않은 채 쓰이는 빈도가 충분히 높아 인용 가능한 형태로 정리해 두는 것이 실무적 가치를 갖는다.

배열 순서는 개념 사이의 의존관계가 결정한다. 이질재가 뿌리이다. 그것은 부동산(그리고 중고차, 노동시장, 그 밖의 몇몇 재화)이 지닌 속성, 곧 헤도닉 체계 전체를 필요하게 만드는 바로 그 속성에 이름을 붙인다. 특성 공간은 이질재를 벡터로 표현하기 위한 형식적 장치이며, 그것 없이는 그 이상 어떤 것도 정밀하게 말할 수 없다. 헤도닉 가격함수는 특성 공간의 점을 가격으로 사상하는 균형 대상이다. 잠재가격은 그 함수의 편도함수이다. 거래사례비교법 격자의 모든 보정은, 감정평가사가 그렇게 부르든 부르지 않든, 이 한계기여분에 대한 주장으로 읽을 수 있다. 헤도닉 회귀분석은 관측된 거래로부터 잠재가격을 복원하는 경험적 추정량이며, 이론이 자료 분석으로 바뀌는 지점이 바로 이 항목이다. 잠재변수는 회귀분석의 잔차가 시작되는 곳에서 이어받는다. 관측되지 않은 속성들의 효과가 버려도 좋은 잡음이 아니라 구조를 가진 신호로서 잔차 속에 살아남는 것이다.

이 순서대로 읽어 나간 독자는 “감정평가 방법론은 애초에 왜 존재하는가?”에서 “적합된 모형의 잔차를 우리는 어떻게 보아야 하는가?”에 이르는 하나의 연속된 실마리를 얻게 될 것이다. 필자의 경험에 비추어 보면, 대부분의 감정평가사는 여러 해에 걸친 실무를 통해 이 여정을 단편적으로만 통과한다.

관통 예제

캘리포니아주 퍼시픽카의 주거용 비교사례 여덟 건으로 이루어진 하나의 데이터셋이 여섯 항목 전체를 관통한다:

비교사례	GLA(sf)	대지 (sf)	조망	상태	거래가격
A	1,650	9,000	유	양호	\$1,425,000
B	1,650	7,200	무	양호	\$1,280,000
C	1,850	7,500	무	양호	\$1,407,500
D	1,450	7,000	무	양호	\$1,155,000
E	1,700	8,000	유	우수	\$1,470,000
F	1,550	7,100	무	보통	\$1,177,500
G	1,700	10,500	무	양호	\$1,392,500
H	1,600	7,300	무	양호	\$1,312,500

이 예제는 손으로 따져 볼 수 있을 만큼 의도적으로 작게, 그러면서도 각 개념을 모두 연습할 수 있을 만큼 풍부하게 설계되었다. 항목 001 은 비교사례 A 와 B 사이의 \$145,000 차이를, 하나의 짝지는 매매 관측치 속에 뒤엉킨 두 특성 (조망과 대지) 의 결합 기여분으로 검토한다. 항목 002 는 여덟 개 비교사례 전부를 4 차원 특성 공간의 점으로 놓고, 소박한 거리 척도가 이들의 순위를 잘못 매긴다는 점을 보인다. 항목 003 은 이 근린의 헤도닉 가격함수에 대한 하나의 함수형태를 제안한다. 항목 004 는 잠재가격을 해석적으로 복원하고, 애초의 A 대 B 차이를 조망과 대지의 기여분 \$127,039 와 설명되지 않은 \$18,000 으로 분해한다. 항목 005 는 R 에서 모형을 정식으로 적합시키고 표준오차와 잔차를 포함한 완전한 $1m()$ 출력을 보고한다. 항목 006 은 모형이 설명하지 못하는 단 하나의 잔차 — 비교사례 H 의 +\$46,455 편차 — 를 검토하고, 이를 필자가 다른 곳에서 발전시킨 잔차제약접근법 (Residual Constraint Approach) 방법론의 바탕이 되는 잠재변수적 해석들과 연결한다.

이 데이터셋은 계산을 재현하고자 하는 독자를 위해 본 호의 저장소 `shared/data/pacifica_comps.csv` 경로에서 이용할 수 있다.

본 호를 읽는 법

헤도닉 방법론을 처음 접하는 독자는 여섯 항목을 순서대로 읽어야 한다. 항목 사이의 의존관계는 실재한다. 각 항목은 앞선 항목에서 확립된 어휘를 전제하며, 건너뛰어 읽으면 항목 자체가 미리 대비할 수 없는 혼란을 자초하게 된다.

헤도닉 회귀분석을 이미 접해 본 독자는 — 고급 통계학 강의를 수강했거나, 자동가치산정모형 (AVM) 을 다루어 보았거나, 대량평가 문헌을 읽어 본 대다수 감정평가사가 여기에 해당한다 — 본 호를 복습과 참고 자료로 쓸 수 있다. 각 항목은 개별적으로 인용 가능하도록 집필되었다. 본지의 다른 곳에 발표된 논문이 잠재가격을 참조할 때, 개념을 다시 유도하는 대신 DOI 로 항목 004 에 연결하면 된다.

잔차제약접근법 방법론에 특히 관심 있는 독자는 먼저 항목 006 을 읽어 잠재변수적 해석들을 이해한 뒤, 항목 003-005 로 돌아와 RCA 가 확장하는 헤도닉 장치를 살피고, 이어서 항목 001-002 에서 그 모든 것이 딛고 선 기초 어휘를 확인하는 것이 좋다.

무엇이 감정평가 보정을 방어 가능하게 만드는가라는 물음에 관심 있는 독자라면, 각 항목의 “왜 방어가능성에 중요한가” 절만 순서대로 읽어도 그것이 보정에 기반한 평가의 방법론적 위상에 관한 편의 짧은 논고에 해당한다는 사실을 알게 될 것이다. 그 논고는 본 호 어디에도 명시적으로 조립되어 있지 않지만, 모든 항목의 잠재된 주제이다.

편집 방침

《기초 개념》은 교과서 지면이 아니다. 각 항목은 한자리에서 읽을 수 있을 만큼 짧고, 이미 감정평가 실무에 능숙한 독자를 상정해 집필되며, 포괄적인 해설이 아니라 개념의 정밀한 정식화를 제시하도록 설계된다. 그 목표는 본지의 다른 논문들이 인용할 수 있는 안정적인 참조점을 제공하는 데 있다.

각 항목은 편집부의 검토를 거치며, 해당 분야 전문가의 초청 논평이 붙을 수 있다. 항목은 크리에이티브 커먼즈 저작자표시 (Creative Commons Attribution) 라이선스로 발행되고, Crossref(정본 인용) 와 Zenodo(판본별 보존) 를 통해 각각 DOI 를 부여받는다. 각 항목은 시간이 지나면서 더 다듬어질 수 있다. 소규모 개정은 Zenodo 판본을 갱신하고, 대규모 개정은 변경 내용을 기술한 편집 주기와 함께 두 DOI 를 모두 갱신한다. Crossref DOI 는 항상 최신 판본으로 연결되며, 특정 과거 판본은 Zenodo 를 통해 지정할 수 있다.

앞으로의 호에서는 《기초 개념》의 어휘를 인접 영역으로 확장할 것이다. 공간적 의존, 조건부 기댓값, 내생성, 누락변수 편익, 식별, 그리고 대리변수와 잠재변수를 구별하는 형식적 장치 등이 그것이다. 다루어야 할 개념에 대한 제안은 언제든지 환영하며, 편집장에게 보내 주시기 바란다.

본 호의 여섯 항목:

1. 이질재 (Heterogeneous Good) — 감정평가 방법론이 애초에 왜 존재해야 하는가.
2. 특성 공간 (Characteristics Space) — 묶음 속성의 벡터 표현.
3. 헤도닉 가격함수 (Hedonic Price Function) — 특성에서 가격으로의 균형 사상.
4. 잠재가격 (Implicit Price) — 각 특성의 한계 시장가격.
5. 헤도닉 회귀분석 (Hedonic Regression) — 관측된 거래로부터 함수를 추정하기, 완전한 R 실무 예제 수록.
6. 잠재변수 (Latent Variable) — 잔차로부터 구조화된 신호를 복원하기.

참고문헌.

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)

부록: Prolog 형식화

아래의 기계 판독 가능한 부속 파일 (000-overview.pl) 은 본지의 Prolog 지식베이스에서 이 항목을 형식화한 것이다. 이 파일은 본문과 함께 열람 가능한 Prolog 갤리로도 발행된다.

```
%% 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).
```

이질재

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor • 2026년 8월 3일

판본 1.0.0

DOI: [10.67330/slscsv65](https://doi.org/10.67330/slscsv65)

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

형식적 정의. 이질재 (heterogeneous good) 란, 각각이 구매자가 개별적으로 가치를 매기는 서로 다른 특성 묶음이고 그 특성들이 따로 거래되지도 않기 때문에 서로 대체될 수 없는 재화를 말한다.

직관적 이해. 동질재 (homogeneous good) 란 어느 개체를 가져와도 무방한 재화이다. WTI 원유 한 배럴, 2 등급 황색 옥수수 한 부셸, 전력망에서 나온 1 킬로와트시, 어떤 주식의 한 주가 그렇다. 이런 재화 뒤에 있는 등급 기준이나 규격은 바로 개별 개체를 서로 대체 가능하게 만들기 위해 존재한다. 그 결과 가격은 특정 개체가 아니라 범주 전체에 적용된다. 이질재는 특정 개체가 문제가 되는 재화이다. 저 집이 아니라 이 집, 옆에 있는 중고차가 아니라 이 중고차인 것이다.

기술적 의미에서 어떤 재화를 이질적으로 만드는 특징은 두 가지이다. 첫째, 구매자가 시장의 개체들 사이에서 서로 다르게 나타나는 여러 구별 가능한 속성에 관심을 둔다. 둘째, 그 속성들은 해당 재화의 생산기술에 의해 하나로 묶여 있어 따로 구입할 수 없다. 총거주면적 200 제곱피트만 따로 살 수는 없다. 학군을 토지와 분리해서 취득할 수도 없다. 특성들은 묶인 채로 도착하며, 시장은 특성 자체가 아니라 묶음 전체에 가격을 매긴다.

부동산은 이질재의 전형이며 그 속성을 유난히 강하게 드러낸다. 위치적 특성은 완벽히 복제 불가능하다. 두 필지가 같은 위치를 차지할 수 없으므로, 모든 부동산은 시장의 다른 모든 부동산과 자신을 구별해 주는 속성 (자신의 좌표) 을 최소한 하나는 가진다. 물리적, 법적, 환경적 속성이 그 위에 변이의 차원을 더 쌓아 올려, 전체로 보면 각 부동산마다 고유한 벡터를 만들어 낸다.

감정평가사가 마주치는 지점. 거래사례비교법에서 이루어지는 모든 보정은 대상 부동산과 비교사례가 이질재라는 사실에 대한 암묵적 인정이다. 부동산이 동질적이라면 거래가격이 이미 직접 비교 가능할 터이므로 격자 (grid) 가 필요 없을 것이다. 격자는 서로 다른 묶음의 가격을, 묶음 구성의 차이를 보정함으로써 조정하려는 시도이다. 격자가 하지 않는 일도 짙어 둘 만하다. 격자는 거래가격을 개별 특성의 가치기여분으로 분해하는 법이 없다. 전통적 격자에는 가치기여분을 적는 열이 없으며, 오직 특성별로 직접 추정된 보정액만 있을 뿐이다.

원가법은 묶음을 그 구성요소 — 기초, 골조, 마감, 감가상각 — 로부터 재구성하고 각각에 가격을 매기는 방식으로 이질성을 다룬다. 수익환원법은 묶음을 하나의 현금흐름 프로파일로 추상화하고, 환원율과 수익률이 묶음 수준의 이질성을 시장에서 도출된 배수 속으로 흡수하게 함으로써 이질성을 다룬다. 세 가지 평가방식 모두 부동산이 이질재라는 근본적 사실에 대한 방법론적 대응이다. 이들이 갈리는 지점은 그 사실을 어떻게 헤쳐 나가는가에 있지, 그것과 대면하는지 여부에 있지 않다.

감정평가사는 비교사례 선정, 짝지은 매매 분석 (paired-sales analysis), “같은 것끼리의 비교” 를 두고 궁리할 때마다 비공식적으로도 이질성과 마주친다. 가능한 한 많은 속성에서 유사한 비교사례를 찾으려는 본능은 이질성에 대한 실무자의 대응이다. 대상과 비교사례 사이의 묶음 거리를 최소화하면, 묶음 가운데 관측되지 않은 부분이 거래가격 차이에 기여하는 몫이 줄어든다.

왜 방어가능성에 중요한가. 부동산의 이질성은 감정평가 방법론이 하나의 분과로서 존재해야 하는 이유 그 자체이다. 부동산이 서로 대체 가능하다면 시장가치는 게시된 가격에서 직접 관측될 것이고, 감정평가는 조회 (lookup) 로 환원될 것이다. 그렇지 않기 때문에 가치는 관련은 있으나 동일하지는 않은 부동산의 가격으로부터 추론되어야 하며, 추론의 모든 단계는 정당화되어야 할 방법론적 선택이 된다.

이질성을 형식적으로 인정하면 USPAP 과 기준규칙 (Standards Rule) 의 방어가능성이 걸려 있는 여러 실무 문제가 분명해진다. 비교사례 선정이란, 그로부터 복원한 잠재가격이 일반화될 만큼 대상과 충분히 가까운 특성 공간상의 개체들을 고르는 일이다. 보정의 타당성은 그 바탕에 있는 잠재가격이 시장 증거로 뒷받침될 수 있는가에 달려 있다. 평가방식 간 조정 (reconciliation) 은 이 부동산의 이질성을 어느 방식이 가장 신빙성 있게 헤쳐 나가는지에 대한 감정평가사의 판단을 반영한다. 이 방어가능성 문제들은 “이질재” 를 그 바탕에 놓인 조건으로 명명하는 순간 하나같이 더 날카로워진다.

감정평가 실무 예제. 캘리포니아주 퍼시피카 (Pacifica) 의 같은 근린에서 최근 이루어진 두 건의 매매:

속성	비교사례 A	비교사례 B
GLA(sf)	1,650	1,650
침실/욕실	3 / 2	3 / 2
대지 (sf)	9,000	7,200
조망	유	무
거래가격	\$1,425,000	\$1,280,000

동질재의 틀에서 보면 이 둘은 같은 부동산이다. GLA 가 같고, 방 수가 같으며, 같은 근린에 있다. 이질재의 틀에서 보면 이 둘은 서로 다른 묶음이며, 구매자가 개별적으로 가치를 매기는 두 특성 (대지와 조망) 에서 갈린다. \$145,000 의 가격 차이는 그 두 속성의 결합 기여분이며, 하나의 짝 안에 두 잠재가격이 뒤엉켜 있는 것이다.

잠재가격을 개별적으로 복원하려면 더 많은 짝이, 이상적으로는 한 번에 하나의 속성만 달라지는 짝이 필요하다. 조망 유/무와 대지 대/소의 조합을 아우르는 서너 건의 추가 매매가 있으면 잠재가격을 따로 식별하기 시작할 수 있다. 충분히 많은 비교사례에 걸쳐 충분한 변이가 있으면 헤도닉 회귀분석이 그 분리를 정식으로 수행한다. 변이가 제한적이면 감정평가사는 괄호 씌우기식 짝지는 매매 추론에 묶이게 되며, 그 신뢰도는 묶음들이 특성 공간에서 서로 얼마나 가까운가에 명시적으로 좌우된다.

이 예제는 이질성이 왜 근본 조건인지를 보여 준다. “동일한” 두 채의 집조차 동일하지 않으며, 감정평가의 방법론적 장치는 바로 그 둘 사이의 거리를 다루기 위해 존재한다.

참고문헌.

- 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.

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

부록: Prolog 형식화

아래의 기계 판독 가능한 부속 파일 (001-heterogeneous-good.pl) 은 본지의 Prolog 지식베이스에서 이 항목을 형식화한 것이다. 이 파일은 본문과 함께 열람 가능한 Prolog 갤리로도 발행된다.

```
%% 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).

```

특성 공간

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor • 2026년 8월 3일

판본 1.0.0

DOI: 10.67330/3yymh010

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

형식적 정의. 어떤 이질재 부류의 특성 공간 (characteristics space) 이란, 구매자가 가치를 매기는 속성들을 좌표로 삼는 벡터공간이며, 그 안에서 각 재화는 자신의 속성 수준으로 이루어진 벡터로 표현된다.

직관적 이해. 앞선 항목에서 이질재란 서로 구별되는 특성 묶음이라는 점을 확립했다. 자연히 다음 물음이 뒤따른다. 그런 묶음들을 어떻게 형식적으로 기술하고 비교할 것인가? 랭커스터 (Lancaster, 1966) 에서 비롯한 답은, 각 재화를 구매자가 관심을 두는 특성들을 축으로 삼는 벡터공간상의 한 점으로 취급하는 것이다.

이 표현에서 퍼시픽카의 주거용 부동산은 더 이상 “파크 퍼시픽카에 있는 집” 이 아니라 총거주면적, 대지면적, 조망, 상태, 학군, 해안까지의 거리 등에 대한 좌표를 가진 벡터가 된다. 이 속성값을 모두 공유하는 두 부동산은 특성 공간의 같은 점을 차지할 것이고, 어느 한 속성이라도 다른 부동산은 다른 점을 차지한다. 이 공간의 기하는 두 부동산이 얼마나 가깝고 얼마나 다른지를, 구매자가 구별을 행하는 모든 차원을 포착하는 방식으로 부호화한다.

공간의 차원 수는 연구 대상 시장에서 어떤 속성이 유의미한가에 달려 있다. 전형적인 주거용 감정평가라면 관련 공간은 열 몇 개의 차원을 가질 수 있다. 물리적 차원 (GLA, 대지, 침실 수, 욕실 수, 연식, 품질, 상태), 위치적 차원 (학군, 조망, 편의시설 또는 혐오시설과의 근접성), 그리고 법적·경제적 차원 (용도지역, 지역권) 이 그것이다. 상업용 감정평가라면 관련 차원은 현금흐름 속성 (임대차 구조, 임차인 신용도, 공실률) 쪽으로 이동하지만, 바탕에 놓인 기하학적 발상은 동일하다.

특성 공간이 유용한 까닭은, 질적인 실무자의 직관 (“이 비교사례들은 대상과 비슷하다”) 을 양적인 주장 (“이 비교사례들은 특성 공간에서 대상과 가깝다”) 으로 바꾸어 주기 때문이다. 일단 거리가 정의되면, 거래사례비교와 관련된 모든 결정은 공간상 위치에 관한 진술로서 검토될 수 있다.

감정평가사가 마주치는 지점. 감정평가사는 이름을 붙이지 않은 채 늘 특성 공간 안에서 일한다. 거래사례비교 격자의 모든 행은 하나의 좌표축이고, 모든 비교사례는 좌표값의 행벡터이며, 모든 보정은 나머지를 고정한 채 한 축을 따라 이동하려는 시도이다.

본 호 전체에서 사용할 여덟 개의 퍼시픽카 비교사례가 이 구성을 직접 보여 준다. 각 비교사례는 GLA, 대지면적, 조망, 상태가 펼쳐는 4 차원 특성 공간의 한 점이다:

비교사례	GLA(sf)	대지 (sf)	조망	상태	거래가격
A	1,650	9,000	유	양호	\$1,425,000
B	1,650	7,200	무	양호	\$1,280,000
C	1,850	7,500	무	양호	\$1,407,500
D	1,450	7,000	무	양호	\$1,155,000
E	1,700	8,000	유	우수	\$1,470,000
F	1,550	7,100	무	보통	\$1,177,500
G	1,700	10,500	무	양호	\$1,392,500
H	1,600	7,300	무	양호	\$1,312,500

앞선 항목의 비교사례 A 와 B 가 첫 두 행이며, 나머지 여섯 건은 본 호의 남은 부분에서 활용할 방식으로 예제를 확장한다.

각 비교사례는 $\mathbb{R}^4$ 의 한 점이다:

- 비교사례 A 는 (1650, 9000, 1, 2) 에 있다.
- 비교사례 B 는 (1650, 7200, 0, 2) 에 있다.
- 비교사례 G 는 (1700, 10500, 0, 2) 에 있다.

조망과 상태 좌표는 수치로 부호화되어 있다 (조망은 $\{0, 1\}$, 상태는 $\{1, 2, 3\}$). 가격은 특성 공간의 좌표가 아니라 그 공간 안의 위치에 대한 함수이다.

왜 방어가능성에 중요한가. 특성 공간의 틀은 비교사례 선정이 실제로 무엇인지를 날카롭게 만든다. 비교사례란 단순히 “비슷한 부동산” 이 아니라, 특성 공간의 관련 영역에서 대상과 가까운 하나의 점이다. 비교사례가 공간상 대상에 가까울수록 필요한 보정은 작아지고, 도출된 가치의 신뢰도는 높아진다.

이 재구성은 비공식적 격자가 가려 버리는 여러 실무 문제를 드러낸다:

- 어느 차원이 중요한가? 축의 선택은 모형화의 결정이다. “조망” 을 축으로 포함하는 것은 시장이 조망에 별도의 가격을 매길 때에만 의미가 있고, “풍수 방위” 를 포함하는 것도 어떤 시장 부문이 그것에 가격을 매길 때에만 의미가 있다. 방어 가능한 감정평가는 축 위의 값뿐만 아니라 축 자체를 방어할 것을 요구한다.
- 어떤 거리 척도를 적용할 것인가? 두 비교사례의 가격이 \$200,000 차이가 나는 것은 그들이 특성 공간에서 멀리 떨어져 있기 때문일 수 있으나, 거리 계산에는 각 축에 대한 척도 조정이 필요하다. 변환하지 않은 유클리드 거리는 GLA 1 제곱피트와 대지 1 제곱피트를 동등하게 취급하는데, 시장은 그렇게 하지 않는다.
- 비교사례들이 가격함수가 안정적으로 작동하는 영역에 있는가? 대상이 특성 공간에서 표본이 성긴 구석에 자리한다면, 가까이 놓인 비교사례가 없으므로 가치 도출은 필연적으로 외삽이 된다.

감정평가 실무 예제. 여덟 개의 퍼시픽카 비교사례는 4 차원 특성 공간에서 비교적 조밀한 영역을 차지한다. (1,700, 7,500, 1, 2) 에 있는 가상의 대상 부동산 — 1,700 sf, 대지 7,500 sf, 조망 있음, 양호한 상태 — 에 어느 비교사례가 “가장 가까운가” 라는 물음을 살펴보자.

축의 척도를 조정하지 않은 유클리드 거리를 소박하게 사용하면, 대지 차원의 값 (7,000–10,500) 이 조망 (0–1) 이나 상태 (1–3) 보다 수치상 훨씬 크기 때문에 대지 차원이 계산을 지배한다. 대상까지의 거리는 다음과 같다:

비교사례	거리 (척도 미조정 유클리드)	비고
A	1,501	조망 일치, 대지가 지배
B	308	조망 불일치, 대지 근접
C	230	조망 불일치, 대지 근접
D	559	조망 불일치, GLA 차이
E	502	조망 일치, 상태 상이
F	422	조망과 상태 모두 상이
G	3,000	조망 불일치, 대지 멀리 떨어짐
H	224	조망 불일치, 전체 중 가장 가까움

원시 거리 기준으로는 비교사례 H 가 최근접 이웃이다. 그러나 비교사례 H 에는 조망이 없고 대상에는 있다. 소박한 척도가 우리를 오도한 것이다. 대지 차원의 수치적 규모가 범주형 조망 차원을 밀어냈는데, 정작 가격에 더 큰 영향을 미치는 것은 조망이다.

방어 가능한 거리 척도라면 각 축을 그 잠재가격으로 — 대략, 그 축의 한 단위 변이가 몇 달러의 가치를 갖는지에 따라 — 조정할 것이다. 축을 잠재가격으로 조정하고 나면, 조망 (약 ~\$100,000) 은 더 이상 대지 (제곱피트당 ~\$25) 에 눌리지 않는다. 그러면 대상과 조망이 일치하는 비교사례 A 가 적절한 주요 비교사례로 떠오른다.

이것이 바로 뒤따르는 항목들에서 전개되는 잠재가격 장치가 제공하는 바이다. “가까움” 을 기하학적 어림짐작에서 시장에서 도출된 계산으로 바꾸어 주는, 특성 공간의 원칙 있는 척도 조정이 그것이다.

참고문헌.

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.

상호 참조: 이질재 (*heterogeneous good*); 헤도닉 가격함수 (*hedonic price function*); 잠재가격 (*implicit price*).

부록: Prolog 형식화

아래의 기계 판독 가능한 부속 파일 (002-characteristics-space.pl) 은 본지의 Prolog 지식베이스에서 이 항목을 형식화한 것이다. 이 파일은 본문과 함께 열람 가능한 Prolog 갤리로도 발행된다.

```
%% 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).

```

헤도닉 가격함수

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor • 2026년 8월 3일

판본 1.0.0

DOI: [10.67330/9x1a0489](https://doi.org/10.67330/9x1a0489)

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

형식적 정의. 헤도닉 가격함수 (hedonic price function) 란 이질재의 특성 공간상의 점에서 거래가격으로 이어지는 균형 사상 (寫像, mapping) 이며, 경쟁시장에서 구매자의 입찰함수와 매도자의 제안함수의 포락선으로 형성된다.

직관적 이해. 이제 두 가지 재료가 갖추어졌다. 이질재 (특성의 묶음) 와 특성 공간 (각 묶음이 하나의 점이 되는 벡터공간) 이다. 헤도닉 가격함수는 이 둘을 결합한다. 그것은 특성 공간의 각 점 $\mathbf{z}$ 에 가격을 부여하는 함수 $p(\mathbf{z})$ 로, 그러한 속성 조합을 가진 묶음에 시장이 얼마를 지불하는지를 알려 준다.

로젠 (Rosen, 1974) 의 핵심 통찰은 $p(\mathbf{z})$ 가 어느 한 구매자나 매도자에 의해 선택되는 것이 아니라는 점이다. 그것은 시장 균형에서 창발한다. 각 구매자는 자신의 선호와 소득을 전제로 어떤 묶음에 지불할 용의가 있는 최대액을 나타내는 입찰함수 (bid function) 를 가진다. 각 매도자는 자신의 비용과 대안을 전제로 수용 가능한 최소액을 나타내는 제안함수 (offer function) 를 가진다. 헤도닉 가격함수는 이 입찰과 제안의 포락선이며, 구매자와 매도자가 실제로 거래하는 경계선이다.

부동산의 경우 $p(\mathbf{z})$ 는 다음 물음에 대한 답이다. 특성의 완전한 묶음을 지정한다면 시장은 그것에 어떤 가격을 만들어 낼 것인가? 그 답은 임의적이지 않다. 그것은 시장 내 모든 구매자의 선호 총합과 모든 매도자의 비용 및 대안의 총합을, 구매자와 묶음의 균형적 짝짓기를 통해 매개하여 반영한다.

헤도닉 가격함수는 헤도닉 분석의 중심 이론적 대상이다. 잠재가격은 그것의 편도함수이다. 헤도닉 회귀분석은 그것의 경험적 추정이다. 거래사례비교 보정의 방어가능성은 궁극적으로 이 함수에 관한 주장으로 환원된다.

감정평가사가 마주치는 지점. 감정평가사는 이번에도 대개 이름을 붙이지 않은 채 $p(\mathbf{z})$ 를 두고 끊임없이 궁리한다. 여러 실무 활동이 헤도닉 가격함수에 대한 암묵적 조작이다:

거래사례비교법에서의 보정. 대상에는 조망이 있고 비교사례 B 에는 없다는 이유로 감정평가사가 비교사례 B 를 \$100,000 상향 보정할 때, 그는 $p(\mathbf{z})$ 에 관한 주장을 하고 있는 것이다. 즉, 다른 특성을 모두 고정한 채 조망 특성이 0 에서 1 로 이동하면 함수값이 \$100,000 만큼 증가한다는 주장이다. 이는 조망 차원에서 p 의 국소 기울기에 관한 진술이다.

평가방식 간 조정. 원가법은 복제된 묶음의 구성요소 비용을 합산하여 아래로부터 가치를 구성한다. 수익환원법은 현금흐름에서 가치를 도출하는데, 균형에서 현금흐름은 유사한 현금흐름을 산출하는 묶음의 가격을 반영한다. 거래사례비교법은 유사한 묶음의 관측된 거래로부터 가격을 직접 읽어낸다. 잘 작동하는 시장이라면 세 방식 모두 $p(\mathbf{z})$ 의 같은 점으로 수렴해야 한다. 결과가 같릴 때 감정평가사는 대상이 특성 공간에서 차지하는 위치에 대해 어느 방식이 함수를 가장 신뢰성 있게 측정했는지를 판단해야 한다.

시장분석과 시점 보정. 헤도닉 가격함수는 공급, 수요, 금융, 기대가 변함에 따라 시간에 걸쳐 이동한다. 연 6 퍼센트의 상승률이라는 주장은 $p(\mathbf{z})$ 가 전체적으로 어떻게 이동했는지에 관한 주장이다. 조망 프리미엄이 있는 부동산이 조망 없는 부동산보다 더 빠르게 상승했다는 발견은, 그 함수가 특성 공간 전반에 걸쳐 고르지 않게 이동했다는 주장이다.

왜 방어가능성에 중요한가. 헤도닉 가격함수를 명시적으로 명명하면 여러 근본적 방어가능성 문제가 분명해진다:

그 함수는 존재하는가? 로젠의 틀은 균형 함수가 유의미할 만큼 거래가 충분한 경쟁시장을 전제한다. 거래가 희박한 시장, 급변하는 시장, 또는 특이한 구매자 집단을 가진 하위시장에서는 그 함수가 제대로 정의되지 않거나 불안정할 수 있다. 그러한 시장에서 일하는 감정평가사는 자신의 추론 장치가 흔들리는 이론적 토대 위에서 작동하고 있음을 인정해야 한다.

그 함수는 어디서나 동일한가? 로젠의 틀은 시장 부문마다 하나의 함수를 낳는 것이지, 전 지구적으로 하나의 함수를 낳는 것이 아니다. 퍼시픽과 샌프란시스코는 서로 다른 헤도닉 함수를 가진 별개의 시장이며, 단독주택과 콘도미니엄은 같은 도시 안에서도 별개의 하위시장이다. 시장 부문을 뒤섞는 감정평가사는 하나의 함수가 양쪽을 아우른다고 암묵적으로 가정하는 것이며, 그 가정은 대개 틀렸다.

p 는 어떤 함수형태를 취하는가? 그 함수는 선형일 필요도, 로그선형일 필요도, 그 밖의 특정 형태일 필요도 없다. 로젠은 p 가 입찰과 제안의 포락선이라는 점만을 보였을 뿐, 그 모양을 지정하지는 않았다. 경험적으로 p 는 비선형성 (침실 하나를 더할 때의 한계가치는 GLA 에 따라 일정하지 않은 경우가 많다), 상호작용 (조망 프리미엄은 대지의 향에 따라 달라진다), 문턱효과 (주방이 일정 품질 기준에 못 미치면 상태 점수에 비해 불균형하게 큰 할인을 받을 수 있다) 를 보인다. 방어 가능한 추정을 위해서는 시장에서 뒷받침되는 함수형태를 쓰거나, 자료로부터 그 모양을 복원할 수 있는 유연한 추정량을 쓸 필요가 있다.

감정평가 실무 예제. 연구 대상인 퍼시픽 근린에 대해서는, 다음 항목에서 더 충분히 다룰 단순한 가법적 함수형태를 사용하기로 한다:

$$p(\mathbf{z}) = \beta_0 + \beta_{\text{GLA}} z_{\text{GLA}} + \beta_{\text{lot}} z_{\text{lot}} + \beta_{\text{view}} z_{\text{view}} + \beta_{\text{cond}} z_{\text{cond}}.$$

이 형태에는 두 가지 강한 가정이 담겨 있다. 각 특성의 기여분이 그 수준에 대해 선형이라는 가정과, 특성들이 서로 상호작용하지 않는다는 가정이다. 두 가정 모두 검정 가능하며 자주 위배되지만, 여기서는 출발점으로 채택한다.

본 호 관통 예제의 여덟 개 비교사례에 대해, 다음 항목들의 분석은 다음과 같은 함수를 낳는 β 값을 추정하게 된다:

모수	추정값
β_0 (절편)	\$80,777
β_{GLA} (sf 당)	\$575
β_{lot} (sf 당)	\$22.70
β_{view} (유/무)	\$86,179
β_{cond} (한 단계당)	\$49,824

가법적 형태에 대입하면 이 함수는 4 차원 특성 공간의 임의의 점에 대해 가격을 예측한다. 비교 사례 A 의 좌표를 대입하면 관측된 \$1,425,000 에 가까운 예측가격이 나오고, (1700, 7500, 1, 2) 의 가상 대상을 대입하면 그 부동산에 대한 가치 도출값이 나온다.

이 항목의 요점은 이 구체적인 수치가 아니라 개념적 이행에 있다. 그 함수는 우리가 추정하기 이전에 원리적으로 존재한다. 함수는 시장의 속성이다. 우리의 추정은 그것을 복원하려는 시도이다. 그 복원은 부분적이거나, 편의가 있거나, 잡음이 섞여 있을 수 있으며, 이어지는 세 항목은 그 복원을 어떻게 책임 있게 수행할 것인가에 관한 것이다.

참고문헌.

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.

상호 참조: 이질재 (*heterogeneous good*); 특성 공간 (*characteristics space*); 잠재가격 (*implicit price*); 헤도닉 회귀분석 (*hedonic regression*).

부록: Prolog 형식화

아래의 기계 판독 가능한 부속 파일 (003-hedonic-price-function.pl) 은 본지의 Prolog 지식베이스에서 이 항목을 형식화한 것이다. 이 파일은 본문과 함께 열람 가능한 Prolog 갤리로도 발행된다.

```
%% 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:
%  $p(z) = \beta_0 + \beta_{gla} * z_{gla} + \beta_{lot} * z_{lot} + \beta_{view} * z_{view} + \beta_{cond} * z_{cond}$ 
%
% 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 == yes -> N = 1
    ; 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).

```

잠재가격

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor • 2026년 8월 3일

판본 1.0.0

DOI: 10.67330/md6wke17

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

형식적 정의. 어떤 특성의 잠재가격 (implicit price) 이란 헤도닉 가격함수를 그 특성에 대해 편미분한 값을 특성 공간의 한 점에서 평가한 것이다. 그것은 다른 특성을 모두 고정한 채 해당 특성이 한 단위 증가할 때의 한계 시장가격을 나타낸다.

직관적 이해. 헤도닉 가격함수 $p(\mathbf{z})$ 는 각 묶음에 가격을 부여한다. 그러나 감정평가의 실무적 물음은 대개 묶음 전체의 가격이 아니라 묶음의 특정 구성요소의 가격에 관한 것이다. 조망은 얼마의 가치가 있는가? 100 제곱피트를 더하면 얼마가 늘어나는가? “양호” 와 “우수” 상태의 차이는 얼마를 기여하는가?

잠재가격이 이 물음들에 답한다. 이 용어는 로젠 (Rosen, 1974) 에서 비롯한다. 그는 차별화된 상품의 관측 가격과 그 특성이 함께 “일련의 잠재적 혹은 ‘헤도닉’ 가격을 정의한다” 는 점을 보였고, 랭커스터 (Lancaster, 1966) 는 재화를 특성의 묶음으로 보는 바탕의 관점을 제공했다. 각 특성 z_i 에 대해 잠재가격은 다른 특성을 모두 고정한 채 z_i 가 변할 때 $p(\mathbf{z})$ 가 변하는 비율이다:

$$\pi_i(\mathbf{z}) = \frac{\partial p(\mathbf{z})}{\partial z_i}.$$

잠재가격의 단위는 특성 한 단위당 달러이다. GLA 제곱피트당 달러, 대지 제곱피트당 달러, 상태 한 단계당 달러가 그것이다. 조망 같은 이항 특성이라면 잠재가격은 조망 없음 (0) 에서 조망 있음 (1) 으로 이동하는 데 지불되는 프리미엄 전액이다.

“잠재적” 이라는 말은 핵심적 사실을 포착한다. 이 가격들은 게시되지 않는다. “GLA 1 제곱피트를 \$575 에 팔겠다고 제안하는 사람은 없다. 그 가격은 실제로 게시되는 균형 묶음 가격 속에 잠재해 있으며, 헤도닉 장치는 바로 그것을 추출하기 위해 존재한다.

잠재가격은 헤도닉 체계의 감정평가적 산출물이다. 거래사례비교 격자의 보정은 원리적으로 잠재가격에 특성 차이를 곱한 추정치이다. 모든 보정은 특성 공간의 해당 지점에서 π_i 의 값이 얼마인가에 대한 내기이다.

감정평가사가 마주치는 지점. 거래사례비교 보정 격자는, 감정평가사가 그렇게 부르지 않더라도, 잠재가격 적용의 표이다. “대상보다 GLA 가 200sf 작으므로 +\$10,000” 이라고 적힌 격자 항목은 해당 특성 공간 위치에서 GLA 의 잠재가격이 제곱피트당 \$50 라고 주장하는 것이다. “열등한 조망에 대해 \$100,000” 이라고 적힌 항목은 이항 조망 특성의 잠재가격이 \$100,000 이라고 주장하는 것이다. 거래사례비교법 전체의 방어가능성은 이 잠재가격들이 시장에서 뒷받침될 수 있는가에 달려 있다. 다만 실무에서 감정평가사는 짝지은 매매, 경험, 동료 실무 관행으로부터 보정액을 직접 추정할 뿐, 그 바탕의 가치기여분을 계산하는 일은 없다. 전통적 격자는 보정액 자체만을 기록한다. 잠재가격이라는 독법은 그 보정이 무엇을 주장하는지에 대한 해석이지, 그것이 어떻게 도출되었는지에 대한 서술이 아니다.

이 명명은 또한 짝지은 매매 분석이 왜 보정액 도출의 기초 기법인지를 분명히 해 준다. 정확히 하나의 특성에서만 다른 두 비교사례는 그 특성의 잠재가격에 대한 직접적 추정치를 준다. 가격 차이를 특성 차이로 나누면 되는 것이다. 짝지은 매매 분석의 신뢰도는 두 비교사례가 특성 공간에서 정말로 충분히 가까워, 그로부터 복원한 잠재가격이 대상에 일반화될 수 있는가에 달려 있다.

실무에 중요한 잠재가격의 세 가지 성질:

국소적이다. 1,650sf 지점에서의 GLA 잠재가격이 2,500sf 지점에서와 반드시 같지는 않다. 해도 닉 가격함수는 일반적으로 비선형이므로 그 편도함수는 특성 공간의 위치에 따라 달라진다. 1,500sf 짝지는 매매 분석에서 도출한 보정액이 3,000sf 대상에 반드시 적용되지는 않는다.

조건부이다. 해변 하위시장에서 조망의 잠재가격은 내륙 하위시장에서와 같지 않다. 편도함수는 다른 특성을 고정한 채 취하지만, 그 값은 다른 특성이 어디에 고정되어 있는가에 달려 있다.

시장 의존적이다. 잠재가격은 균형가격이므로 거래 시점 구매자 집단의 선호를 반영한다. 특성 자체가 변하지 않아도 시장이 변하면 잠재가격이 달라질 수 있다. 시점 보정은 어제의 잠재가격이 오늘의 그것이 아닐 수 있다는 사실을 암묵적으로 인정하는 것이다.

왜 방어가능성에 중요한가. 잠재가격의 틀은 비공식적 격자가 가릴 수 있는 보정 방어가능성 문제들을 드러낸다:

특성 공간의 어디에 있는가? 잠재가격은 대상의 특성 공간상 위치 가까이에서 추정되었을 때에만 방어 가능하다. 내륙 비교사례에서 도출한 조망 프리미엄을 해변 대상에 적용하는 것은, 함수의 거동이 알려지지 않은 특성 공간 영역을 가로질러 외삽하는 일이다.

국소 기하는 어떠한가? 두 특성이 상호작용한다면 — 예컨대 대지가 클수록 조망의 가치가 크다면 — 이들을 가법적으로 따로 다루는 것은 설정 오류이다. 조망의 잠재가격은 대지면적에 의존하며, 상호작용을 무시한 보정은 함수의 실제 구조를 과소평가한다.

잠재가격은 어떻게 복원되었는가? 짝지는 매매 분석, 회귀계수, 업계 경험칙, 그리고 “전문가적 판단”은 모두 같은 양 — 잠재가격 — 을 가리키지만 인식론적 위상은 매우 다르다. 감정평가사의 방어가능성은 각 보정에 사용된 잠재가격이 어느 출처에서 나왔는지를 밝히는 데 달려 있다.

감정평가 실무 예제. 앞선 항목에서 채택한 가법적 설정을 쓰는 퍼시픽가 관통 예제에서, 잠재가격은 곧 선형함수의 계수이다:

$$\pi_{\text{GLA}} = \beta_{\text{GLA}}, \quad \pi_{\text{lot}} = \beta_{\text{lot}}, \quad \pi_{\text{view}} = \beta_{\text{view}}, \quad \pi_{\text{cond}} = \beta_{\text{cond}}.$$

항목 003 에서 추정한 계수를 쓰면:

특성	잠재가격	단위
GLA	\$575	sf 당
대지	\$22.70	sf 당
조망	\$86,179	유/무당
상태	\$49,824	한 단계당

이 잠재가격들은 하나의 짝지는 매매 분석을 함의한다. 항목 001 의 비교사례 A 대 비교사례 B 의 차이로 돌아가 보면:

- 비교사례 A 에는 조망이 있다 ($\Delta z_{\text{view}} = +1$). 기여분은 +\$86,179.
- 비교사례 A 의 대지가 1,800sf 더 크다 ($\Delta z_{\text{lot}} = +1,800$). 기여분은 $1,800 \times \$22.70 = +\$40,860$.
- 다른 특성은 모두 동일하다.
- 예측된 가격 차이: $\$86,179 + \$40,860 = \$127,039$.

관측된 가격 차이는 \$145,000 이었다. 잠재가격 분해는 그 가운데 약 \$127,000 을 설명하며, 남은 \$18,000 은 소규모 데이터셋의 표본변동을 (그리고 더 일반적으로는 잠재변수 항목에서 다룰 종류의 미관측 요인을) 반영한다.

실무상 짚어 둘 점이 두 가지 있다. 첫째, 선형성 가정이 실제로 일을 하고 있다. $p(\mathbf{z})$ 가 비선형이라면 같은 물리적 대지 차이라도 작은 대지에서 더 큰 가치를 가질 수 있고, 위의 가법적 분해는 편의를 갖게 된다. 둘째, 대지 제곱피트당 \$22.70 이라는 잠재가격은 추가 1 제곱피트의 한계가치이지 대

지 전체의 가치가 아니다. 9,000sf 대지가 묶음 가격에 기여하는 총액은 $9,000 \times \$22.70 = \$204,300$ 이지만, 1 제곱피트를 더 늘리면 묶음 가격은 \$22.70 만 오른다.

한계와 총액의 이 구별은 일상적인 감정평가 언어에서 상당한 혼동의 원천이며, 잠재가격을 명시적으로 명명함으로써 얻는 가장 중요한 개념적 소득 가운데 하나이다.

참고문헌.

- 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.
- Triplet, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

상호 참조: 이질재 (*heterogeneous good*); 특성 공간 (*characteristics space*); 헤도닉 가격함수 (*hedonic price function*); 헤도닉 회귀분석 (*hedonic regression*).

부록: Prolog 형식화

아래의 기계 판독 가능한 부속 파일 (004-implicit-price.pl) 은 본지의 Prolog 지식베이스에서 이 항목을 형식화한 것이다. 이 파일은 본문과 함께 열람 가능한 Prolog 갤리로도 발행된다.

```
%% 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).
```

헤도닉 회귀분석

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor • 2026년 8월 3일

판본 1.0.0

DOI: [10.67330/ams1t139](https://doi.org/10.67330/ams1t139)

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

형식적 정의. 헤도닉 회귀분석 (hedonic regression)이란 이질재 시장에서 관측된 거래로부터 헤도닉 가격함수를 추정하는 것으로, 모형에 들어간 특성들의 잠재가격에 대한 경험적 추정치를 산출한다.

직관적 이해. 앞선 항목들은 헤도닉 가격함수와 그 잠재가격을 이론적 대상으로 서술했다. 그 함수는 원리적으로 존재하지만, 실제로 우리가 관측하는 것은 유한한 거래 집합뿐이다. 거래된 묶음과 거래된 가격이 그것이다. 헤도닉 회귀분석은 이 둘을 잇는 경험적 다리이다.

설정은 간명하다. 각 거래는 $(\mathbf{z}_i, p_i)$ 쌍을 제공한다. 거래된 묶음의 특성 벡터와 그 거래가격이다. 이 쌍들의 모임은 (관측되지 않은) 헤도닉 가격함수로부터 얻은 하나의 표본이다. 회귀분석 (regression)은 그 표본에 함수형태를 적합시켜 함수를 특징짓는 모수를 추정한다. 모수가 추정되고 나면 적합한 함수의 편도함수가 잠재가격 추정치를 준다.

회귀분석을 돌리기 전에 세 가지 모형화 선택을 해야 한다:

- 어떤 특성을 모형에 넣을 것인가? 이 선택이 추정되는 함수의 차원을 결정한다. 관련 있는 특성을 누락하면 그 효과가 상관된 포함 특성들의 계수로 스며든다 (누락변수 편이).
- 어떤 함수형태인가? 선형, 로그선형, 이중로그, 다항식, 기저확장, 준모수적, 기계학습 등이 있다. 이 선택은 추정된 함수가 취할 수 있는 모양과 복원 가능한 잠재가격 양상을 제약한다.
- 어떤 추정량인가? OLS, 일반화최소제곱, 로버스트 M-추정량, 분위수회귀, 공간계량경제 추정량, 벌점회귀 등이 있다. 이 선택은 추정량이 무엇을 최적화하는지, 그리고 어떤 가정이 그 추론을 뒷받침하는지를 결정한다.

각 선택은 방법론적으로 실질적이며, 각각 방어가능성에 결과를 남긴다.

감정평가사가 마주치는 지점. 헤도닉 회귀분석은 대량평가 (mass appraisal)의 기초이며 대부분의 자동가치산정모형 (AVM)의 엔진이다. 카운티 과세평가관이 해마다 10만 필지를 재평가할 때, 그 일을 해내는 것은 최근 거래에 대한 헤도닉 회귀분석이다. Zillow 나 Redfin이 자동 추정치를 내놓을 때, 그것을 산출하는 것은 헤도닉 계열 모형이다 (OLS가 아니라 트리 앙상블인 경우가 많지만 개념적으로는 여전히 헤도닉 추정량이다). 대출기관의 유통시장 AVM이 어떤 부동산의 채용자 가부를 판정할 때에도 같은 장치가 관여한다.

개별 부동산 감정평가도 점점 더 헤도닉 회귀분석에 기대고 있다. 최종 가치 도출이 거래사례비교에서 나오는 경우에도 그렇다. 근린 표본에 대한 회귀분석에서 추정한 잠재가격은 감정평가사의 직관이나 경험적 백분율보다 훨씬 방어 가능하게 보정 크기를 뒷받침할 수 있다. 회귀계수는 격자에 적용되는 보정에 대해 의견이 아니라 증거가 된다.

왜 방어가능성에 중요한가. 회귀분석에 기반한 추정은 짝지은 매매 분석과는 다른 방어가능성 프로파일을 가진다. 장점은 이렇다. 회귀분석은 한 번에 두 건이 아니라 이용 가능한 모든 거래를 사용한다. 다른 여러 특성을 동시에 통제하면서 한 특성의 효과를 분리할 수 있다. 그리고 각 잠재가격의 불확실성을 정량화하는 표준오차를 산출한다. 대가는 이렇다. 회귀분석은 참된 가격함수와 맞지 않을 수도 있는 함수형태를 부과한다. 짝지은 매매 분석보다 더 많은 자료를 요구한다. 그리고 그 결과는 표본과 모형설정의 선택에 민감할 수 있다.

회귀분석 기반 감정평가에서는 몇 가지 방어가능성 문제가 되풀이된다:

표본은 적절한가? 회귀분석은 표본이 속한 시장 부문과 기간에서 헤도닉 함수가 작동하는 방식을 추정한다. 내륙 거래로 돌린 회귀분석을 해변 대상에 적용하는 것은, 통계 결과가 아무리 깔끔해 보여도 방어가능성의 실패이다.

모형설정은 적절한가? 선형 설정은 강한 가정이다. 참된 함수가 어떤 문턱 위에서 GLA 에 대해 수확체감을 보인다면, 선형 적합은 그 구간에서 잠재가격을 잘못 추정할 것이다. 진단, 잔차도, 대안적 설정은 방어 가능한 작업흐름의 일부이다.

표준오차는 무엇을 뜻하는가? 표준오차는 가정된 모형설정 아래에서의 표본변동을 포착한다. 그것은 모형설정 오류, 누락변수 편의, 표본선택 문제를 포착하지 않는다. 잘못된 설정에서의 작은 표준오차는 올바른 설정에서의 큰 표준오차보다 더 위험하다.

결과가 다른 방식과 삼각검증되는가? 방어 가능한 회귀분석 기반 잠재가격이라면 짝지은 매매 분석이 시사하는 바, 원가법의 구성요소 원가가 함의하는 바, 그리고 해당 시장에 밝은 감정평가사들이 믿는 바와 일관되어야 한다. 이 셋 모두와 어긋나는 회귀계수는 발표하라는 신호가 아니라 조사하라는 초대장이다.

감정평가 실무 예제. 이제 여덟 개의 퍼시픽카 비교사례에 대해 헤도닉 회귀분석 장치 전체를 돌려 볼 수 있다. 데이터셋은 `shared/data/pacifica_comps.csv` 에 있다.

R 코드:

```
comps <- read.csv("shared/data/pacifica_comps.csv")
fit <- lm(price ~ gla + lot + view + cond, data = comps)
summary(fit)
```

출력:

```
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
```

해석. 적합한 모형은 헤도닉 가격함수를 다음과 같이 추정한다:

$$\hat{p}(\mathbf{z}) = 80,777 + 575 z_{\text{GLA}} + 22.70 z_{\text{lot}} + 86,179 z_{\text{view}} + 49,824 z_{\text{cond}}.$$

각 계수는 해당 특성의 추정된 잠재가격이다. GLA 는 제곱피트당 \$575 를 기여하며 (표준오차가 가장 좁고 유의성이 가장 강하다), 조망은 퍼시픽카 부동산에 약 \$86,000 을 기여한다. 상태가 한 단계 오르면 (예컨대 “양호” 에서 “우수” 로) 대략 \$50,000 의 가치가 있다.

몇 가지 유의점이 곧바로 눈에 띈다:

표본이 매우 작다. 관측치 여덟 개에 예측변수 네 개이므로 잔차 자유도는 셋뿐이다. 표준오차는 그만큼 넓고, 대지·조망·상태의 p -값은 경제적으로 그럴듯함에도 통상적인 0.05 문턱에 이르지 못한다.

이는 소표본의 인공물이지, 이 특성들이 중요하지 않다는 증거가 아니다. 비교사례가 8 건이 아니라 80 건이라면 p -값은 극적으로 줄어들 것이다.

전체 적합도는 강하다. 조정 $R^2 = 0.93$ 과 모형 p -값 0.012 는 포함된 특성들이 표본 내 가격 변동의 대부분을 공동으로 설명함을 나타낸다. 강한 결합 적합도와 약한 개별 유의성의 조합은 다중공선성의 교과서적 신호이며, 여기서는 구조적이다. GLA, 대지, 상태가 표본 안에서 함께 움직이기 때문이다.

잔차 하나가 유독 두드러진다. 최대 잔차는 +\$46,455 이며 비교사례 H 에 해당한다. 나머지 일곱 개 잔차는 모두 $\pm \$18,000$ 대역 안에 든다. 비교사례 H 의 무언가가 적합된 함수에 포착되지 않은 것이다. 모형은 그 가격을 두 번째로 큰 잔차의 거의 세 배에 달하는 크기로 체계적으로 과소 예측하고 있다.

이 이상치 잔차는 잡음이 아니다. 그것은 비교사례 H 의 어떤 미관측 특성이, 포함된 변수들로는 복원할 수 없는 방식으로 가격에 기여하고 있다는 신호이다. 잠재변수 항목이 다루는 것이 바로 이 문제이다.

짜지은 매매 추론과의 검증. 항목 004 의 짜지은 매매 분석은 조망과 대지만으로 비교사례 A 와 B 사이에 \$127,039 의 차이를 예측했다. 회귀분석의 잠재가격은 그것과 동일하다 (그 계산의 출처가 바로 이 잠재가격이기 때문이다). 회귀분석은 짜지은 매매의 논리를 여덟 개 비교사례 전채로 동시에 일반화하지만, 이 표본 크기에서는 본질적으로 새로운 것을 낳지 않는다. 짜지은 분석을 확인해 주면서, 짜지은 매매 추론으로는 얻을 수 없는 표준오차를 보고할 뿐이다.

참고문헌.

- 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.

상호 참조: 이질재 (*heterogeneous good*); 특성 공간 (*characteristics space*); 헤도닉 가격함수 (*hedonic price function*); 잠재가격 (*implicit price*); 잠재변수 (*latent variable*).

부록: Prolog 형식화

아래의 기계 판독 가능한 부속 파일 (005-hedonic-regression.pl) 은 본지의 Prolog 지식베이스에서 이 항목을 형식화한 것이다. 이 파일은 본문과 함께 열람 가능한 Prolog 갤러로도 발행된다.

```
%% 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).
```

잠재변수

(한국어판)

기초 개념 (Foundations) • The Valuation Engineer

Bert Craytor • 2026년 8월 3일

판본 1.0.0

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

여기서 세 가지 실무적 귀결이 따라 나온다:

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

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

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

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

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

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

참고문헌.

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *American Journal of Psychology*, 15(2), 201–292.

Bollen, K. A. (2002). Latent Variables in Psychology and the Social Sciences. *Annual Review of Psychology*, 53, 605–634.

Craytor, B. (2025). Residual Constraint Approach: Framework & Protocol. Zenodo. [doi:10.5281/zenodo.14787917](https://doi.org/10.5281/zenodo.14787917)

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

부록: Prolog 형식화

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

```
%% 006-latent-variable.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 006
%% Title: Latent Variable
%% Author: Bert Craytor
%% Concept: latent_variable
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(latent_variable, "Latent Variable", measurement_theory,
%      "A latent variable is an unobserved attribute that influences
%      observed outcomes, whose value must be inferred from its
%      effects on observable variables rather than measured directly.",
%      published, 'dictionary-2026.1').
%
% term(residual, ...).
% term(omitted_variable_bias, ...).
% term(proxy_variable, ...).
% term(residual_constraint_approach, ...). % synonym: rca
%
% introduced_by(latent_variable, spearman_1904).
% formalized_by(latent_variable, bollen_2002).
% introduced_by(residual_constraint_approach, craytor_2025).
% depends_on(latent_variable, characteristics_space).
% -----

:- use_terms([ latent_variable, residual, omitted_variable_bias,
               proxy_variable, residual_constraint_approach,
               characteristics_space, hedonic_price_function,
```

```

        implicit_price, hedonic_regression ]).
:- use_relations([related/2, depends_on/2,
                 introduced_by/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(latent_variable).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.006').

% =====
% LOCAL VOCABULARY
% =====

local_term(unobserved).
local_term(noise).
local_term(signal).
local_term(structured_signal).
local_term(uad_3_6).
local_term(c3_condition_rating).
local_term(mls).
local_term(permit_records).
local_term(neighborhood_micro_effect).

% =====
% THE CENTRAL CLAIM
% =====

claim(latent_variables_accumulate_in_residuals,
      "Unobserved-but-relevant attributes do not vanish when omitted from the model;
      ↳ their effects accumulate in the residual, appearing as noise from the model's
      ↳ perspective but as structured signal from the data-generating process's
      ↳ perspective." ).

claim(residuals_are_mixture,
      "Standard regression theory treats residuals as iid Gaussian noise; the latent-
      ↳ variable view treats them as a mixture of sampling noise AND signal from
      ↳ unobserved attributes." ).

% =====
% THREE CONTEXTS WHERE APPRAISERS ENCOUNTER LATENT VARIABLES
% =====

appraiser_encounter(quality_and_condition_proxies,
      "UAD 3.6 'C3 condition' is a single ordinal scale collapsing a multi-dimensional
      ↳ construct (deferred maintenance, finish wear, mechanical-systems age, presentation
      ↳ ). The rating is a proxy_variable for a latent construct." ).

appraiser_encounter(unexplained_grid_disagreement,
      "When three adjusted comps disagree by more than expected, the disagreement is
      ↳ rarely all sampling noise; usually reflects an attribute not yet identified or
      ↳ adjusted for." ).

appraiser_encounter(unobserved_recent_improvement,
      "A recent remodel, roof replacement, or permitted ADU buildout may be observable
      ↳ in person but not in MLS/assessor data — observable to the appraiser but latent
      ↳ to the regression." ).

```

```
% =====
% DEFENSIBILITY ISSUES
% =====

defensibility_issue(omitted_variable_bias,
    "If a latent variable is correlated with an observed variable, its effect leaks
    ↪ into the observed variable's coefficient. Observed-variable coefficients are
    ↪ conditional on what is missing from the model.").

defensibility_issue(residual_interpretation,
    "Treat residuals as a mixture, not pure noise. A residual that correlates with
    ↪ location, sale date, appraiser identity, or any other observable suggests the
    ↪ latent variable can be partially recovered.").

defensibility_issue(model_selection_limits,
    "The latent variables that matter most are often hardest to add (condition fine-
    ↪ grain, quality, micro-locational effects). Sometimes the defensible response is to
    ↪ allocate the unexplained variance to the residual rather than force-fit
    ↪ additional variables.").

% =====
% WORKED EXAMPLE: COMP H'S ANOMALOUS RESIDUAL
% =====

:- consult('../..//kb/pacifica_comps').

% Comp H's regression residual from entry 005.
comp_h_residual_dollars(46455).

% The latent attribute revealed by investigation.
comp_h_latent_attribute(kitchen_remodel).
comp_h_latent_attribute_value_estimate(60000).

% Why the estimated residual is smaller than the latent attribute's
% true contribution: regression spreads small portions of the latent
% effect across correlated coefficients, biasing them upward.
residual_smaller_than_truth_explanation(
    "Regression spreads small portions of the latent effect across the other
    ↪ coefficients via correlations, biasing them slightly upward in correlated
    ↪ dimensions.").

% =====
% THREE OPERATIONAL CONSEQUENCES
% =====

operational_consequence(outlier_residual_is_evidence,
    "The first response to Comp H's residual is to investigate, not to drop. An in-
    ↪ person inspection or permit search reveals the remodel; the model's failure has
    ↪ diagnostic value.").

operational_consequence(other_coefficients_slightly_biased,
    "View, condition, and lot coefficients absorbed small fractions of the remodel
    ↪ effect via correlations with Comp H's observed characteristics. Bias is small at n
    ↪ =8 but not zero.").

operational_consequence(residuals_carry_recoverable_structure,
    "Residuals in residential appraisal data are systematically not iid Gaussian. They
    ↪ reflect structured contributions of unobserved attributes that can be partially
```

```

    ↪ recovered: this is the Residual Constraint Approach intuition.").

% =====
% CONNECTION TO RESIDUAL CONSTRAINT APPROACH (RCA)
% =====
%
% The single-comp illustration here is the simplest case of the RCA
% intuition. With richer modeling, residual structure is partitioned
% into recognizable latent factors.

framework_reference(residual_constraint_approach, craytor_2025).

rca_intuition(
    "Treat the residual itself as a latent-variable signal to be modeled rather than
    ↪ discarded. Latent variables are not a nuisance to be controlled away; they are the
    ↪ next frontier of defensible valuation modeling.").

rca_factor_examples(
    [condition_fine_grain, quality_fine_grain, functional_utility]).

% Two-level condition/quality structure (RCA protocol; mirrors the prose).
rating_levels(condition, condition_coarse_grain, condition_fine_grain).
rating_levels(quality, quality_coarse_grain, quality_fine_grain).

coarse_to_fine_grain_protocol(
    "Regress on the appraiser's coarse-grain C1-C6/Q1-Q6 ratings to obtain approximate
    ↪ value contributions; move those contributions out of the measured attributes back
    ↪ into the residual; reassign them to the fine-grain latent variables and refine
    ↪ against MLS public remarks. In the end, the fine-grain value contribution replaces
    ↪ the coarse-grain value contribution.").

% =====
% INFERENCE RULE: anomalous-residual detection
% =====
%
% A property's residual is "anomalous" if its magnitude exceeds the
% typical residual band reported in entry 005 (+/- $18,000 for the other
% seven comps). Useful for flagging properties whose hedonic prediction
% suggests a latent attribute warrants investigation.

typical_residual_band(18000).

anomalous_residual(P, R) :-
    comp_h_residual_dollars(R0), % stand-in: this entry references the
                                % residual computed in entry 005; future
                                % integration would pull from there.

    P == comp_h,
    R = R0.

% =====
% CROSS-REFERENCES
% =====

cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).
cross_reference(hedonic_regression).

```

최근 판례

최근 법원 판결 — 미국

최근 법원 판결 • *The Valuation Engineer*

William Bert Craytor • 제 1 권 제 1 호 • 2026 년 여름

DOI: 10.67330/fz34f711

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

관할별 판례 요약

California(캘리포니아주)

The Retail Property Trust v. Orange County Assessment Appeals Board No. 1. California Court of Appeal(캘리포니아주 항소법원), 2026 년 4 월 판결.

항소법원은 COVID-19 팬데믹 기간에 발령된 정부 명령이 부과한 출입 제한에 대하여 California Revenue and Taxation Code section 170(캘리포니아주 세입세무법 제 170 조) 에 따른 재산세 과세 평가 경감을 받을 수 있는지를 다루었다. 원고인 쇼핑센터는 팬데믹 기간의 출입 제한이 제 170 조가 말하는 재산에 대한 “손해”(damage) 에 해당하므로 과세평가 경감을 받을 자격이 있다고 주장하였다.

감정평가사에 대한 시사점. 제 170 조상 “손해” 에 관한 이 판결의 해석은, 공중보건 명령이든 기반 시설 사업이든 그 밖에 출입에 영향을 주는 정부 행위이든, 외부 원인으로 인한 부동산 효용의 감소를 근거로 한 향후의 평가 청구를 좌우하게 될 것이다. 출입 제한이 상업용 부동산 가치에 미치는 영향에 관하여 의견을 제시하도록 의뢰받은 감정평가사는, 효용의 제한이 — 물리적 훼손과는 구별되는 것으로서 — 과세평가 경감의 목적상 보상 대상인 손해에 해당하는지에 관한 법원의 논거를 검토하여야 한다.

Colorado(콜로라도주)

Cheroutes v. Rocket Mortgage & Solidifi. U.S. District Court, D. Colorado(미국 콜로라도 연방지방법원); 계속 중, 2026 년 진행 상황 갱신.

이 사건의 소장은 2021 년 Denver(덴버) 소재 2 세대 주택에 대한 감정평가와 관련하여 Fair Housing Act(공정주거법) 에 위반되는 감정평가 편향이 있었다고 주장한다. 해당 부동산은 약 \$640,000 로 평가되었는데, 이는 전년도 감정평가액보다 \$220,000 이상 낮은 금액이다. 앞서 한 연방판사는 피고들의 소각하 신청을 기각하면서, 대출기관과 감정평가관리회사 (AMC) 는 독립적인 수수료 감정평가사의 행위에 대하여 책임을 지지 않는다는 논리로 책임을 부인할 수 없고 Rocket Mortgage 와 감정평가관리회사 Solidifi 가 그 주장에 대하여 방어하여야 한다고 판시하였다. 이 사건은 2026 년 증거개시 단계에 있다.

감정평가사에 대한 시사점. 이 사건은 수수료 감정평가사의 행위에 대하여 감정평가관리회사와 대출 기관이 공정주거법상 책임을 지는지를 검증하는 현재 가장 주된 사건이다. 법원이 최종적으로 본안에서 감정평가관리회사와 대출기관의 책임을 인정한다면, 감정평가에서 비롯되는 공정주거 청구에 노출되는 당사자의 범위가 크게 넓어질 것이며, 감정평가관리회사와 감정평가사 간의 업무위탁 계약에도 파급될 가능성이 높다.

Florida(플로리다주)

주택담보 차주에게 부과된 AMC 수수료에 대한 플로리다주 집단소송. 2026 년 1 월 2 일 제기.

이 소장은 감정평가관리회사가 주택담보대출 차주에게 부과하는 수수료에 이의를 제기하면서, 수수료 감정평가사에게 지급되는 금액과 결제 시 차주에게 부과되는 금액 사이의 가산액이 연방법이 정한 한도를 초과하며 그 공시도 불충분하다고 주장한다. 이 사건은 Dodd-Frank(도드-프랭크법)의 “관행적이고 합리적인” 수수료 요건 (12 U.S.C. § 1639e) 과, Truth in Lending Act(대부진실법) 및 Real Estate Settlement Procedures Act(부동산결제절차법) 상의 소비자 공시 제도가 교차하는 지점에 놓여 있다.

감정평가사에 대한 시사점. 이 소송은 감정평가관리회사 보수 구조를 검증하는 것으로, 집단소송으로 인가되어 본안 판단이 내려질 경우 업계 전반의 공시 실무를 재편할 수 있다. 그 결과는 감정평가관리회사가 지급하는 금액과 차주에게 부과되는 금액 사이의 차액을 사법적 심사에 노출시킴으로써, 감정평가관리회사를 통해 업무를 수행하는 수수료 감정평가사의 보수 형성 환경에 간접적으로 영향을 줄 수 있다.

Georgia(조지아주)

Green Rock 보전지역권 사건들. United States Tax Court(미국 조세법원), *Tax Notes* 2026 년 3 월 2 일자에서 논평된 판결들.

조세법원은 조지아주 Piedmont(피드먼트) 지역의 인접 필지 (Green Rock Property 및 Green Rock South) 에 설정된 보전지역권을 평가하면서 최유효이용 (highest and best use) 심사의 각 요건을 차례로 검토하였다. 신청인 측 전문가들은 지역권 설정 이전 해당 부동산의 최유효이용이 화강암 채석장 운영이라고 주장하면서, 현금흐름할인분석을 사용하여 지역권 설정 전 공정시장가치를 약 \$13.75 million 로, 지역권 설정 후 가치를 \$180,000 로 추정하였다.

법원은 심사의 각 요건마다 유보적인 입장을 밝혔다. 법률적 허용성에 관하여 법원은 채석 작업을 허용하는 용도지역 변경이 합리적으로 개연성이 있다고 보지 않았으며, 특히 Green Rock Property 에 대한 용도지역 분류 변경이 새로운 보전지역권 감정평가를 의뢰한 직후에 이루어졌다는 점을 지적하였다 — 법원은 이러한 시간적 순서를 믿기 어렵다고 보았다. 물리적 가능성에 관하여 법원은, Piedmont 지역에 채굴 가능한 화강암이 일반적으로 존재한다는 사실이 알려져 있음에도 불구하고, 신청인의 지질학적 증거가 Harman Road 필지에서 실현 가능한 채석장을 개발할 수 있음을 입증하기에 불충분하다고 판단하였다. 법원은 나아가 법률적 허용성과 물리적 가능성이 인정된다고 가정하더라도, 그 분석의 나머지 요건이 충족되지 않는 한 그것만으로는 결정적일 수 없다고 판시하였다.

감정평가사에 대한 시사점. 이 판결은 4 요건 최유효이용 심사를 규율에 따라 순차적으로 적용한 하나의 모범 사례이다. 이는 *Ranch Springs, LLC v. Commissioner*, 164 T.C. No. 6 (2025) 및 *Beaverdam Creek Holdings, LLC v. Commissioner*, T.C. Memo. 2025-53 에서 확립된 입장 — 제안된 용도의 일반적 실현 가능성이 주변 지역에서 문서로 뒷받침되더라도 조세법원은 투기적인 최유효이용 주장을 받아들이지 않는다는 입장 — 을 강화한다. 지역권 감정평가를 작성하는 감정평가사는 지역 차원의 또는 유추에 의해 뒷받침되는 실현 가능성에 의존할 것이 아니라 각 요건마다 해당 필지에 특유한 증거를 마련하여야 하며, 감정평가 의뢰와 같은 시기에 취득된 용도지역 변경과 같은 시간적 패턴에 대해서는 사법적 회의를 불러올 수 있으므로 특히 신중하여야 한다.

Maryland(메릴랜드주)

Baltimore(볼티모어시) 의 Pimlico Race Course(핍리코 경마장) 취득. 토지수용 절차, 2025-2026 년 진행 중.

Baltimore 시가 수용권을 통해 Pimlico Race Course 를 취득한 사안은 특수용도 시설에서 계속기업 가치와 부동산 가치의 관계에 관한 정당보상 문제를 제기한다. Pimlico 의 영업은 그 물적 시설과 기능적으로 분리될 수 없기 때문에 통상적인 부동산 비교매매 분석은 효용이 제한적이며, 수익방식 평가는 경마 영업이 계속된다는 가정에 크게 의존한다.

감정평가사에 대한 시사점. 경마장, 영화 촬영소, 테마파크, 마리나, 그리고 일부 의료 또는 숙박 자산과 같이 영업과 부동산이 기능적으로 분리될 수 없는 단일목적 시설의 평가를 의뢰받는 감정평가사라면 이 절차를 주시할 만하다. 소송 절차에서 도출되는 방법론은 향후 수용 사건에서 정당보상 산정이 계속기업 증분을 어떻게 다룰지에 영향을 미칠 가능성이 높다.

감정평가사를 상대로 한 연방 공정주거법 청구

Braxton v. Bank of America. 2026 년 4 월 제기.

한 주택 소유자가 Bank of America(뱅크오브아메리카) 와 Virginia(버지니아주) 소재 감정평가사를 상대로 Fair Housing Act(공정주거법) 위반의 인종차별을 주장하며 소를 제기하였다. 원고는 감정평가 현장에 본인과 임차인들이 있었으며 감정평가사의 행위가 인종적 적의를 드러냈다고 주장한다. 이 사건은 Maryland 연방지방법원의 *Connolly/Mott v. loanDepot* 사건과 캘리포니아주 Marin County(마린 카운티) 의 *Austin v. Miller* 사건을 비롯하여 앞선 감정평가 편향 소송에서 형성된 유형, 즉 개인 원고가 감정평가사, 감정평가관리회사, 대출기관을 상대로 공정주거법 청구를 제기하는 유형을 이어 간다.

감정평가사에 대한 시사점. 개별 감정평가사를 상대로 한 감정평가 편향 사건의 제기가 계속되는 추세에 더하여 차별적 영향 (disparate impact) 집행에서 연방 차원의 후퇴가 겹치면서, 책임 노출의 지형이 달라졌음을 알 수 있다. 이제 감정평가 편향 책임의 주된 원천은 연방기관이 아니라 사인 원고와 주 차원의 공정주거 집행기관이다. 감정평가사는 비교사례 선정, 보정치 도출, 최종가치 조정이 사후 심사에 대하여 독립적으로 방어될 수 있도록 문서화 실무를 점검하여야 하며, 조사 현장에서의 발언이나 행동 자체가 원고 주장의 사실적 근거가 되는 경우가 빈번하다는 점을 인식하여야 한다.

최근 법원 판결 — 유럽연합

최근 법원 판결 • *The Valuation Engineer*

William Bert Craytor • 제 1 권 제 1 호 • 2026 년 여름

DOI: 10.67330/rlynly02

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

범위와 출처

이 개관은 접근 가능한 경우 공개된 1 차 자료에 근거하였다 — 특히 유럽연합사법재판소의 공식 판례 데이터베이스 (curia.europa.eu), 유럽인권재판소의 HUDOC, 연방재정법원 (Bundesfinanzhof) 의 공간 (公刊) 판결 (bundesfinanzhof.de), 독일 연방헌법재판소 (bverfg.de), 프랑스 판례에 대한 Légifrance, 그리고 영국의 Upper Tribunal (Lands Chamber) — 아울러 전문 및 학술 간행물의 2 차 논평을 참조하였다. 각국 언어로 된 판결은 원문으로 읽고 영문으로 서술하였으며, 사건명은 원래의 철자와 발음구별부호를 그대로 유지한다.

유료 구독 전용 법률 조사 데이터베이스 (juris.de, beck-online, Westlaw UK, Lamy, Aranzadi) 는 사용하지 않았다. 최고법원 아래 단계의 국내 법원 판결은 선별적으로 다루었으며, 선례적 또는 방법론적 중요성이 분명한 판결에 우선순위를 두었다. 여기에 실린 판례 인용은 실무나 추가 출판에 의존하기에 앞서 해당 법원의 공식 판례집과 대조하여 확인하여야 한다.

이 기준 회차는 Germany(독일), France(프랑스), United Kingdom(영국) 을 주된 국가별 관할로 다룬다. Italy(이탈리아), Spain(스페인), Netherlands(네덜란드) 및 북유럽 국가들은 이번 회차에서 별도로 다루지 않았으나 자료가 축적되는 대로 이후의 분기별 회차에서 추가될 것이다. Switzerland(스위스) 는 유럽연합 밖에 있으나 스위스 감정평가 기준이 유럽 실무 전반에 미치는 방법론적 영향이 크므로 향후 회차에서 다룰 예정이다.

유럽의 규제 환경

유럽연합 차원의 두 가지 규제 동향은 그 자체가 법원 판결은 아니지만 감정평가사의 업무 환경에 실질적인 영향을 미친다.

EU AI Act 와 자동가치산정모형. Regulation (EU) 2024/1689(EU AI Act, 유럽연합 인공지능법) 은 2024 년에 발효되었으며, 고위험 인공지능 시스템에 적용되는 규정은 2026 년 8 월까지 단계적으로 시행된다. 2026 년 5 월 19 일 유럽집행위원회는 고위험 인공지능 시스템의 분류에 관한 지침 초안을 이해관계자 의견 수렴을 위해 공표하였다. AI Act 부속서 III 은 자연인과 관련된 신용도 평가를 고위험으로 추정되는 여덟 개 사용 사례 범주 가운데 하나로 열거하고 있으며, 집행위원회의 지침 초안은 이 범주를 상세히 전개한다. 저당대출 실행, 저당대출 포트폴리오 모니터링, 보험 인수에 사용되는 자동가치산정모형 (AVM) 은 해당 대출의 신용도 평가와 기능적으로 분리될 수 없으므로, 유럽연합 내 신용기관 또는 보험자가 운용하는 AVM 은 대부분의 구성에서 고위험 인공지능 시스템으로 분류될 것이다. 준수에 따르는 결과로는 의무적 위험관리체계, 데이터 거버넌스 통제, 최소 6 개월의 로그 보존을 수반하는 자동 로깅, 운영자가 시스템의 결정을 이해하고 무효화할 수 있도록 하는 인적 감독 장치, 그리고 시판 후 모니터링 의무가 포함된다.

감정평가사에 대한 시사점. 유럽의 감정평가사는 대출기관 고객이 운용하는 AVM 기반 업무 흐름에 문서화 및 설명가능성 요건이 크게 강화될 것을 예상하여야 한다. 감정평가사가 모형 입력값, 보정 데

이터, 또는 AVM 산출물에 대한 인적 감독 검토를 제공하는 경우, 그러한 활동 자체가 대출기관의 고 위험 AI 준수 태세를 구성하는 요소가 된다.

건물에너지성능지침 (개정). Directive (EU) 2024/1275, 즉 개정 건물에너지성능지침 (Energy Performance of Buildings Directive, EPBD) 은 2024 년 5 월 8 일 관보에 게재되어 2024 년 5 월 28 일 발효되었다. 회원국의 국내법 전환 기한은 2026 년 5 월 29 일이다. 개정 EPBD 는 여러 규정 가운데 회원국으로 하여금 성능이 가장 낮은 비주거 건물 재고의 16 퍼센트와 주거 건물 재고의 43 퍼센트에 대한 개보수를 우선 추진하도록 하고, 최소에너지성능기준 (MEPS) 을 설정하며, 2028 년부터 신축 공공건물에, 2030 년부터 모든 신축 건물에 무배출 건물을 요구하고, 2030 년부터 신축 건물에 대한 전생애탄소 평가를 의무화한다.

감정평가사에 대한 시사점. 개정 EPBD 는 여러 경로를 통해 부동산 평가에 영향을 미칠 것이다. 최소 에너지성능 기준치는 규제 준수에 기인하는 새로운 유형의 진부화를 낳고, 에너지성능인증서 (EPC) 등급은 주거용과 상업용 재고 모두에서 가격을 결정하는 중요한 요인이 되며, 규제 준수를 위한 개보수 비용은 원가방식과 잔존 경제적 내용연수의 공제에 반영되어야 한다. 감정평가사는 또한 주택신용지침 (Mortgage Credit Directive) 의 적용을 받는 대출기관들이 담보평가에서 EPC 등급과 MEPS 관련 자본적 지출 요건을 명시적으로 다룰 것을 점점 더 요구하리라는 점을 예상하여야 한다.

다음 회차를 위한 메모

이 기준 회차에 이어지는 분기별 회차는 변동 사항에 초점을 맞춘다. 새로운 판결, 여기에서 확인된 계속 중인 사안의 처리 (특히 독일 연방헌법재판소의 1 BvR 472/26 처리), 그리고 개정 EPBD 를 국내법으로 전환하는 회원국의 이행 입법이 그것이다. 스위스의 법체계, 이탈리아와 스페인 최고법원 판결의 취급, 프랑스 *Conseil constitutionnel* (헌법위원회) 관련 동향은 자료가 축적되는 대로 추가될 것이다.

관할별 판례 요약

Court of Justice of the European Union(유럽연합사법재판소)

GR Real (Case C-351/23). Court of Justice of the European Union(유럽연합사법재판소), 2025 년 6 월 24 일 판결.

사법재판소는 슬로바키아 법원의 선결적 부탁에 대하여 판단하였다. 이 사건은 잠재적으로 불공정한 계약 조항을 다투는 소송이 아직 계속 중인 상태에서 소비자들이 경매를 통해 가족의 주거를 상실한 재판 외 저당권 실행 절차에서 비롯되었다. 재판소는 Directive 93/13/EEC(불공정계약조항지침) 상의 소비자보호 보장은 저당 잡힌 주거가 재판 외 실행을 통해 매각되기 전에 계약 조항의 공정성에 대한 실효적인 사법심사가 가능하도록 국내 절차규범이 정하여질 것을 요구하며, 그러한 심사가 완료되기 전에 매각이 되돌릴 수 없게 진행되도록 허용하는 절차는 위 지침과 양립할 수 없다고 판시하였다.

감정평가사에 대한 시사점. 이 판결은 직접적인 평가 사건은 아니지만, 저당담보 평가가 의뢰되고 사용되는 절차적 환경을 재편한다. 재판 외 실행 절차를 두고 있는 여러 회원국 (특히 Slovakia(슬로바키아), Czech Republic(체코), Hungary(헝가리) 및 발트 3 국) 에서 실질적인 효과는 실행 경매 절차를 지연시키고 대출기관의 담보평가가 원용되는 기간을 늘리는 것이다. 이들 관할에서 실행 목적의 담보평가를 작성하는 감정평가사는 평가보고서가 종전보다 더 오랜 기간 사용될 수 있고, 최초 기준 시점과 최종 매각 사이의 시장 변동에 비추어 검증받을 수 있음을 예상하여야 한다.

European Court of Human Rights(유럽인권재판소)

Baku(바쿠) 건물 철거 사건 (Azerbaijan, 아제르바이잔), 2025 년 7 월 판결. European Court of Human Rights(유럽인권재판소), 제 1 추가의정서 제 1 조.

유럽인권재판소는 당국이 지진 위험을 들어 명령한 Baku 소재 건물의 철거 및 수용과 관련하여 재산의 평온한 향유권 침해를 인정하였다. 스트라스부르 재판소는 국내 당국이 보상을 결정하면서 신청인들이 제출한 평가보고서를 배척한 데 대해 충분한 이유를 제시하지 않았고, 법원이 스스로 의뢰한 감정의견의 금액만을 정당한 근거 없이 채택하였다고 판단하였다. 재판소는 재산적 손해에 대하여 총 €75,000, 비재산적 손해에 대하여 €3,000, 소송비용에 대하여 €1,055 를 지급하도록 하였다.

감정평가사에 대한 시사점. 이 판결은 수용 절차에서 국내 법원이 상충하는 평가 증거를 이유를 갖추어 고려하여야 하고, 법원이 의뢰한 감정의견을 택하여 신청인의 평가보고서를 배척하는 것 자체에도 이유가 제시되어야 한다는 제 1 추가의정서 제 1 조 법리의 흐름을 강화한다. 수용 및 정당보상 영역에서 활동하는 유럽의 감정평가사에게 이 사건은, 평가보고서의 문서상 방어가능성 — 비교사례 선정, 보정치 도출, 최종가치 조정을 뒷받침하는 명시적 논거 — 이 그 보고서가 국가를 상대로 한 청구를 뒷받침하기 위해 제출될 때에는 그 자체로 기본권의 문제가 된다는 점을 부각한다.

France(프랑스)

Cour de cassation, Chambre commerciale, 19 June 2024, n° 22-24.169. (프랑스 파기원 상사부. 2025 년까지 전문가 논평에서 다루어짐.)

파기원은 상속세상 *valeur vénale*(발뢰르 베날, 시장가치) 분쟁의 맥락에서, 과세 원인 사실 (피상속인 *de cujus* 의 사망) 이후에 이루어진 상속 부동산의 매매는 상속인들이 당초 신고한 *valeur vénale* 를 하향 수정하는 데 사용될 수 없으며, 이는 사후 매매 가격이 현저히 낮고 매매가 사망 후 수개월 내에 이루어진 경우에도 마찬가지라고 판시하였다. 이 판결은 상속세 목적의 *valeur vénale* 가 과세 원인 사실의 시점을 기준으로 결정되며 같은 부동산의 사후 거래를 참조하여 수정되지 않는다는 원칙을 적용한 것이다.

감정평가사에 대한 시사점. 이 판결은 프랑스 상속세 실무에서 엄격한 기준시점 원칙을 확인하고, 가치에 관한 소급 증거로서 사후 매매가 갖는 관련성을 제한한다. 프랑스 상속세 평가를 작성하는 감정평가사는 기준시점 또는 그 이전의 거래에 기초하여 포괄적인 비교매매 증거를 확보하여야 하며, 대상 부동산 자체의 사후 매매를 확인 증거로 제시하여서는 안 된다. 같은 논리는 증여세 (*droits de donation*, 드루아 드 도나시옹) 와 부유세 (*IFI*) 영역으로도 확장될 가능성이 높다.

Cour de cassation, 3^e Chambre civile, 8 January 2026, n° 23-22.803. 화해적 (비사법적) 감정보고서.

프랑스 파기원 제 3 민사부는 프랑스 민사절차에서 화해적 (비사법적) 감정보고서의 법적 가치를 재정의하는 판결을 내렸다. 이 판결은 조정과 화해적 분쟁해결을 민사절차의 원칙으로 삼은 2025 년 7 월 18 일 *décret*(데크레, 시행령) 에 의한 보다 폭넓은 절차 개혁에 뒤이은 것이다. 재판소는 일방 당사자의 계약상 의뢰로 작성된 화해적 감정보고서도 대심주의 원칙과 감정인의 독립성 및 방법론에 대한 법원의 평가를 조건으로 증거로 채택되고 증명력을 인정받을 수 있음을 분명히 하였다.

감정평가사에 대한 시사점. 이 판결은 법원이 선임한 사법감정인 (*experts judiciaires*, 엑스페르 쥐디시에르) 의 틀 밖에서 작성된 *expertises immobilières*(엑스페르티즈 이모빌리에르, 부동산 감정) 의 증거법상 지위를, 보고서가 직역의 방법론 기준을 준수하고 감정인의 독립성이 문서로 뒷받침되는 것을 조건으로 강화한다. 상속, 이혼, 상사 분쟁에서 다투어지는 평가를 수행하는 프랑스 감정평가사에게 이 판결은 고객 의뢰 보고서가 후속 소송에서 인용되고 원용될 수 있는 더 탄탄한 기반을 마련해 준

다. 그만큼 독립성 (고객과의 종속관계, 경제적 이해관계 또는 친족관계의 부존재) 을 문서화하고 직역 기준이 정한 방법을 적용하는 데 각별한 주의를 기울이는 것이 더욱 중요해진다.

Germany(독일)

Bundesfinanzhof, 2025 년 11 월 12 일 판결 (II R 25/24, II R 31/24, II R 3/25). Grundsteuer Bundesmodell(그룬트슈토이어 연방모델) 3 부작 — 개편된 재산세 평가 체계의 합헌성. 독일 연방재정법원 (Bundesfinanzhof, BFH) 제 2 부는 세 건의 병행 상고 절차에서, 2025 년 1 월 1 일부터 시행된 개편 재산세 (Grundsteuer, 그룬트슈토이어) 의 목적상 독일 11 개 주에서 주거용 부동산 평가에 적용되는 이른바 Bundesmodell(연방모델) 하의 수익환원 평가방법 (Ertragswertverfahren, 에르트라크스베르트페어파렌) 에 관한 규정이 합헌이라고 판시하였다. 원고들은 North Rhine-Westphalia(노르트라인베스트팔렌), Saxony(작센), Berlin(베를린) 의 아파트 소유자들로서, 획일적인 평가 규정이 일반적 평등원칙 (기본법 제 3 조 제 1 항) 에 위반되고 체계적으로 부정확한 가액을 산출한다는 이유로 자신들에 대한 과세평가를 다투었다.

BFH 는 입법자가 개념적으로 시장가치 지향을 따르는 (konzeptionell einer Verkehrswertorientierung folgt) 평가 체계를 만들었고, 이는 평가 대상 전체에 걸쳐 평균적으로 통상가치 (gemeiner Wert, 게마이너 베르트) 의 범위 안에서 “객관적으로 현실적인 부동산 가액” 에 근접하도록 설계되었다고 판시하였다. 재판소는 입법자가 평가 제도의 목표와 그러한 제도에 필연적으로 수반되는 불균등을 형량함에 있어 자신에게 허용된 재량의 범위를 넘지 않았으며, 자동화된 주기적 갱신을 통해 새로운 “평가 적체” 를 피하는 데 상당한 비중을 둔 것이 헌법상 허용된다고 보았다. BFH 는 또한 납세자가 2024 년 5 월 27 일자 종전 BFH Beschlüsse(베슬뤼세, 결정) (II B 78/23 및 II B 79/23) 로 확립된 바에 따라 § 198 BewG(평가법 제 198 조) 에 의한 적격 감정평가를 통해 더 낮은 통상가치 (gemeiner Wert) 를 입증할 수 있는 선택지를 보유한다는 점을 강조하면서, 다만 그 낮은 가액은 공적으로 선임된 부동산 감정평가사가 인정된 평가방법을 적용하여 작성한 감정평가보고서로 뒷받침되어야 한다는 제한이 따른다고 하였다.

Bund der Steuerzahler Deutschland(독일납세자연맹) 와 Haus & Grund Deutschland(독일주택토지 소유자협회) 가 지원하는 헌법소원 (Verfassungsbeschwerde, 페어파송스베슈베르테) 이 현재 사건번호 1 BvR 472/26 으로 연방헌법재판소 (Bundesverfassungsgericht) 에 계속 중이다.

감정평가사에 대한 시사점. 이 3 부작은 개편 대상 기간 중 가장 중요한 독일 부동산 평가 판례이다. 감정평가사에게 주된 실무적 결과는, Bundesmodell 에 따라 산정된 Grundsteuerwert(그룬트슈토이어베르트, 재산세 평가액) 가 대상 부동산의 실제 시장가치를 초과하는 것으로 보이는 사안에서 § 198 BewG 상의 “저가” 감정평가 (Verkehrswertgutachten zum Nachweis des niedrigeren gemeinen Werts) 에 대한 수요가 높아진다는 점이다. BFH 가 이 경로를 명시적으로 승인하면서도 Bundesmodell 자체에는 일반적으로 존중을 표한 결과, 감정평가보고서는 이제 체계적 과대평가로부터 개별 납세자를 구제하는 실효적 수단이 되었다. 감정평가사는 그러한 보고서의 의뢰가 지속적으로 늘어날 것을 예상하고 ImmoWertV 2021(부동산가치평가령 2021) 방법론과 § 198 BewG 의 요건을 준수하도록 하여야 한다. BVerfG 에 계속 중인 헌법소원 (1 BvR 472/26) 은 이 틀에 추가 변화가 있을지를 판단하는 데 중요하므로 주시할 필요가 있으며, 다음 분기 회차의 주요 주제가 될 것으로 예상된다.

Bundesfinanzhof, 2025 년 10 월 7 일 판결, IX R 26/24. 토지와 건물 간 매매대금 배분 (Kaufpreisaufteilung, 카우프프라이스아우프타일룽). 주거용 임대 부동산의 감가상각 (Absetzung für Abnutzung, AfA) 기초가액이 문제된 소득세 사건에서, 연방재정법원 제 9 부는 계약상 매매대금을 감가상각 대상이 아닌 토지 부분 (Grund und Boden) 과 감가상각 대상인 건물 부분 (Gebäude) 으로 배분하는 방법론을 다투었다. 원고들은 대상 부동산의 특정 위치에 대하여 기초 자료가 신뢰할 만하게 문서화되어 있지 않다는 이유로, 원심이 Bodenrichtwerte (보덴리히트베르테, 기준지가) 에 의존한 것

을 다투었다.

감정평가사에 대한 시사점. 이 판결은 매매대금 배분에 관하여 ImmoWertV 가 인정된 평가 원칙의 근거임을 확인하는 BFH 판례의 흐름 (특히 앞서 인용한 2025 년 11 월 12 일 BFH II R 25/24, 그리고 그 이전의 2022 년 9 월 20 일 BFH IX R 12/21) 을 잇는다. 이 사건은 또한 *Bodenrichtwerte* 의 기초 자료가 갖는 중요성을 보여준다. 즉 *Gutachterausschüsse*(구타흐터아우스쉬세, 감정평가위원회) 가 특정 위치에 대하여 충분한 비교매매 자료를 보유하고 있지 않은 경우, *Bodenrichtwert* 는 자료의 질을 이유로 다투어질 수 있다. 매매대금 배분이나 감가상각 기초가액 의견을 작성하는 감정평가사에게 이는 *Bodenrichtwert* 의 자료 기반 — 그 바탕이 되는 *Kaufpreissammlung*(카우프프라이스잠룽, 매매가격 집적자료) 을 포함하여 — 을 문서화하는 것이 방어 가능한 분석의 일부가 된다는 뜻이다.

United Kingdom(영국)

Mehrunnisa Ali Khan Ahmed Khan v Wolverhampton City Council. Upper Tribunal (Lands Chamber)(상급심판소 토지부), 2025 년.

Upper Tribunal (Lands Chamber) 는 강제매수 보상액을 산정하기 위한 정당한 기준시점이 일반귀속선언 (general vesting declaration, GVD) 에 따라 부동산이 취득기관에 귀속되는 날임을 확인하였다. 심판소는 동시대 평가와 소급 평가의 관계도 다루었는데, 이 사건에서 지방정부가 GVD 직후에 실시한 최초 평가는 £150,000 였던 반면, 소유자의 통지 3 개월 후를 귀속일로 가정하여 심판 절차 중인 2025 년에 작성된 두 번째 평가는 £120,000 였다. 심판소는 법률상 기준시점을 확인하였음에도, GVD 가 이루어진 시점에 시간적으로 더 가까운 시기에 산정된 금액을 채택하여 청구인에게 더 높은 보상이 지급되는 결과가 되었다.

감정평가사에 대한 시사점. 이 판결은 강제매수 영역에서 활동하는 영국 감정평가사에게 방법론적으로 시사하는 바가 크다. 법률상 기준시점은 귀속일이지만, 동일한 법률상 기준시점에 대하여 수년 후에 작성된 소급 평가보다 동시대 평가를 심판소가 선호하였다는 점은, 동시대의 시장 증거가 귀속일 자체에서는 형식적으로 채택될 수 없는 것이라 하더라도 실질적인 비중을 갖는다는 것을 보여준다. 감정평가사는 GVD 에 가까운 시기에 작성된 평가보고서가 1 차 기록으로 기능할 수 있도록 충분한 뒷받침 증거와 함께 문서화되도록 하여야 하며, 이후의 심판 절차가 소급적 재구성보다 동시대 보고서를 우선하여 원용할 수 있음을 예상하여야 한다.

최근 법원 판결 — 라틴아메리카

최근 법원 판결 • *The Valuation Engineer*

William Bert Craytor • 제 1 권 제 1 호 • 2026 년 여름

DOI: 10.67330/ywghzg47

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

범위와 출처

이 개관은 접근 가능한 경우 공개된 1 차 자료에 근거하였다 — 특히 미주인권재판소 판례 저장소 (corteidh.or.cr), 브라질 연방대법원 (STF) 과 고등법원 (STJ) 포털 (portal.stf.jus.br, stj.jus.br), 멕시코 *Semanario Judicial de la Federación* (세마나리오 후디시알 데 라 페데라시온, 연방사법주보, sjfsemanal.scjn.gob.mx), 아르헨티나 *Corte Suprema de Justicia de la Nación*(국가대법원, csjn.gov.ar), 칠레 *Poder Judicial*(사법부) 포털 (poderjudicial.cl) 및 *Diario Constitucional*, 그리고 콜롬비아 *Corte Constitucional* (헌법재판소, corteconstitucional.gov.co) — 아울러 전문 및 학술 간행물의 2 차 논평을 참조하였다. 스페인어와 포르투갈어로 된 판결은 원문으로 읽고 영문으로 서술하였으며, 사건명은 원래의 철자와 발음구별부호를 그대로 유지한다.

유료 구독 전용 법률 조사 데이터베이스 (vLex, La Ley, Thomson Reuters Proview Latin America) 는 사용하지 않았다. 최고법원 아래 단계의 국내 법원 판결은 선별적으로 다루었으며, 선례적 또는 방법론적 중요성이 분명한 판결에 우선순위를 두었다. 여기에 실린 판례 인용은 실무나 추가 출판에 의존하기에 앞서 해당 법원의 공식 판례집과 대조하여 확인하여야 한다.

이 기준 회차는 Argentina(아르헨티나), Brazil(브라질), Chile(칠레), Colombia(콜롬비아), Mexico(멕시코) 를 주된 국가별 관할로 다룬다. Peru(페루), Venezuela(베네수엘라), Ecuador(에콰도르), Bolivia(볼리비아), Paraguay(파라과이), Uruguay(우루과이) 및 중미와 카리브해 관할은 이번 회차에서 별도로 다루지 않았으나 자료가 축적되는 대로 이후의 분기별 회차에서 추가될 것이다. 베네수엘라의 재산 원상회복과 쿠바 제재 관련 평가 업무는 그 방법론적 특수성과 국경 간 실무에 대한 관련성 때문에 향후 회차에서 다룰 예정이다.

지역적 맥락

유럽연합 개관과 달리, 이 라틴아메리카 개관에는 EU AI Act 와 개정 건물에너지성능지침이 대표하 는 초국가적 규제 층위에 직접 대응하는 것이 없다. 라틴아메리카의 부동산 감정평가 규제는 압도적으로 국가 단위이며, 주된 초국가적 영향은 미주인권협약 제 21 조 (재산권) 에 관한 미주인권재판소의 기본권 법리와, 재무보고 목적으로 역내 대부분의 관할이 채택한 국제회계기준 (특히 IFRS 13, 공정가치 측정) 이다. 그럼에도 두 가지 방법론적 주제가 역내 판례 전반에 걸쳐 흐르고 있어 여기에서 언급할 만하다.

원주민 공동체 재산의 평가. 라틴아메리카 감정평가 실무에서 가장 특징적인 방법론적 과제는 원상회복, 수용, 보상의 맥락에서 원주민과 아프리카계 후손 공동체 재산을 평가하는 일이다. 미주인권재판소의 판례는 원주민 공동체 재산이 개인의 사유재산과 동일한 지위에서 헌법적으로 보호된다고 일관되게 판시해 왔으나, 그러한 재산을 평가하는 방법 — 통상 독립당사자 간 비교거래가 존재하지 않고, 재산이 공동으로 보유되며 양도가 불가능하고, 영역이 갖는 문화적·정신적 의미가 그 가치에 본질적으로 결합되어 있는 상황에서 — 은 여전히 충분히 발전되지 못하였다. 2025 년 말 기준으로 원주

민에게 유리한 미주인권재판소 판결 24 건이 미이행이거나 부분적으로만 이행된 상태로 남아 있으며, 그중 여럿은 해결되지 않은 평가 방법론에 좌우되는 보상액 산정을 포함하고 있다.

수용 보상에서의 인플레이션, 통화, 물가연동. 역내 여러 관할, 특히 아르헨티나에서는 평가 방법론과 보상액 산정이 고인플레이션, 외화 표시 *tasaciones*(타사시오네스, 감정평가), 복잡한 물가연동 제도에 대응하여야 한다. 아르헨티나 대법원은 당초 외화로 표시된 수용 보상을 인위적인 등가 환율로 자국 통화로 환산할 수 없으며, 평가는 보상 시점의 부동산의 실질 경제가치를 반영하여야 한다고 판시하였다. 유사한 문제는 베네수엘라 (원상회복 및 제재 관련 평가와 관련하여) 에서, 그리고 브라질에서도 고인플레이션 시기마다 주기적으로 제기된다. 역내 여러 관할에 걸쳐 활동하는 감정평가사에게 이러한 인플레이션·통화 고려사항은 유럽이나 미국에는 대응물이 없는 방식으로 평가 방법론에 본질적으로 결합되어 있다.

다음 회차를 위한 메모

이 기준 회차에 이어지는 분기별 회차는 변동 사항에 초점을 맞춘다. 새로운 미주인권재판소 판결, 여기에서 확인된 계속 중인 사안의 처리, 2027 년 운영 개시를 향한 콜롬비아 *Jurisdicción Agraria y Rural*(후리스딕시온 아그라리아 이 루랄, 농지·농촌 관할) 시행의 진전, 그리고 IFRS 13 공정가치 측정 실무에 관한 각국의 동향이 그것이다. 페루, 베네수엘라, 에콰도르 및 중미 관할은 자료가 축적되는 대로 추가될 것이다. 특히 원주민 공동체 재산 평가의 방법론적 발전에 주목할 것인데, 이 영역에서는 미주인권재판소의 실체적 판시가 감정평가사와 보상 당국이 활용할 수 있는 실무적 평가 방법론을 앞질러 가고 있다.

관할별 판례 요약

Inter-American Court of Human Rights(미주인권재판소)

***Maya Q'eqchi' Agua Caliente vs. Guatemala* 및 *Comunidad Garífuna de San Juan y sus Miembros vs. Honduras*.** Inter-American Court of Human Rights(미주인권재판소), 2024 년 판결, 2025 년까지의 역내 판례에서 논의됨.

미주인권재판소는 두 건의 동시대 판결에서 원주민 및 아프리카계 후손 공동체에 속하는 조상 대대의 영역을 경계 확정하고 권원을 부여하며 그 박탈을 방지할 Guatemala(과테말라) 와 Honduras(온두라스) 의 의무에 관하여 판단하였다. Maya Q'eqchi' Agua Caliente 공동체는 45 년에 걸친 행정적·사법적 조치를 거친 끝에 비로소 자기 영역에 대한 확정적 소유권원을 취득하였고, 온두라스 San Juan 의 Garífuna 공동체는 조상 대대의 연안 토지를 계속 박탈당하는 상황에 놓여 있었다. 두 판결 모두 원주민 공동체 재산을 인정하고 경계를 확정하며 보호할 미주인권협약 제 21 조 (재산권) 상 국가의 의무를 다루었다. 역내 논평자들은 이들 판결에서 미주인권재판소의 원상회복 틀이 이전 판결들보다 다소 제한적이며, 특히 원주민의 권리 주장을 알지 못한 채 중첩되는 권리를 취득한 “선의의 제 3 자” (*terceros inocentes*, 테르세로스 이노센테스) 의 취급에서 그러하다고 지적하였다.

감정평가사에 대한 시사점. 이들 판결은 원주민과 아프리카계 후손의 공동체 재산이 미주 체계 전반에 걸쳐 헌법적으로 보호되며, 원상회복이나 보상의 맥락에서 그러한 재산의 평가가 방법론적 보완 없이 통상적인 독립당사자 간 비교매매 (대개 존재하지 않는다) 에 의존할 수 없음을 재확인한다. 원주민 원상회복 맥락에서 의뢰를 받는 감정평가사는 평가가 세 가지 별개의 요소를 통합하여야 함을 예상하여야 한다. 즉 해당 지역에서 물리적 특성이 비교 가능한 토지의 기초 시장가치, 원주민의 이용에 기인하는 기회비용과 개량, 그리고 미주인권재판소 판례가 재산권의 보상 가능한 한 차원으로 다루는, 그 영역이 갖는 비경제적인 문화적·정신적 의미가 그것이다. 세 번째 요소에 관한 방법론은 여전히 충분히 발전되지 못하였으며 미주인권재판소 논증의 지속적인 주제가 되고 있다.

Mapuche / Consejo de Todas las Tierras vs. Chile. Inter-American Court of Human Rights(미주인권재판소), 2024 년 판결 (2025 년 판례에서 논의됨).

미주인권재판소는 Consejo de Todas las Tierras(콘세호 데 토다스 라스 티에라스, 모든 토지 평의회)의 영역 회복 운동 (1989-1992) 과 관련하여, 그리고 원주민의 영역 주장에 반테러법을 적용한 것과 관련하여 Mapuche(마푸체) 족 구성원 135 인에 대한 인권 침해를 이유로 칠레 정부를 단죄하였다. 이 판결은 *Norín Catrimán and others vs. Chile*(2014) 법리를 확장하고, 영역에 관한 주장과 결부된 원주민의 자결권을 명시적으로 인정한다.

감정평가사에 대한 시사점. 이 사건의 주된 판시는 평가 방법론 자체보다 형사절차상 보호에 관한 것이지만, 영역과 결부된 원주민의 자결권 인정은 칠레의 원주민 영역 원상회복 및 관련 맥락에서의 평가 업무에 영향을 미치는 절차적 충위를 만들어 낸다. 그러한 맥락에서 의뢰를 받는 감정평가사는 방법론이 적절한 협의 절차를 포함하고, 자기 영역의 가치가 어떻게 성격 규정될 것인지에 대한 공동체의 자결권을 존중하도록 하여야 한다.

Argentina(아르헨티나)

Tribunal de Tasaciones de la Nación 에 관한 아르헨티나 법리. 2025 년까지의 판례를 통해 정착됨.

아르헨티나 국가대법원 (CSJN) 과 하급 연방법원들은 수용 및 *expropiación inversa*(엑스프로피아시온 인베르사, 역수용) 사안에서 Tribunal de Tasaciones de la Nación(트리부날 데 타사시오네스 데 라 나시오, 국가감정평가심판원, TTN) 이 수행한 평가를 둘러싼 법리를 계속 발전시켜 왔다. 확립된 CSJN 법리 (Fallos 313:82, 2024-2025 년까지 반복 적용) 는 TTN 평가가 심판원 구성원의 기술적 역량, 그 평가가 근거하는 증거 기반, 그리고 실무의 통일성을 이유로 추정적 비중을 가지며, TTN 평가에서 벗어나려면 명백한 오류나 누락의 입증이 필요하다고 본다. 2024-2025 년 이 법리를 적용한 연방법원들 — Yacyretá(야시레타) 수력발전 저수지에서 비롯된 *expropiación inversa* 사건인 *Sosa v. Entidad Binacional Yacyretá*(2024 년 6 월) 의 Juzgado Federal de Oberá(오베라 연방법원) 를 포함하여 — 은 보상의 실효적 기준으로 TTN 평가에 계속 의존하고 있다.

감정평가사에 대한 시사점. 아르헨티나 감정평가사에게 실무적 결과는 TTN 평가에 반대하여 제출되는 사실 감정평가보고서가 상당한 증명 부담을 진다는 점이다. TTN 평가를 배척하거나 수정하려면 사실 보고서는 단지 대안적인 평가 결론이 아니라 TTN 작업의 방법론적 오류나 누락을 입증하여야 한다. 그러한 보고서를 작성하는 감정평가사는 가액 차이의 크기만을 다룰 것이 아니라 TTN 의 비교 매매 선정, 보정치 도출, 지역 시장 상황의 취급에 방법론적 비판을 집중하여야 한다.

외화 표시 tasaciones 와 보상의 등가성. 당초 외화로 표시된 수용 보상을 인위적인 1 대 1 등가로 아르헨티나 페소로 환산할 수 없다는 CSJN 법리는 2025 년까지 유효하게 유지되고 있다. 이 법리는 인플레이션이나 통화정책으로 인하여 보상 시점의 부동산 실질 경제가치를 반영하지 못하는 보상으로부터 수용된 소유자를 보호한다. 이 원칙은 장기간 진행되는 수용 절차와 관련하여 2024-2025 년에도 주기적으로 재확인되었다.

감정평가사에 대한 시사점. 상당한 인플레이션이나 통화정책 변화 기간에 걸친 수용 사안의 평가를 작성하는 아르헨티나 감정평가사는 평가의 통화 기준을 명시적으로 문서화하고, 기준시점과 장래의 지급 예정 시점 사이에 적용할 물가연동 방법론을 검토하여야 한다. 인플레이션과 통화 위험을 다루지 않은 보고서는 CSJN 보상 법리를 적용하는 법원을 설득하기 어려울 것이다.

Brazil(브라질)

Supremo Tribunal Federal, RE 922.144(일반적 파급효 Tema 865). 브라질 연방대법원, 2024 년 판결 및 2025 년 계속 적용.

STF 는 *desapropriação*(데자프로프리아상, 수용) 절차에서의 보상이 연방헌법 제 100 조가 정한 연방 *precatório*(프레카토리우, 확정판결 채무 지급명령) 제도와 헌법적으로 양립한다고 판시하였다. 구체적으로, 공공기관이 *imissão provisória na posse*(이미상 프로비조리아 나 포시, 잠정적 점유이전) 를 위하여 최초로 공탁한 금액과 사법 감정으로 최종 확정된 금액의 차액은 즉시 이루어지는 추가 공탁 이 아니라 *precatório* 제도를 통해 지급되어야 한다. 이 판결은 *justa, prévia e em dinheiro*(정당하고 사전적이며 금전으로 지급되는) 보상이라는 헌법상 명령 (제 5 조 제 24 호) 과 *precatório* 제도의 질 서 있는 지급 요건을 형량한 것이다.

감정평가사에 대한 시사점. 이 판결은 수용기관을 위하여 활동하는 감정평가사가 작성하는 최초 공탁금 평가에 직접적인 실무적 결과를 가져온다. 최초 공탁액이 과소평가되면 그 차액은 상당한 지연 이 따를 수 있는 *precatório* 제도를 통해 나중에 지급되므로, 원고 측에는 최초 평가가 과소하다고 다 틀 구조적 유인이 있다. 최초 공탁금 평가를 작성하는 감정평가사는 최초 공탁액과 최종 사법 판단 사이의 격차가 STF 판례와 Decreto-Lei 3.365/1941(법률명령 제 3.365/1941 호) 체계상의 보상이자 (*juros compensatórios*, 주루스 콤펙사토리우스) 부과 대상이 되리라는 점을 예상하면서, 이후의 사법 감정에 대하여 평가를 방어할 수 있도록 방법론을 각별히 주의하여 문서화하여야 한다.

Superior Tribunal de Justiça, REsp 2.129.162(반복 쟁점 Tema). 브라질 고등법원, 2025 년 5 월 판결.

STJ 는 수용기관이 *desapropriação* 소를 취하하는 경우의 변호사 보수에 관한 반복적 쟁점을 정리하 였다. 재판소는 취하 사건의 보수가 *Código de Processo Civil*(코디구 지 프로세수 시빌, 민사소송법) 의 일반 절차규정이 아니라 Decreto-Lei 3.365/1941 제 27 조 제 1 항의 체계를 따라야 한다고 확립 하였다. 이 판결은 배전선로를 위한 행정 지역권을 구하는 Cemig(Companhia Energética de Minas Gerais, 미나스제라이스 전력공사) 의 소에서 비롯되었다.

감정평가사에 대한 시사점. 이 판결은 방법론이 아니라 절차에 관한 것이지만, 행정 지역권 및 *desapropriação* 사안에서 감정평가사의 수임 구조에 시사점을 갖는다. 그러한 절차에서 감정인 (*peritos judiciais*, 페리투스 주디시아이스) 으로 선임되는 감정평가사는 보수 약정이 일반 절차규정 이 아니라 해당 법률명령 체계를 근거로 문서화되도록 하여야 한다.

Superior Tribunal de Justiça, *juros compensatórios* 에 관한 법리. 2025 년 계속 적용.

STJ 는 *desapropriação* 사안에서 보상이자 (*juros compensatórios*) 가 수용된 부동산의 권원에 관한 사법적 판단 이후에야 *incidem*(발생) 한다는 법리를 유지해 왔다. 이자는 최초 공탁액과 최종 보상액 의 차액에 대하여 역사적으로 확립된 연 12 퍼센트의 이율 (STF 의 *Súmula* 618, 수물라 제 618 호) 로 적용되며, 최종 보상이 지급되기 전에 점유를 미리 상실하는 데 대하여 수용된 소유자를 보상하는 기능을 한다.

감정평가사에 대한 시사점. 브라질 감정평가사는 사법 감정에서 자신이 확인하는 평가 격차 — 공공 기관의 최초 공탁액과 감정된 시장가치 사이의 차이 — 가 보상이자 산정의 기초가 된다는 점을 예상 하여야 한다. 따라서 그 격차의 경제적 의미는, 특히 절차가 장기화되는 경우, 명목상의 차이가 시사 하는 것보다 크다.

Chile(칠레)

Corte Suprema, Rol N° 31.423-2025. 칠레 대법원, 2025 년 9 월 판결.

칠레 대법원은 Decree-Law N° 2.186(법률명령 제 2.186 호) 제 12 조에 따른 특별 수용 절차에서 부동 산 개발업자 (*inmobiliaria*, 인모빌리아리아) 가 국고 (*Fisco de Chile*, 피스코 데 칠레) 를 상대로 제 기한 상고를 기각하였다. 원고는 전문가위원회 (*Comisión de Peritos*, 코미시온 데 페리토스) 가 정 한 평가액이 해당 부동산의 최유효이용 (*mayor y mejor uso*, 마요르 이 메호르 우소) 과 기존 건축물

및 부대공작물의 가치를 반영하지 못하였다고 주장하고, 나아가 장래 임대수익을 발생시켰을 기존 임대차계약을 근거로 일실이익 (*lucro cesante*, 루크로 세산테)에 대한 보상을 구하였다. 재판소는 원고의 주장이 평가를 규율하는 법규범에 대한 진정한 위반이 아니라 원심의 증거 판단에 대한 이견에 지나지 않으며, 전문가위원회의 증거를 서증 및 인증과 함께 원용한 원심의 논거가 충분히 근거를 갖추었다고 판시하였다.

감정평가사에 대한 시사점. 이 판결은 수용 절차에서 전문가위원회의 평가에 칠레 대법원 법리가 상당한 존중을 부여한다는 점과, 방법론적 이견만을 이유로 한 평가 결론에 대한 상고심 심사의 범위가 제한적이라는 점을 보여준다. 칠레 수용 사안에서 반대 평가를 작성하는 감정평가사는 대안적 평가 논거보다 위원회 작업의 입증 가능한 방법론적 오류에 초점을 맞추어야 하며, 기존 임대차 관계에 기초한 *lucro cesante* 주장은 상당한 증명 요건에 직면한다는 점을 인식하여야 한다.

Corte Suprema(Coquimbo(코킴보) 수용 사건). 칠레 대법원, 2025 년 7 월 17 일 판결.

칠레 대법원은 Coquimbo 수용 사안에서 보상액을 (사실 감정평가사의 보고서에 근거한) UF 2,969.75로 인정한 항소심 판단을, (*Conservador de Bienes Raíces* (콘세르바도르 데 비에네스 라이세스, 부동산등기소)의 등기된 매매에 근거한) 국가 측 주장 금액 UF 2,349.07에 반하여 인용하였다. 재판소는 항소심이 인증 및 인근 부동산에 관한 추가 감정보고서와 함께 고려하여 사실 감정평가사의 보고서에 의존한 것은 *sana crítica*(사나 크리티카, 합리적 증거 판단)의 허용된 행사이며, 상고심 심사는 그러한 증거 판단을 재차 다루는 데까지 미치지 않는다고 강조하였다.

감정평가사에 대한 시사점. Coquimbo 판결은 Rol N° 31.423-2025의 대응물이다. 개발업자가 다룰 때에는 위원회의 평가가 인정되지만, 적절한 방법론적 논거와 보강 증거로 뒷받침될 때에는 사실 감정평가사의 더 높은 평가도 인정될 수 있다. 칠레 감정평가사에게 주는 실무적 교훈은 사실 감정평가의 방법론과 문서상 뒷받침이 결정적이며, 등기된 매매 자료만으로는 추가 비교사례와 맥락 증거를 원용하여 잘 뒷받침된 사실 감정평가를 배척하기에 충분하지 않다는 것이다.

Tribunal Constitucional(헌법재판소), 2025 년 10 월. 조세법에 따른 *bienes raíces*(비에네스 라이세스, 부동산)의 *tasación*(타사시온, 평가).

칠레 헌법재판소는 *impuesto territorial*(임푸에스토 테리토리알, 토지세)의 목적상 부동산 (*bienes raíces*)의 *tasación*을 규율하는 *Código Tributario*(코디고 트리부타리오, 조세법)규정에 대한 위헌 주장을 배척하였다. 반대의견의 재판관들은 문제된 규범이 집행을 받는 당사자가 *tasación*에 이의를 제기할 수 있는 장치를 충분히 두지 않아 합리적이고 정당한 절차의 헌법적 보장을 훼손한다고 주장하였다.

감정평가사에 대한 시사점. 이 판결은 칠레 부동산 세제의 기존 행정적 *tasación* 체계를 유지시키지만, 반대의견의 논거는 향후 헌법적 또는 행정법적 다툼에서 다시 문제될 수 있는 현행 체계의 취약점을 지적한다. 부동산 세금 분쟁과 관련한 평가를 작성하는 칠레 감정평가사는 납세자의 이익제기 장치를 개선하기 위한 행정적 개혁이 이루어질 가능성이 있어 보이므로 판결 이후의 정책 환경을 주시하여야 한다.

Colombia(콜롬비아)

Corte Constitucional, Jurisdicción Agraria y Rural의 승인. 콜롬비아 헌법재판소, 2025년 결정.

콜롬비아 헌법재판소는 콜롬비아 사법부 내에 농촌 토지에 관한 전문 관할인 새로운 *Jurisdicción Agraria y Rural*(후리스딕시온 아그라리아 이 루랄, 농지·농촌 관할)을 창설하는 법률에 대한 위헌법률심사를 통과시켰다. 재판소는 개혁의 대상, 사법부로의 편입, 전문 법원 및 재판부의 설치, 예산에 관한 사항을 정한 핵심 조항들을 승인하였다. 새 관할은 2027년 업무 개시가 예정되어 있으며, 토지

원상회복 분쟁, 농지 재산 분쟁, 2016 년 평화협정 이행에서 비롯되는 사안을 해결하도록 설계되었다. 승인 시점 기준으로 61,000 헥타르가 넘는 면적에 관한 농지 분쟁이 기존 행정·사법 체계에 계속 중이다.

감정평가사에 대한 시사점. 새 관할은 콜롬비아에서 토지 원상회복 절차와 농촌 부동산 사안에서 비롯되는 평가 분쟁의 주된 재판 기관이 될 것이다. 원상회복 맥락에서 활동하는 콜롬비아 감정평가사 (Unidad de Restitución de Tierras(우니다드 데 레스티투시온 데 티에라스, 토지원상회복청) 와 Instituto Geográfico Agustín Codazzi(인스티투토 헤오그라피코 아구스틴 코다시) 를 위해 일하는 이들을 포함하여) 는 새로운 법원들이 자체 평가 판례를 형성할 것이고 방법론적 기준이 현재의 행정 주도 체계로부터 변화할 수 있음을 예상하여야 한다. 실무자들은 2027 년의 운영 개시와 새 법원들의 초기 사건 처리 양상을 주시하여야 한다.

토지 원상회복을 위한 기존 IGAC 기반 평가 체계. 2025 년 계속 적용.

콜롬비아의 토지 원상회복 실무는 원상회복과 대체 보상의 맥락에서 보상의 주된 근거로 Instituto Geográfico Agustín Codazzi(IGAC) 가 작성하는 *avalúo comercial* (아발루오 코메르시아, 상업적 감정평가) 에 계속 의존하고 있으며, 필요한 경우 Lonjas de Propiedad Raíz(론하스 데 프로피에다드 라이스, 부동산협회) 가 작성하는 *avalúo* 로 보완된다. 이 체계는 (이행의 절차적 요건에 관한) 헌법 재판소 2018 년 SU-034 결정과, (현물 보상 상황에서 IGAC 평가와 개별 수혜자의 선호 사이의 관계에 관한) 2024 년 T-120 결정을 비롯한 후속 결정들에서 정리되어 있다.

감정평가사에 대한 시사점. 콜롬비아 토지 원상회복 업무를 수행하는 감정평가사에게 이 제도적 체계는 IGAC 평가를 보상 산정의 실효적 중심에 두고, Lonja de Propiedad Raíz 의 *avalúo* 는 보완적 또는 분쟁 해결 수단으로 기능하게 한다. 앞으로의 Jurisdicción Agraria y Rural 은 이러한 제도적 구조를 바꿀 수 있으며, IGAC 평가와 Lonja 평가의 상대적 비중은 새 법원들이 발전시킬 쟁점 가운데 하나가 될 가능성이 높다.

Mexico(멕시코)

Suprema Corte de Justicia de la Nación, avalúo 와 가격의 차액에 대한 ISR 납세의무. 멕시코 대법원, 2026 년 초 공표된 결정.

멕시코 대법원 (SCJN) 전원부는 Ley del Impuesto sobre la Renta(레이 텔 임푸에스토 소브레 라 렌타, 소득세법, ISR) 제 125 조, 제 130 조 제 4 호, 제 132 조를 합헌으로 선언하였다. 이들 조항은 부동산 취득의 *avalúo*(아발루오, 감정평가액) 가 계약상 매수가격을 10 퍼센트 넘게 초과하는 경우 그 차액을 매수인의 합산 소득으로 보아 ISR 의 과세 대상으로 삼는다. 재판소는 시장가치보다 현저히 낮은 가격으로 재산을 취득하는 것은 매수인의 재산에 객관적 증가를 가져오므로 과세 대상 경제적 이익을 적절히 구성한다는 논거로, 이 규정이 조세의 비례성과 형평성이라는 헌법 원칙에 위반되지 않는다고 판시하였다.

감정평가사에 대한 시사점. 이 결정은 멕시코의 부동산 거래와 이를 뒷받침하는 감정평가 실무에 상당한 실무적 결과를 가져온다. *corredores públicos*(코레도레스 푸블리코스, 공인중개인) 또는 그 밖의 인가된 전문가가 수행하는 *avalúo* 는, 그 금액이 계약가격을 10 퍼센트 넘게 초과하는 경우 이제 매수인에게 직접적인 조세상 결과를 낳는다. 감정평가사는 거래 관련 *avalúo* 의 방법론에 대한 심사가 강화되고, 비교매매 근거를 각별히 엄밀하게 문서화한 *avalúo* 에 대한 수요가 늘어날 것을 예상하여야 한다. 이 결정은 또한 방법론적 선택이 이제 매수인의 납세의무에 영향을 줄 수 있으므로, 대안적 평가 방식 가운데 무엇을 택할 것인지의 중요성을 암묵적으로 높인다.

Suprema Corte de Justicia de la Nación, Amparo en revisión 391/2024. 멕시코 대법원 제 2 부, 2025 년 6 월 25 일 판결.

멕시코 대법원 제 2 부는 광구 (鑛區) 및 관련 부수적 권리 (임시 점용, 지역권, 수리권 우선권, 임대차 갱신 조항) 의 수용에 관한 법적 틀에 영향을 미치는 명령 (Decreto) 에 대하여 한 광업회사가 제기한 위헌 주장을 판단하였다. 재판소는 문제된 규정들이 취득한 권리가 아니라 장래 권리에 대한 단순한 기대에 관한 것이므로 소급효 금지 원칙에 위반되지 않는다고 판시하였다.

감정평가사에 대한 시사점. 이 판결은 멕시코의 광업권 및 광산 부동산 평가에 중요하다. 광구 연장권과 관련 규정을 취득한 권리가 아니라 “단순한 기대” 로 성격 규정한 재판소의 태도는 광업 특허권의 평가와 그 평가에서의 규제 위험 산정에 영향을 미친다. 멕시코에서 광산 부동산이나 관련 권리의 평가를 작성하는 감정평가사는 평가의 규제 위험 차원을 명시적으로 다루어야 하며, 조건부이거나 연장 여부에 좌우되는 가치를 완전히 확정된 것처럼 취급하는 일을 피하여야 한다.