

Valuation Engineer Journal

La ingeniería del valor: medida, disciplinada, reproducible.

Volumen 1 · Número 1 · Verano de 2026



EN ESTE NÚMERO

La brecha entre doctrina y práctica en la valoración inmobiliaria
Residual Constraint Approach (RCA): marco y protocolo
earthUI, glmnetUI & mgcvUI: interfaces abiertas de modelización

FUNDAMENTOS (F0–F6) Bien heterogéneo · Espacio de características ·
Función de precios hedónicos · Precio implícito · Regresión hedónica · Variable latente

RESOLUCIONES JUDICIALES RECIENTES Estados Unidos · Unión Europea · América Latina

Valuation Engineer Journal

Volumen 1 · Número 1
Verano de 2026
ISSN 3143-5033

Editor jefe William Bert Craytor, SRA

Índice

Misión	2
Editorial	3
Bienvenida al número inaugural	4
Cartas al editor	6
Una invitación a la correspondencia	7
Artículos	9
La brecha entre doctrina y práctica en la valoración inmobiliaria . .	10
Residual Constraint Approach (RCA): marco y protocolo	42
earthUI: una interfaz interactiva para el paquete earth (MARS) . .	103
glmnetUI: una interfaz interactiva para el paquete glmnet (lasso y elastic net)	150
mgcvUI: una interfaz interactiva para el paquete mgcv (modelos aditivos generalizados)	188
Fundamentos	229
Panorama general (F0)	230
Bien heterogéneo (F1)	236
Espacio de características (F2)	244
Función de precios hedónicos (F3)	251
Precio implícito (F4)	257
Regresión hedónica (F5)	264
Variable latente (F6)	272
Resoluciones judiciales recientes	279
Resoluciones judiciales recientes – Estados Unidos (R1)	280
Resoluciones judiciales recientes – Unión Europea (R2)	283
Resoluciones judiciales recientes – América Latina (R3)	289

Misión

La *Valuation Engineer Journal* publica investigación y estudios de caso rigurosos y orientados a la práctica sobre la ingeniería del valor: métodos, modelos y mediciones que conectan las decisiones técnicas con los resultados económicos. Damos prioridad a la reproducibilidad, a la transparencia de los supuestos y a la claridad de la exposición.

Editorial

SECCIÓN: EDITORIAL

Bienvenidos al número inaugural

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor, SRA — Editor jefe

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

DOI: 10.67330/0tymfa32

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

Este artículo de acceso abierto se distribuye bajo los términos de la licencia Creative Commons Atribución 4.0.

Resumen. El Mensaje del Editor para el Volumen 1, Número 1: por qué existe esta revista, qué aportan los cinco artículos y las tres secciones, y una invitación a participar.

Palabras clave: editorial, ingeniería de valuación, número inaugural

Bienvenidos al número inaugural de *The Valuation Engineer Journal*.

Esta revista existe a causa de una brecha. La doctrina de la tasación lleva mucho tiempo describiendo la valuación como la medición disciplinada del comportamiento del mercado y, sin embargo, la práctica cotidiana sigue apoyándose en gran medida en la selección manual de comparables y en ajustes estimados por criterio profesional únicamente. Entretanto, las herramientas estadísticas capaces de medir lo que los mercados pagan realmente por las características de un inmueble — la regresión hedónica, los modelos modernos de splines, el aprendizaje automático con validación cruzada — han madurado a lo largo de décadas con escasa adopción en la tasación en ejercicio. *Ingeniero de valuación* es el nombre que damos al profesional que cierra esa brecha: competencia acreditada en tasación unida a ingeniería de datos, estadística y protocolo reproducible.

Los cinco artículos de este número sientan la base de trabajo. El artículo inicial examina la propia brecha entre doctrina y práctica y sostiene que las normas de la profesión ya exigen más medición de la que ofrecen sus métodos cotidianos. En segundo lugar — y central para todo lo que esta revista se propone — se presenta en su totalidad el Residual Constraint Approach (Enfoque de Restricción Residual): un protocolo que permite al canal de regresión medir lo que puede medirse, disciplina lo que no puede medirse en una descomposición residual explícita y restringe el total de modo que el criterio profesional pueda refinar el valor, pero nunca inventarlo. Siguen tres artículos de software, que documentan earthUI y sus interfaces afines

para regresión penalizada y modelos aditivos generalizados — el instrumental abierto sobre el que se construye esta revista —, consolidados en una sola aplicación poco antes del cierre de edición, tal como describen las notas del editor de cada artículo.

Tres secciones acompañan a los artículos. *Foundations* (Fundamentos) se abre con siete entradas breves (F0–F6) que trazan la cadena conceptual desde los bienes heterogéneos, pasando por el espacio de características, la función de precios hedónicos, los precios implícitos y la regresión hedónica, hasta las variables latentes que habitan en los residuos. Cada entrada cita su linaje académico, lleva su propio DOI y termina con una formalización en Prolog legible por máquina — el comienzo de una base de conocimiento que pretendemos ampliar, número a número, hasta convertirla en un vocabulario compartido y preciso para el campo. *Recent Court Decisions* (Decisiones judiciales recientes) reseña resoluciones de los Estados Unidos, la Unión Europea y América Latina que inciden en la práctica valuatoria, porque el entorno jurídico forma parte del problema de ingeniería. *Letters to the Editor* (Cartas al director) inaugura la columna de correspondencia: los métodos concebidos para usarse se ponen a prueba de maneras que sus autores nunca previeron, y es aquí donde esos hallazgos — y los desacuerdos que suscitan — tienen su lugar.

Todos los contenidos de este número se publican en acceso abierto bajo licencia CC BY 4.0, tanto en PDF como en HTML, con un DOI de Crossref, y se preservan de forma independiente en Zenodo. Invitamos a enviar contribuciones — artículos, entradas de *Foundations*, correspondencia y discrepancias — a tasadores, economistas e ingenieros por igual.

William Bert Craytor, SRA

Editor jefe

Cartas al editor

SECCIÓN: CARTAS AL DIRECTOR

Una invitación a la correspondencia

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor, SRA — Editor jefe

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

DOI: 10.67330/r55mw896

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

Este artículo de acceso abierto se distribuye bajo los términos de la licencia Creative Commons Atribución 4.0.

Resumen. El editor inaugura la sección de cartas: qué espera recibir esta revista de sus lectores y cómo enviarlo.

Palabras clave: cartas, correspondencia, comentario entre pares

Una revista que solo habla es media revista. Esta sección existe para la otra mitad.

The Valuation Engineer Journal publica métodos pensados para ser utilizados, y los métodos que se utilizan se ponen a prueba de maneras que sus autores nunca previeron. Si usted ha aplicado el Residual Constraint Approach (Enfoque de Restricción Residual) y ha encontrado un paso que falla en su mercado, queremos esa carta. Si una entrada de *Foundations* define un término de forma más restringida de lo que su práctica exige, queremos esa carta. Si una resolución judicial que hemos resumido se lee de otro modo para un profesional de esa jurisdicción, también queremos esa carta. La discrepancia, la replicación, el refinamiento y la simple corrección entran todos dentro del alcance; también los informes breves surgidos de la práctica — una técnica que funcionó, un problema de datos resuelto, un fracaso del que conviene advertir a los demás.

Las reglas básicas son pocas. Las cartas deben ir firmadas, ser concisas (una página es lo ideal) y ser específicas: cite el artículo, el apartado o la entrada a la que se refiere y aporte pruebas del mismo modo en que lo haría en un expediente de trabajo. Podemos editarlas por extensión y claridad, siempre con la aprobación del autor. Las cartas se publican en acceso abierto bajo licencia CC BY 4.0, como todo lo demás en estas páginas, y a los autores cuyo trabajo se comente se les ofrecerá la posibilidad de responder en el mismo número.

Envíe las cartas a través del sitio de la revista (journal.valuation-engineer.com) o directamente al editor. Las primeras cartas reales que reciba esta revista

marcarán el momento en que se convierta en una conversación y no solo en una publicación: las esperamos con interés.

William Bert Craytor, SRA
Editor jefe

Artículos

La brecha entre doctrina y práctica en la tasación inmobiliaria: un análisis estructurado de funciones, límites y tensiones

Traducción al español del artículo publicado originalmente en inglés en este número. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor

Volumen 1, Número 1 · 21 de mayo de 2026

Publicado el 1 de julio de 2026

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

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

Este artículo de acceso abierto se distribuye bajo los términos de la [licencia Creative Commons Atribución 4.0](#).

Resumen. La tasación inmobiliaria ocupa una posición de considerable interés público; sin embargo, su doctrina oficial —centrada en opiniones de valor de mercado bajo condiciones pasadas y actuales— se encuentra en creciente tensión con la manera en que los productos de la tasación se utilizan realmente en los circuitos financieros contemporáneos. Este artículo construye ese argumento en cinco etapas, cada una acompañada de una base de conocimiento explícita y consultable en Prolog: las dos primeras establecen las funciones sociales y económicas que cumple la tasación; la tercera identifica su producto final principal, una opinión de valor de mercado; y la cuarta examina la orientación temporal y metodológica de la tasación frente al pronóstico cuantitativo y financiero, junto con el límite regulatorio que los separa. Todas convergen en la afirmación central de la quinta: en el crédito hipotecario, el valor de mercado escalado por una relación préstamo-valor (LTV) funciona como un proxy implícito del valor futuro de ejecución hipotecaria —un uso orientado al futuro que la doctrina del valor presente no reconoce—. Muchos tasadores, al desconocer este papel de valuación de la garantía, no tratan el valor de mercado con la cautela hacia las tendencias futuras del valor que este amerita; una cuestión que el presente análisis busca poner de manifiesto y abrir a trabajos posteriores.

Palabras clave: tasación inmobiliaria, valor de mercado, doctrina de valuación, representación del conocimiento, normas de tasación, valor de ejecución hipotecaria

Índice

1. Importancia social	2
2. Importancia económica	6
3. La importancia de las estimaciones de valor de mercado como producto final de las tasaciones	10
4. La orientación temporal de la actividad de tasación	14
5. El valor de mercado $\times$ LTV como proxy del valor de ejecución hipotecaria	21
6. Conclusión	27

1 Importancia social

Las tasaciones de bienes inmuebles y de otros activos cumplen una función social y económica esencial al ayudar a compradores, vendedores, prestamistas, aseguradoras, inversionistas, tribunales, autoridades fiscales y al público en general a tomar decisiones informadas y equitativas sobre la transferencia, el financiamiento, la tributación y la administración de la propiedad y de otras formas de riqueza. Las tasaciones de alta calidad contribuyen a la transparencia del mercado, la confianza pública, la asignación eficiente del capital y la estabilidad de las instituciones financieras, al proporcionar opiniones de valor independientes y bien fundamentadas, basadas en la evidencia, el análisis y el juicio profesional. (LaCour-Little y Malpezzi 2003; K'Akumu y Larsen 2024; Hendershott y Kane 1995; Lee, Peng y Liao 2020)

Las tasaciones de baja calidad, en cambio, pueden producir consecuencias graves en múltiples horizontes temporales. En el corto plazo, las tasaciones inexactas pueden causar a compradores o vendedores individuales un daño financiero inmediato por sobrepago, venta por debajo del valor, tributación excesiva o decisiones de crédito inapropiadas. En el mediano plazo, las deficiencias sistemáticas de tasación pueden distorsionar las señales del mercado, deteriorar la suscripción de crédito, fomentar el comportamiento especulativo, contribuir a una tributación o cobertura de seguros inequitativas y aumentar los litigios y la desconfianza transaccional. En el largo plazo, las fallas persistentes en la calidad de las tasaciones pueden socavar la confianza en los mercados y las instituciones, contribuir a la inestabilidad financiera, la asignación deficiente de recursos, el declive o la sobreinversión en los vecindarios, y a inequidades sociales más amplias que afectan a hogares, empresas y comunidades. (LaCour-Little y Malpezzi 2003; Mayer y Nothaft 2022)

Dado que los bienes inmuebles y otros activos representan a menudo una porción sustancial de la riqueza personal, corporativa y pública, la calidad de la tasación no es meramente una cuestión técnica privada, sino un asunto de considerable interés público ligado a la resiliencia económica, la equidad y el bien público a largo plazo. (Mayer y 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
    , social_effect/4
    , social_contribution/1
    , social_hallmark/1
    , social_equity_principle/1
    ],
    % DecisionType, Party
    % Quality, Horizon, Domain, Effect
```

```

        , 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
    ]).

:- discontinuous social_effect/4.
:- discontinuous 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 Importancia económica

Las tasaciones de bienes inmuebles y de otros activos desempeñan un papel económico fundamental al apoyar el funcionamiento eficiente de los mercados de crédito, las decisiones de inversión, los sistemas tributarios, la información financiera, la suscripción de seguros, la administración de sucesiones, las transacciones corporativas y la planificación de políticas públicas. Las tasaciones confiables ayudan a establecer medidas creíbles del valor de la garantía y del valor de mercado que permiten que el capital fluya con mayor eficiencia entre prestatarios, prestamistas, inversionistas, gobiernos y empresas. (LaCour-Little y Malpezzi 2003; Ebrahim y Hussain 2010)

En el corto plazo, las tasaciones precisas reducen la incertidumbre transaccional, mejoran las decisiones de crédito, facilitan negociaciones justas y ayudan a prevenir pérdidas financieras inmediatas derivadas de la sobrevaluación o la subvaluación. En el mediano plazo, la calidad de la tasación influye en la estabilidad y la eficiencia de los mercados en general, al afectar el riesgo hipotecario, la asignación de inversiones, la equidad tributaria, la suficiencia de los seguros y el comportamiento de fijación de precios de los participantes del mercado. Las tasaciones débiles o sistemáticamente sesgadas durante este período pueden contribuir a burbujas de activos infladas, carteras de préstamos deterioradas, incentivos de desarrollo distorsionados y un mayor riesgo de incumplimiento o de litigio. (LaCour-Little y Malpezzi 2003; Hendershott y Kane 1992; Griffin y Maturana 2016)

En el largo plazo, las consecuencias económicas acumuladas de la calidad de la tasación se vuelven aún más significativas, al influir en los patrones de formación de capital, la inversión en infraestructura, la transferencia intergeneracional de riqueza, los ingresos públicos, la asequibilidad de la vivienda y la resiliencia general del sistema financiero. Las deficiencias sostenidas en las prácticas de valuación pueden contribuir a una asignación crónicamente deficiente del capital, a la inestabilidad bancaria, a una menor confianza de los inversionistas y a un uso ineficiente de la tierra y de los activos productivos, mientras que unas normas de tasación sólidas ayudan a sostener mercados estables, una planificación económica sana y el crecimiento económico a largo plazo. (Ghent, Torous y Valkanov 2019; Hendershott y Kane 1992; Ebrahim y Hussain 2010)

```

% appraisal_economics.pl

```

```

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

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

:- 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 La importancia de las estimaciones de valor de mercado como producto final de las tasaciones

Las tasaciones de valor de mercado desempeñan un papel importante en el análisis del riesgo de la garantía porque proporcionan una estimación elaborada profesionalmente del precio probable que una propiedad o un activo podría alcanzar en condiciones normales de mercado, sirviendo así como referencia fundamental para prestamistas, inversionistas y analistas de garantías que evalúan la exposición crediticia y los posibles resultados de una ejecución hipotecaria. (Agarwal, Ben-David y Yao 2015) Aunque el valor de ejecución hipotecaria o de liquidación puede diferir sustancialmente del valor de mercado debido a condiciones de comercialización en situación de dificultad, períodos de venta acelerados, costos legales, deterioro físico o un sentimiento de mercado adverso, los analistas de garantías recurren con frecuencia a las tasaciones de valor de mercado como punto de partida para estimar el potencial de recuperación en escenarios de estrés. (Campbell, Giglio y Pathak 2011; Pennington-Cross 2006; Clauretie y Daneshvary 2009)

Al analizar la relación entre el valor de mercado, los saldos de los préstamos, las relaciones préstamo-valor, las tendencias del mercado, el estado de la propiedad, la liquidez y las condiciones locales de oferta y demanda, los analistas de garantías evalúan la probabilidad y la severidad de las pérdidas en caso de incumplimiento del prestatario. (Qi y Yang 2009)

En el corto plazo, las tasaciones precisas de valor de mercado ayudan a los prestamistas a

tomar decisiones de suscripción prudentes y a establecer márgenes de garantía apropiados. En el mediano plazo, apoyan la gestión del riesgo de cartera, la titulización de préstamos, la asignación de reservas y el cumplimiento regulatorio, al ayudar a las instituciones a evaluar la suficiencia cambiante de la garantía a lo largo de los ciclos económicos. En el largo plazo, unas prácticas de tasación consistentemente confiables contribuyen a la estabilidad del sistema financiero al reducir el riesgo de apalancamiento excesivo, de préstamos mal garantizados y de una fijación sistémicamente errónea de los precios de los activos en garantía, capaz de amplificar las pérdidas por ejecución hipotecaria y perturbaciones económicas más amplias durante períodos de tensión en el mercado.(Eriksen et al. 2019)

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

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

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

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

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

```

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

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

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

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

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

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

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

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

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

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

risk_input(market_value).
risk_input(loan_balance).
risk_input(loan_to_value_ratio).

```

```

risk_input(market_trends).
risk_input(property_condition).
risk_input(liquidity).
risk_input(local_supply_demand).

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

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

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

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

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

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

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

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

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

```

```

collateral_effect(low, long_term, system, systemic_collateral_mispricing).
collateral_effect(low, long_term, system, amplified_foreclosure_losses).
collateral_effect(low, long_term, system,
↳ amplified_economic_disruption_under_stress).

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

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

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

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

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

```

4 La orientación temporal de la actividad de tasación

Los tasadores generalmente desarrollan opiniones de valor de mercado en un punto del tiempo, a una fecha efectiva específica —retrospectiva (pasada), actual o, con limitaciones definidas, prospectiva (futura)—, en lugar de pronósticos probabilísticos sobre períodos prolongados. Esta orientación temporal y metodológica representa una de las distinciones centrales entre la práctica regulada de la tasación y el trabajo de los analistas cuantitativos, los analistas financieros, los analistas de inversiones y los economistas. (Quan y Quigley 1991)

En la mayoría de los encargos de tasación, en particular los que involucran valor de mercado bajo normas como los Uniform Standards of Professional Appraisal Practice (USPAP, normas uniformes de práctica profesional de tasación) o las International Valuation Standards (IVS, normas internacionales de valuación), el tasador produce una estimación puntual conciliada (o, en los casos permitidos, un rango estrecho) anclada en evidencia de mercado observable, condiciones imperantes y metodologías reconocidas a una única fecha efectiva. Los USPAP permiten explícitamente conclusiones de valor expresadas como una cantidad específica, un rango de cifras o una relación con un valor de referencia; aun así, la opinión conciliada sigue siendo una indicación puntual creíble ligada a esa fecha. (The Appraisal Foundation 2024; International Valuation Standards Council 2025)

Aunque los tasadores incorporan necesariamente las expectativas razonables de los participantes del mercado respecto de tendencias, oferta y demanda, depreciación, tasas de capitalización o de descuento y absorción, lo hacen para sustentar un valor respaldado por el mercado a la fecha efectiva, no para pronosticar movimientos especulativos. Las opiniones de valor prospectivo (por ejemplo, el valor estabilizado al término de un proyecto) requieren supuestos extraordinarios o condiciones hipotéticas explícitos y permanecen restringidas a una fecha efectiva futura definida, con respaldo derivado del mercado.

En contraste, los analistas cuantitativos, financieros y de inversiones abordan con frecuencia la incertidumbre mediante estimaciones que cubren un rango de resultados posibles a lo largo de períodos, empleando distribuciones de probabilidad, simulaciones de Monte Carlo, análisis de

escenarios, pruebas de sensibilidad o métricas de valor en riesgo. Su enfoque suele centrarse en el desempeño futuro esperado, los rendimientos ajustados por riesgo o el valor de inversión para una parte específica en horizontes variables, en lugar de un único punto de valor de mercado a una fecha.(Hull 2018)

Ese trabajo también presupone competencias técnicas que una licencia de tasador ni exige ni certifica. Ajustar un modelo de multivariate adaptive regression splines (MARS, regresión adaptativa multivariante por splines), por ejemplo, requiere el juicio necesario para controlar la complejidad del modelo de modo que un fuerte ajuste dentro de la muestra (un R^2 alto) no se compre al precio del sobreajuste, sino que se traslade a los datos reservados (un R^2 alto en validación cruzada), lo cual a su vez presupone una comprensión funcional de los algoritmos subyacentes.(Friedman 1991; Steurer, Hill y Pfeifer 2021) Junto a esa facilidad cuantitativa, el trabajo premia la capacidad conceptual de manejar conceptos altamente estructurados, de discernir distinciones finas de significado y de asimilar métodos desconocidos. Se trata de habilidades y no de credenciales: un tasador con licencia puede poseerlas y un analista sin licencia puede carecer de ellas, que es exactamente la razón por la cual, por sí solas, no determinan de qué lado de la línea regulatoria cae una tarea determinada.

Los tasadores también pueden abordar elementos limitados orientados al futuro, como proyectar flujos de ingresos, vacancia, gastos y valores de reversión en análisis de flujo de caja descontado, o estimar valores prospectivos en contextos de desarrollo o de litigio. Estos componentes permanecen dentro de los límites de la tasación cuando se fundamentan en evidencia de mercado actual y en normas profesionales que enfatizan la fundamentación y la evitación de la especulación sin respaldo.(Diaz y Wolverton 1998; Gallimore 1996; Crosby, Lizieri y McAllister 2010) La práctica moderna de la tasación integra cada vez más herramientas estadísticas, consideraciones de escenarios y analítica de riesgos —particularmente en la tasación masiva, los modelos de valuación automatizada o el trabajo de cartera—, pero el producto central sigue siendo, por lo general, una opinión de valor de mercado en un punto del tiempo.(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 y Neal 2023)

Podría caracterizarse la práctica de la tasación como primordialmente probatoria y orientada a estimaciones puntuales del pasado al presente (o prospectivas limitadas), mientras que el análisis financiero y cuantitativo es más probabilístico, distribucional y explícitamente orientado al futuro a través de horizontes temporales.(Geltner 1991) Los tasadores preguntan: «¿Cuál es la indicación de valor de mercado más fundamentada a la fecha efectiva especificada, dada la evidencia actual y los supuestos definidos?» Los analistas financieros y cuantitativos preguntan con mayor frecuencia: «¿Qué rango de resultados es probable en un período futuro, con qué probabilidades, y cómo deberían asignarse los recursos?»

El límite no es absoluto y continúa evolucionando con la complejidad del mercado. No obstante, persisten diferencias fundamentales en el propósito (valor de mercado frente a valor de inversión), los requisitos probatorios y los marcos regulatorios.

La línea operativa, sin embargo, no la traza la habilidad sino la ley: la pregunta definitoria es simplemente si una tarea determinada es una de aquellas para las que la licencia de tasador se exige legalmente. Los analistas cuantitativos, los analistas financieros, los economistas, los analistas de inversiones, los gestores de cartera y los profesionales afines pueden legalmente estimar, pronosticar, modelar o analizar valores futuros de activos —incluidas proyecciones distribucionales o por rangos— sin una licencia de tasador inmobiliario, siempre que el trabajo no constituya una tasación regulada.(Murphy 2012) Las distinciones clave descansan en:

1. el tipo de valor (por ejemplo, una estimación puntual de valor de mercado frente a un valor de inversión o un pronóstico probabilístico);
2. el enfoque temporal (una fecha efectiva específica frente a un período u horizonte temporal);

3. la forma del producto (un punto o un rango estrecho frente a una distribución de probabilidad completa);
4. el propósito previsto y si el trabajo se presenta como una tasación; y
5. la participación de contextos regulados de crédito o contextos legales que exijan el cumplimiento de normas como los USPAP.

Este marco se originó en gran medida en regulaciones bancarias como FIRREA (1989) en los Estados Unidos, que buscó garantizar opiniones de valor de mercado creíbles e independientes para las transacciones relacionadas con el ámbito federal. (United States Congress 1989; K'Akumu y Larsen 2024) Muchas actividades de valuación sofisticadas —precios objetivo en análisis de acciones, estudios de factibilidad, modelización del riesgo de cartera, valuación de opciones o analítica predictiva— quedan fuera de la exigencia de licencia porque sirven a propósitos de inversión, pronóstico o asesoría, y no a funciones reguladas de tasación.

```

%% 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_appraisal).

%% --- 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 El valor de mercado $\times$ LTV como proxy del valor de ejecución hipotecaria

Las secciones precedentes expusieron para qué sirve la tasación, qué produce y dónde se encuentran sus límites doctrinales y regulatorios; esta sección llega a la tensión hacia la cual venían construyendo.

Una de las realidades prácticas más profundas que subyacen al crédito hipotecario y al análisis de garantías es que, en su función económica real, una tasación de valor de mercado utilizada en la suscripción hipotecaria a menudo sirve como proxy implícito de la recuperabilidad esperada de la garantía a lo largo de algún período futuro de tenencia, aun cuando formalmente la tasación se enmarque como una opinión de valor de mercado presente a una fecha efectiva específica. (Foster y Van Order 1984; Kau et al. 1995; Deng, Quigley y Van Order 2000)

Una relación préstamo-valor del 80 % no es arbitraria. Históricamente, evolucionó en parte como un colchón de riesgo destinado a proteger a los prestamistas frente a:

- descensos moderados del mercado,
- gastos de ejecución hipotecaria,
- descuentos de liquidación,
- costos de mantenimiento,
- gastos legales,
- intereses devengados,
- y la incertidumbre durante el tiempo necesario para disponer de la garantía tras el incumplimiento. (Foote, Gerardi y Willen 2008; Qi y Yang 2009)

El colchón es también la razón por la cual el propio monto del préstamo —el producto del valor de mercado y la relación préstamo-valor permitida— puede leerse como la estimación de trabajo que hace el prestamista del valor recuperable de ejecución hipotecaria: limitar el préstamo a una fracción del valor de mercado es, en efecto, una apuesta a que el producto de una venta forzosa no caerá por debajo del saldo del préstamo. Cuán grande *debería* ser el recorte, sin embargo, es mucho más difícil de precisar de lo que sugiere la cifra redonda del 80 %. Las estimaciones del descuento por ejecución hipotecaria varían ampliamente entre mercados, métodos y períodos; un estudio de ventas forzosas citado con frecuencia lo sitúa cerca de un 27 % por debajo de los precios comparables entre partes independientes (Campbell, Giglio y Pathak 2011), mientras que otros análisis —que corrigen por estado de la propiedad, ubicación y selección— reportan magnitudes sustancialmente distintas. (Pennington-Cross 2006; Clauretie y Daneshvary 2009)

El umbral convencional del 80 %, en cambio, es esencialmente un único parámetro de alcance nacional, aplicado a mercados locales enormemente heterogéneos y a fases sucesivas del ciclo crediticio. No puede seguir el descuento de ningún mercado en particular, por lo que funciona

como una aproximación muy burda y de talla única, en lugar de una correspondencia precisa entre el valor de mercado y la recuperación en ejecución hipotecaria. En todo caso, es un colchón delgado: un recorte del 20 % queda *por debajo* del descuento del 27 % que reporta ese estudio, de modo que el préstamo convencional máximo puede exceder el producto bruto de la venta forzosa incluso antes de restar los costos de disposición enumerados arriba.

Cuando el capital propio del prestatario es demasiado escaso para constituir el colchón, se lo sustituye por un seguro hipotecario. Esto se entiende mejor no como una palanca de política, sino como un intercambio en la asignación de riesgos: un segmento de prestatarios dispone de pocos activos pero tiene ingresos estables, y el riesgo incremental de incumplimiento y de pérdida que implica prestarles por encima del 80 % de LTV es pronosticado y asumido por un asegurador —de forma privada, mediante el seguro hipotecario privado en un préstamo convencional de LTV alto, o pública, a través de la Federal Housing Administration (administración federal de vivienda)—. El mercado en el que ese intercambio es rutinario, sin embargo, es una creación de la política federal de vivienda: las government-sponsored enterprises (empresas patrocinadas por el gobierno) que dominan el crédito convencional tienen prohibido por sus estatutos comprar préstamos por encima del 80 % de LTV sin dicha mejora crediticia (United States Code 1970). Los términos en que el seguro termina son igualmente decisiones institucionales y no cálculos de riesgo, y difieren marcadamente entre programas.¹

De modo que, aunque el tasador está oficialmente estimando el «valor de mercado actual», el prestamista interpreta económicamente ese valor a través de una lente de riesgo orientada al futuro. En efecto, el prestamista se pregunta algo más cercano a:

«Si este prestatario incumple en algún momento del futuro previsible, ¿es probable que la garantía siga siendo suficiente para cubrir la deuda pendiente y las pérdidas relacionadas con la ejecución hipotecaria?»

Esa es, inherentemente, una pregunta orientada al futuro.

Esto descansa en una separación conceptual deliberada entre:

- la opinión de «valor de mercado actual» del tasador, y
- el uso práctico que el prestamista hace de esa opinión como colchón de pronóstico frente al deterioro futuro de la garantía.

Histórica y legalmente, los reguladores intentaron impedir que los tasadores se convirtieran en pronosticadores especulativos porque:

- el pronóstico especulativo contribuyó a crisis crediticias pasadas,
- los tribunales querían tasaciones ligadas a evidencia observable,
- y los reguladores bancarios buscaban prácticas de valuación estandarizadas y defendibles.

Así, la doctrina de la tasación evolucionó hacia:

- el valor presente,
- la evidencia de mercado actual,
- fechas efectivas definidas,
- y supuestos acotados.

Mientras tanto, prestamistas, inversionistas y analistas de garantías utilizan rutinariamente esas mismas tasaciones en marcos probabilísticos de riesgo futuro.

¹El seguro hipotecario privado convencional pagado por el prestatario termina automáticamente al 78 % de LTV y su cancelación puede solicitarse al 80 % conforme a la Homeowners Protection Act de 1998. (United States Congress 1998) Las primas de la Federal Housing Administration son más estrictas: para los préstamos originados el 3 de junio de 2013 o después con una relación préstamo-valor original superior al 90 % (por ejemplo, 3.5 % de pago inicial), la prima anual se mantiene durante toda la vida del préstamo, mientras que en los préstamos con un LTV igual o inferior al 90 % se cancela después de once años. (U.S. Department of Housing and Urban Development 2013) Por ello, muchos prestatarios de la FHA solo se liberan del seguro hipotecario refinanciando hacia un préstamo convencional cercano al 80 % de LTV.

En realidad, la suscripción hipotecaria es un sistema integrado que involucra:

1. la valuación de mercado actual,
2. la estimación del riesgo futuro,
3. la probabilidad de incumplimiento,
4. el momento de la ejecución hipotecaria,
5. la severidad de la liquidación,
6. las expectativas macroeconómicas,
7. y el análisis de estrés de cartera.

El tasador contribuye formalmente sobre todo al punto (1), mientras que el prestamista combina internamente ese resultado con los demás.

La distinción se vuelve borrosa en la práctica. Si un prestamista no otorgara un préstamo a menos que creyera que la garantía probablemente conservará suficiente valor en el futuro, entonces la tasación funciona indirectamente como parte de un modelo de protección del valor futuro, aun cuando las normas de tasación eviten formalmente describirla de esa manera.

Esta tensión se ha vuelto aún más pronunciada con:

- los AVM (modelos de valuación automatizada),
- la analítica institucional de carteras,
- los modelos de titulización,
- los sistemas de valuación con aprendizaje automático,
- la modelización de capital de Basilea,
- y las pruebas de estrés hipotecario a gran escala.

La analítica moderna de garantías trata cada vez más los valores de las propiedades de forma probabilística a lo largo del tiempo, en lugar de como estimaciones estáticas de un solo punto. (Basel Committee on Banking Supervision 2017; Steurer, Hill y Pfeifer 2021; Board of Governors of the Federal Reserve System 2021)

Podría argumentarse que la regulación tradicional de la tasación refleja un compromiso histórico:

- los tasadores proporcionan anclas de valor presente disciplinadas y basadas en evidencia,
- mientras que las instituciones financieras asumen la responsabilidad de la interpretación del riesgo futuro.

Si esa separación sigue siendo plenamente coherente en los sistemas financieros modernos es una pregunta legítima y cada vez más debatida.

De ello se sigue una implicación práctica para la propia práctica de la tasación. La mayoría de los tasadores no son conscientes de este papel de valuación de la garantía —el uso que el prestamista hace del valor de mercado como colchón orientado al futuro rara vez es visible desde dentro del encargo de tasación— y, por lo tanto, no tratan sus conclusiones de valor de mercado con la cautela hacia las tendencias futuras del valor que ese papel amerita. Reconocer cómo se consume realmente la opinión de valor presente es un primer paso para cerrar esa brecha, y un área que este análisis pretende abrir a trabajos posteriores.

```
%% 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 Conclusión

Los cinco pasajes codificados en las bases de conocimiento precedentes describen una única profesión desde cinco puntos de vista complementarios. Tomado de forma aislada, cada pasaje se lee como un relato autónomo: lo que la profesión pone en juego en lo social y en el interés

público, las funciones económicas que las tasaciones apoyan, su papel operativo en el análisis del riesgo de la garantía, el límite regulatorio que las separa del trabajo cuantitativo adyacente, y la tensión entre su marco formal y su uso práctico. Tomados en conjunto, y con la estructura hecha explícita en Prolog, describen algo más interesante: una profesión cuya doctrina oficial y cuya función económica se han mantenido deliberadamente separadas, y cuya coherencia interna se ve cada vez más puesta a prueba por las herramientas analíticas que ahora se aplican a sus productos.

Las tres primeras bases de conocimiento comparten un esquema común: **effect(Quality, Horizon, Domain, Effect)**. Cada una enmarca la calidad de la tasación como un insumo causal cuyas consecuencias se propagan a través de horizontes de corto, mediano y largo plazo, difiriendo solo en el dominio a través del cual se leen esas consecuencias. La lente económica las lee a través de los mercados de crédito, la asignación de inversiones, la tributación y la formación de capital; la lente social las lee a través de la equidad, la confianza pública y los resultados comunitarios; la lente de la garantía las lee a través de la suscripción, la gestión de carteras y la estabilidad del sistema financiero. El esquema compartido no es una coincidencia. Refleja una afirmación sustantiva que recorre los tres pasajes: la calidad del trabajo de valuación tiene consecuencias, esas consecuencias escalan con el tiempo, y los mismos modos de falla subyacentes se manifiestan de manera diferente según qué intereses se pongan en primer plano. Los átomos compartidos en el lado de los argumentos de los predicados —en particular los vocabularios de calidad y de horizonte— hacen que los tres vocabularios sean combinables, de modo que una única falla, como la sobrevaluación sistemática, pueda rastrearse simultáneamente a través de la fijación errónea de precios en lo económico, la inequidad social y la medición errónea de la garantía.

La cuarta base de conocimiento cambia de registro. No describe consecuencias de la calidad de la tasación; describe lo que la tasación *es*, metodológica y legalmente. Sus predicados son taxonómicos y no causales, y su maquinaria inferencial responde preguntas sobre jurisdicción y no sobre impacto. Este cambio es necesario porque el marco consecuencial de los tres primeros pasajes presupone una definición de la actividad que se evalúa, y esa definición está, ella misma, en disputa. Los predicados de alcance registran lo que hacen los tasadores —opiniones de valor presente basadas en evidencia, con componentes orientados al futuro estrechamente acotados—, mientras que los predicados de regulación registran las condiciones bajo las cuales esa actividad cae dentro de la práctica con licencia. Las reglas de encadenamiento hacia adelante (**requires_license/1**, **exempt_from_appraisal_license/1**, **regulated_appraisal/1**) proporcionan un mecanismo para preguntar si un trabajo de valuación determinado cae dentro o fuera del perímetro regulado, una pregunta que las tres primeras bases de conocimiento asumen ya respondida.

La quinta base de conocimiento avanza una tesis. Donde las otras describen, clasifican o rastrean consecuencias, la base de conocimiento del proxy argumenta que la cuidadosa separación doctrinal entre la tasación a valor presente y el pronóstico del valor futuro —la separación que documenta la base de conocimiento de la regulación— está en tensión con la función económica real de las tasaciones en el crédito hipotecario. La convención del 80 % de relación préstamo-valor, los siete riesgos específicos contra los que amortigua, la pregunta subyacente del prestamista registrada textualmente («Si este prestatario incumple en algún momento del futuro previsible, ¿es probable que la garantía siga siendo suficiente...?») y el sistema integrado de suscripción de siete componentes apoyan todos la misma afirmación: la tasación es formalmente una opinión de valor presente y económicamente un colchón de valor futuro. El compromiso histórico que produjo este arreglo —los tasadores como anclas disciplinadas de valor presente, los prestamistas como el locus de la interpretación del riesgo futuro— fue una respuesta defendible a las condiciones bajo las cuales surgió la licencia de tasación, pero los intensificadores modernos enumerados en el pasaje (los AVM, los sistemas de valuación con aprendizaje automático, la modelización de capital de Basilea, la analítica de titulización, las pruebas de estrés a gran escala) tratan los valores de las propiedades de forma probabilística a lo largo del tiempo. Esas herramientas

reincorporan capas doctrinalmente separadas a un único flujo computacional, y la separación conceptual entre capas se vuelve, en consecuencia, más difícil de sostener.

Si esa reconciliación es alcanzable, y en qué términos, es la pregunta con la que cierra el pasaje del proxy. Las bases de conocimiento no la responden. Sí proporcionan, en cambio, un relato explícito y consultable de las consideraciones relevantes para responderla: las consecuencias de la calidad de la valuación a través de horizontes y dominios, el límite jurisdiccional que determina qué consecuencias caen dentro de la práctica regulada, y las razones estructurales por las que ese límite está ahora bajo presión. Ese relato es el fundamento sobre el cual pueden construirse argumentos más específicos —sobre metodología, sobre la reforma de las normas, sobre la relación entre la práctica de la tasación y el sistema más amplio de valuación cuantitativa—.

Notas sobre la representación del conocimiento

Vale la pena registrar explícitamente tres patrones estructurales, ya que emergieron del proceso de codificación y no de ningún pasaje individual.

Primero, la familia de bases de conocimiento no es plana. Tres de las cinco comparten un esquema; una proporciona las calificaciones jurisdiccionales que determinan cuándo se aplica ese esquema; una avanza una tesis sobre la brecha entre el marco formal y la función económica. Una capa puente que combine las cinco no las trata como cinco columnas de la misma matriz. Trata a las tres primeras como descripciones de consecuencias, a la cuarta como la restricción que determina dónde operan esas descripciones, y a la quinta como la identificación del punto donde la propia restricción está bajo presión. Esta estructura asimétrica es una virtud y no un defecto: refleja la organización real de la práctica profesional, en la que las afirmaciones descriptivas, las jurisdiccionales y las críticas ocupan papeles epistémicos genuinamente distintos.

*Segundo, la disciplina de prefijos que distingue los módulos (*economic_*, *social_*, *collateral_*, *scope_/regulation_*, *proxy_/underwriting_*) solo hace un trabajo real allí donde existe riesgo de colisión.* Los predicados con contrapartes directas en otros módulos se ganan su prefijo; los predicados únicos de un solo dominio conceptual, no. El vocabulario de valores compartido en el lado de los argumentos —los átomos de calidad, horizonte y dominio— es precisamente lo que permite que los módulos se comuniquen, y prefijar esos valores lo impediría. El patrón que emergió es general: prefijo en los predicados, átomos compartidos en los argumentos, con una prefijación selectiva y no uniforme cuando el predicado es único de una sola lente.

Tercero, la tensión central identificada en la base de conocimiento del proxy está estrechamente relacionada con la asimetría regulatoria identificada en la base de conocimiento de alcance/regulación, pero es distinta de ella. Las dos se registran mediante predicados paralelos de dos argumentos —*regulatory_asymmetry/2* y *proxy_doctrine_practice_gap/2*— con nombres deliberadamente diferentes porque describen ejes distintos del mismo arreglo subyacente. La asimetría regulatoria concierne a quién está regulado y con qué intensidad; la brecha entre doctrina y práctica concierne a lo que formalmente se dice del producto del tasador frente a lo que económicamente se hace con él. Las dos tensiones se refuerzan mutuamente. La regulación empuja a los tasadores hacia la doctrina del valor presente; la función económica arrastra su producto hacia el uso como valor futuro. El cuadro combinado es el de una profesión cuya autodescripción oficial resulta cada vez más difícil de conciliar con los flujos analíticos que sus productos alimentan.

Referencias

Agarwal, Sumit, Itzhak Ben-David y Vincent Yao (2015). «Collateral Valuation and Borrower Financial Constraints: Evidence from the Residential Real Estate Market». En: *Management Science* 61.9, págs. 2220-2240. DOI: [10.1287/mnsc.2014.2002](https://doi.org/10.1287/mnsc.2014.2002).

- Alexandrov, Alexei, Laurie Goodman y Michael Neal (ene. de 2023). *Reengineering the Appraisal Process: Better Leveraging Both Automated Valuation Models and Manual Appraisals*. Inf. téc. 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. de 2021). *Dodd-Frank Act Stress Test 2021: Supervisory Stress Test Methodology*. Inf. téc. 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 y Parag Pathak (2011). «Forced Sales and House Prices». En: *American Economic Review* 101.5, págs. 2108-2131. DOI: [10.1257/aer.101.5.2108](https://doi.org/10.1257/aer.101.5.2108).
- Clauretie, Terrence M. y Nasser Daneshvary (2009). «Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time». En: *Real Estate Economics* 37.1, págs. 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 y Patrick M. McAllister (2010). «Means, Motive and Opportunity? Disentangling Client Influence on Performance Measurement Appraisals». En: *Journal of Property Research* 27.2, págs. 181-201. DOI: [10.1080/09599916.2010.499014](https://doi.org/10.1080/09599916.2010.499014).
- Deng, Yongheng, John M. Quigley y Robert Van Order (2000). «Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options». En: *Econometrica* 68.2, págs. 275-307. DOI: [10.1111/1468-0262.00110](https://doi.org/10.1111/1468-0262.00110).
- Diaz Julian, III y Marvin L. Wolverton (1998). «A Longitudinal Examination of the Appraisal Smoothing Hypothesis». En: *Real Estate Economics* 26.2, págs. 349-358. DOI: [10.1111/1540-6229.00749](https://doi.org/10.1111/1540-6229.00749).
- Ebrahim, M. Shahid y Sikandar Hussain (2010). «Financial Development and Asset Valuation: The Special Case of Real Estate». En: *Journal of Banking & Finance* 34.1, págs. 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». En: *Journal of Urban Economics* 111, págs. 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 y Paul S. Willen (2008). «Negative Equity and Foreclosure: Theory and Evidence». En: *Journal of Urban Economics* 64.2, págs. 234-245. DOI: [10.1016/j.jue.2008.07.006](https://doi.org/10.1016/j.jue.2008.07.006).
- Foster, Chester y Robert Van Order (1984). «An Option-Based Model of Mortgage Default». En: *Housing Finance Review* 3.4, págs. 351-372.
- Friedman, Jerome H. (1991). «Multivariate Adaptive Regression Splines». En: *The Annals of Statistics* 19.1, págs. 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». En: *Journal of Property Research* 13.4, págs. 261-273. DOI: [10.1080/095999196368781](https://doi.org/10.1080/095999196368781).
- Geltner, David M. (1991). «Smoothing in Appraisal-Based Returns». En: *The Journal of Real Estate Finance and Economics* 4.3, págs. 327-345. DOI: [10.1007/BF00161933](https://doi.org/10.1007/BF00161933).
- Ghent, Andra C., Walter N. Torous y Rossen I. Valkanov (2019). «Commercial Real Estate as an Asset Class». En: *Annual Review of Financial Economics* 11, págs. 153-171. DOI: [10.1146/annurev-financial-110118-123121](https://doi.org/10.1146/annurev-financial-110118-123121).
- Griffin, John M. y Gonzalo Maturana (2016). «Who Facilitated Misreporting in Securitized Loans?» En: *The Review of Financial Studies* 29.2. Evidence of appraisal inflation and its consequences, págs. 384-419. DOI: [10.1093/rfs/hhv130](https://doi.org/10.1093/rfs/hhv130).
- Hendershott, Patric H. y Edward J. Kane (jul. de 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>.
- Hendershott, Patric H. y Edward J. Kane (1995). «U.S. Office Market Values During the Past Decade: How Distorted Have Appraisals Been?» En: *Real Estate Economics* 23.2, págs. 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*. 10.^a 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. y James E. Larsen (2024). «A Primer on Regulations and the Practice of Residential Property Appraisal». En: *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». En: *The Journal of Real Estate Finance and Economics* 11.1, págs. 5-36. DOI: [10.1007/BF01097934](https://doi.org/10.1007/BF01097934).
- LaCour-Little, Michael y Stephen Malpezzi (2003). «Appraisal Quality and Residential Mortgage Default: Evidence from Alaska». En: *The Journal of Real Estate Finance and Economics* 27.2, págs. 211-233. DOI: [10.1023/A:1024728420837](https://doi.org/10.1023/A:1024728420837).
- Lee, Ming-Hsiang, Chien-Wen Peng y Hui-Fen Liao (2020). «An Analysis of Objectivity in the Real Estate Appraisal Process». En: *International Real Estate Review* 23.4, págs. 483-504.
- Mayer, Yanling G. y Frank E. Nothaft (2022). «Appraisal Overvaluation: Evidence of Price Adjustment Bias in Sales Comparisons». En: *Real Estate Economics* 50.3, págs. 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». En: *Journal of Real Estate Research* 28.2, págs. 193-214. DOI: [10.1080/10835547.2006.12091177](https://doi.org/10.1080/10835547.2006.12091177).
- Qi, Min y Xiaolong Yang (2009). «Loss Given Default of High Loan-to-Value Residential Mortgages». En: *Journal of Banking & Finance* 33.5, págs. 788-799. DOI: [10.1016/j.jbankfin.2008.09.010](https://doi.org/10.1016/j.jbankfin.2008.09.010).
- Quan, Daniel C. y John M. Quigley (1991). «Price Formation and the Appraisal Function in Real Estate Markets». En: *The Journal of Real Estate Finance and Economics* 4.2, págs. 127-146. DOI: [10.1007/BF00173120](https://doi.org/10.1007/BF00173120).
- Steurer, Miriam, Robert J. Hill y Norbert Pfeifer (2021). «Metrics for Evaluating the Performance of Machine Learning Based Automated Valuation Models». En: *Journal of Property Research* 38.2, págs. 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 %.

Residual Constraint Approach (RCA): marco metodológico y protocolo

Traducción al español de la versión actualizada del artículo publicado originalmente en Zenodo el 5 de marzo de 2025. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor, SRA

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

DOI: [10.67330/kgebx46](https://doi.org/10.67330/kgebx46)

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

Este artículo de acceso abierto se distribuye bajo los términos de la [licencia Creative Commons Atribución 4.0](#).

Resumen. Una limitación importante de los métodos convencionales de valuación radica en su dependencia de técnicas manuales obsoletas, como el análisis de pares comparables (*matched pairs*), que pasan por alto las contribuciones al valor de las características del inmueble como si no existieran: el *Sales Grid* (cuadro de comparación) tradicional no tiene columna para las contribuciones al valor, solo para ajustes. De hecho, las contribuciones al valor proporcionan la base necesaria para las restricciones matemáticas que impiden la sobrevaluación y la subvaluación por parte de los valuadores tradicionales.

El Residual Constraint Approach (RCA, es.: enfoque por restricciones residuales) hace avanzar el Sales Comparison Approach (SCA, enfoque de comparación de ventas) tradicional mediante la integración de un proceso de valuación multifásico que emplea Multivariate Adaptive Regression Splines (MARS) junto con rigurosas restricciones matemáticas sobre las contribuciones al valor de las características — en particular las características no medidas: variables latentes como la estética, el diseño, el estado y la calidad, cuyos valores suelen estar sujetos al juicio subjetivo y al sesgo del valuador tradicional.

Las restricciones matemáticas del RCA se basan en la aplicación experta de la regresión MARS para estimar las contribuciones al valor de las características medibles del inmueble al precio de venta. Sin embargo, esta estimación normalmente representa solo alrededor del 80 % del precio de venta real, excluyendo características evaluadas subjetivamente como el diseño, la utilidad funcional, el estado y la calidad. En la San Francisco Bay Area, el residuo — que representa estos componentes subjetivos — promedia alrededor del 20 % del valor total del inmueble. La cifra de 80/20 refleja la experiencia del autor en la San Francisco Bay Area bajo condiciones de datos típicas y no debe interpretarse como una constante invariante entre mercados. En mercados donde el R^2 alcanzable es menor, la respuesta del marco metodológico consiste en identificar y medir los determinantes del valor no medidos — la distancia al océano en comunidades costeras, la elevación en vecindarios de ladera y variables similares recopiladas de fuentes externas cuando los datos MLS no las proporcionan — en lugar de aceptar la limitación y reportar bandas de error más amplias. Si bien los residuos de MARS suelen tratarse como errores de estimación en otros contextos, en la valuación inmobiliaria sirven como medida indirecta del valor de las variables latentes. Aunque esto pueda parecer poco práctico al principio, una demostración establece que el residuo puede descomponerse de manera significativa en componentes descriptivos, siempre que sus contribuciones sumen el residuo — sin afectar el resultado final de la valuación.

Esta metodología surge de dos décadas de aplicación empírica por parte del autor, desarrollada mediante implementaciones iterativas de MARS en entornos R y Python.

Palabras clave: valuación inmobiliaria, avalúo residencial, Residual Constraint Approach, MARS, Multivariate Adaptive Regression Splines, Sales Comparison Approach, análisis de residuos, variables latentes, valuación automatizada

Índice

1. Introducción	4
2. Antecedentes	6
3. Valuation Engineer	9
4. Supuestos	10
5. Workflow del RCA	13
5.1. Tipos de análisis MARS	13
5.2. Workflow de procesamiento	13
5.3. Workflow de datos	15
5.3.1. Requisitos de datos para MARS	15
5.3.2. Protocolos: listings de ventas cerradas	16
5.4. Estructura de carpetas del proyecto	18
5.4.1. Protocolos: múltiples ejecuciones de MARS (earth)	19
5.5. El libro de Excel principal	21
5.6. Copias de seguridad	21
6. Notación del libro de Excel	22
6.1. Dimensiones descriptivas	22
6.1.1. La dimensión de proyecto (“a”)	23
6.1.2. Dimensión de versión de ejecución (“r”)	24
6.1.3. Etapas del flujo de trabajo (“w”)	24
6.1.4. Hojas del libro de Excel (“s”)	25
6.1.5. Listings de propiedades (“p”) - filas de MLSData	25
6.1.6. Variables MLS (“v”) Columnas de MLSData	26
6.1.7. Ejemplo final de notación	26
6.1.8. Justificación	26
6.1.9. ¿Matrices?	27
6.1.10. Primera simplificación	27
6.1.11. Segunda simplificación	27
7. El Sales Grid del RCA	29

8. Variables	33
8.1. Ontología de origen de las variables	33
8.2. Ontología de aplicación de las variables	34
8.3. Términos de segundo y tercer grado	36
8.3.1. Variables de ajuste	36
8.3.2. Variables de agregación	36
8.3.3. Variables especiales & notación	37
8.3.4. Descomposición de los residuos	37
9. Demostraciones fundamentales del RCA	39
10. Modelo MARS	41
10.1. Funciones “ g ” de Variables del Modelo y de Interacción	41
10.2. Indexación de las Funciones g	41
10.3. Dimensiones de Graficación	42
10.4. La Ecuación Final del Modelo	42
10.5. Número de Gráficas Necesarias para Graficar un Modelo	43
11. Características de los datos	46
11.1. Propósito del residuo	46
11.2. Las calificaciones de estado y calidad como regresores aproximados	47
11.3. Errores de datos	47
11.4. ¿Qué es la exactitud?	49
11.5. El inmueble a valorar como anomalía	50
11.6. La importancia del CVR2 frente a datos inesperados	51
12. Residuos	52
12.1. Ranking & Scoring (clasificación y puntuación)	52
12.2. Curva CQA-Residuo	55
12.3. Características de la curva CQA	55
12.3.1. Análisis de los componentes de la curva CQA	55
12.3.2. Dinámica del mercado e implicaciones sociales	56
12.4. Residuo del inmueble a valorar	56
12.5. Conclusión de valor del RCA	58
12.5.1. Consideraciones generales	58
12.5.2. Revisión, auditoría, quejas	58
13. Conclusión	60

1. Introducción

Este artículo presenta el fundamento lógico y matemático de un enfoque moderno para estimar con exactitud el valor de mercado de inmuebles residenciales y de otros tipos de bienes raíces. El método, desarrollado a lo largo de 20 años de experiencia aplicando el software Multivariate Adaptive Regression Splines (MARS) a viviendas del norte de California — en particular en comunidades del San Francisco Bay Area —, ofrece un avance significativo en la precisión de la valuación. MARS mejora la calidad de la valuación al capturar con destreza las relaciones no lineales entre las características del inmueble y el valor, a la vez que identifica y prioriza de manera eficiente las variables de mayor impacto. Esta capacidad permite a los valuadores descubrir la dinámica del valor en nuevas zonas de mercado con mucha mayor rapidez de la que permiten los métodos tradicionales. MARS constituye la columna vertebral del Residual Constraint Approach (RCA) (es.: enfoque por restricciones residuales) aquí propuesto. El método RCA hace honor a su nombre gracias a su marco riguroso:

- Impone restricciones estrictas sobre las contribuciones al valor de las características del inmueble, tanto medidas como no medidas:
 1. Todas las contribuciones al valor derivadas de MARS, incluido el residuo, deben sumar el precio de venta neto de cada inmueble comparable.
 2. El residuo puede descomponerse en subcomponentes con valores asignados, siempre que su total sea igual al residuo del comparable.
- RCA minimiza el sesgo en la estimación de las características subjetivas — tales como el atractivo estético, el estado y la calidad — al vincular su valor total al residuo, el cual se cuantifica con exactitud cuando un valuation engineer competente construye el modelo MARS.

Aunque está diseñado principalmente para bienes raíces residenciales, donde las características subjetivas influyen de manera significativa en el valor de mercado, RCA tiene potencial para aplicaciones más amplias de valuación de activos. Los avalúos exactos resultan críticos dado el valor estimado de \$45 billones (45 trillion) del mercado residencial de bienes raíces de los EE. UU. Fannie Mae señala que los avalúos típicos se desvían alrededor de un 5 % de los precios de venta reales, aunque esta cifra enmascara el sesgo, ya que los valuadores a menudo enfrentan presiones para alinearse con los precios de venta. Además, dado que el 60 %–70 % de los avalúos respaldan operaciones de refinanciamiento — en las que no existen precios de venta comparables —, la necesidad de precisión, como la que ofrece RCA, se hace evidente. El análisis de la Reserva Federal de 2012 estimó que el 15 %–20 % de los avalúos previos a la crisis de 2008 se desviaban en más de $\pm 10\%$ de los precios de venta, un rango probablemente aún vigente hoy para los avalúos de refinanciamiento, con variaciones según el tipo de inmueble y el mercado. ¿Por qué importa la exactitud? En transacciones con baja relación préstamo-valor (loan-to-value, LTV) — por ejemplo, un comprador con sólido historial crediticio que da \$500,000 de enganche por una vivienda de \$1,000,000 — los prestamistas pueden eximir del avalúo, al suponer un riesgo mínimo de ejecución hipotecaria. Sin embargo, los compradores que omiten el avalúo corren el riesgo de pagar de más y de descubrir defectos solo después de la compra. Esto subraya el valor de métodos de valuación robustos como RCA. Cuantificar la exactitud de los avalúos sigue siendo un desafío. Los investigadores carecen de métricas directas de las desviaciones respecto del “valor de mercado” y recurren, en su lugar, a indicadores indirectos. Los métodos tradicionales de valuación, arraigados en el juicio subjetivo y en la competencia variable de los valuadores, complican el análisis. Entre las estrategias empleadas se encuentran:

1. Tzioumis (2016) estudió a 4,782 valuadores a través de 259,436 avalúos en 11 estados, y encontró que un tercio se ubicaba dentro del 1 % del precio de contrato, la mitad entre el 1 %-3 % por encima, y una minoría en promedio un 5 % por encima — lo que sugiere un sesgo al alza. El estudio identificó dos tipos de valuadores: los “al valor de mercado” (cercaos al precio de contrato) y los “por encima del valor de mercado” (consistentemente más altos).
2. Krivorotov y LaCour-Little (2020) encontró que el 90 % de los avalúos de compra igualaban o superaban los precios de contrato, lo que pone de relieve el “sesgo de anclaje” — la tendencia a alinearse con un valor deseado —, especialmente en contextos de refinanciamiento.

Estos estudios revelan sesgo en las prácticas de valuación, amplificado por las limitaciones de las herramientas tradicionales. Si los valuadores fueran uniformemente competentes e imparciales, sus valuaciones convergerían; la sobrevaluación persistente por parte de ciertos grupos señala problemas sistémicos.

Los estudios sobre sesgo citados anteriormente, y el patrón sistémico que señalan, reflejan un rasgo más profundo de la práctica tradicional de comparación de ventas que RCA está diseñado para abordar. La metodología tradicional opera en dos canales con un estatus epistémico muy distinto. En el canal de las características medibles — GLA (superficie habitable bruta), número de habitaciones, tamaño del lote, antigüedad, ubicación — se espera que el valuador derive los valores de ajuste de manera objetiva, mediante análisis de pares emparejados o regresión lineal local sobre ventas comparables. Los valuadores competentes realizan este trabajo, y los ajustes resultantes son, cuando menos, aproximadamente defendibles frente a los datos. Aun aquí, sin embargo, el Sales Grid tradicional (cuadro de comparación) registra únicamente ajustes; ni siquiera cuenta con una columna para las contribuciones al valor. Tampoco se calculan los ajustes tradicionales como diferencias entre contribuciones al valor — se estiman directamente, a partir de pares emparejados o procedimientos similares, sin que las contribuciones subyacentes lleguen a computarse jamás. En RCA, por el contrario, cada ajuste es la diferencia entre las contribuciones al valor del inmueble a valorar y las del comparable para la característica en cuestión — la estructura cuyas consecuencias precisa la Demostración II. En el canal de las características subjetivas — estado, calidad de construcción, diseño, utilidad funcional, estética — los procedimientos de respaldo disponibles van de endebles a precarios. Los ajustes por características subjetivas pueden respaldarse mediante análisis de pares emparejados de inmuebles ajenos al conjunto inmediato de comparables, mediante apelación a la práctica reconocida entre pares, o mediante referencia a la experiencia documentada del valuador en el mercado. Los tres son formalmente conformes con las normas profesionales. Ninguno alcanza el rigor estadístico que el canal medible logra de manera rutinaria.

El resultado es una asimetría procedimental entre lo que exigen las normas y lo que la metodología puede ofrecer. Las normas exigen respaldo para cada ajuste, sea medible o subjetivo. La metodología ofrece un respaldo robusto en un canal y un espectro de procedimientos más débiles en el otro, que van desde un par emparejado anecdótico de tamaño muestral uno hasta la cita de lo que otros valuadores acostumbra usar. El valuador honesto y competente que realiza un cuidadoso trabajo de pares emparejados para un ajuste por vista produce un respaldo significativamente mejor que el de quien cita la práctica de sus pares sin fundamento; pero un revisor que compare ambos avalúos en el formulario estándar no puede distinguirlos con facilidad, porque el formulario registra el valor del ajuste, no el rigor de su derivación. La metodología tolera la asimetría porque ningún procedimiento más sólido para las características subjetivas ha estado disponible de manera general para los valuadores en ejercicio.

RCA mejora ambos canales, pero de maneras categóricamente distintas. En el canal medible, la regresión MARS captura relaciones no lineales, interacciones entre variables y una estructura de funciones base que ni el análisis de pares emparejados ni la regresión lineal ordinaria pueden

representar. Ajustadas contra el conjunto completo de ventas calificadas de la zona de mercado — típicamente varios cientos — en lugar de contra los tres a nueve comparables del Sales Grid tradicional, las contribuciones al valor derivadas de MARS son a la vez objetivas y sustancialmente más exactas que la metodología a la que reemplazan. En el canal subjetivo, RCA hace algo distinto en especie de cualquiera de los procedimientos tradicionales de respaldo. No intenta en absoluto estimar valores de ajuste para características subjetivas individuales. Las contribuciones de todas las características subjetivas — variables latentes — aparecen colectivamente como el residuo: la diferencia entre el precio de venta neto real de cada comparable y su precio predicho por MARS, calculable a partir de los datos sin intervención del valuador. La descomposición de ese residuo en componentes con nombre — Diseño, Utilidad Funcional, Vista, Estado, Calidad, etcétera — está restringida a sumar el residuo total, y por la Demostración II (Sección 9) cualquier descomposición que satisfaga esa restricción produce la misma conclusión de valor. El juicio que permanece en el flujo de trabajo es la ubicación del inmueble a valorar dentro de la distribución ordenada de los residuos de los comparables, realizada contra una clasificación inspeccionable y documentada explícitamente en el informe. Un avalúo RCA débil sigue siendo posible — un VE inexperto puede producir un modelo MARS sobreajustado, ubicar el inmueble a valorar de manera inverosímil o pasar por alto anomalías en los datos — pero es débil de maneras que un revisor puede identificar al inspeccionar el modelo y la clasificación, y no de maneras que una revisión con el formulario estándar no pueda sacar a la luz.

Más allá del refinanciamiento, los valuadores deben a los propietarios valuaciones exactas, especialmente en compras de alto valor, donde los errores pueden costar decenas o cientos de miles. Sin embargo, muchos se resisten a gastar \$600–\$1,000 en avalúos, dudando de su valor agregado. Los métodos avanzados como RCA, si bien inicialmente intensivos en recursos, prometen mayor precisión. Con refinamiento y experiencia, su eficiencia — y su adopción — deberían aumentar.

2. Antecedentes

La formación académica del autor es en matemáticas, ciencias de la computación y estadística. De profesión, es ingeniero de software con experiencia secundaria en valuación y contabilidad. Ha realizado avalúos, de manera intermitente, desde 2002 y ha aplicado MARS desde aproximadamente 2004 para determinar ajustes de propiedades y tendencias de precios en bienes raíces residenciales.

Valuar con MARS constituye una “experiencia profunda” en valuación. MARS genera modelos que reflejan con mayor detalle que cualquier otro método estadístico tanto las variables que contribuyen al valor de mercado como la manera en que los cambios en dichos valores impactan el precio de venta. Así, al valuar con MARS, el valuador investiga constantemente las relaciones causales entre los valores de las variables y los precios de venta generados por los modelos MARS.

En aras de la transparencia, la metodología RCA (es.: enfoque por restricciones residuales) se aplica mejor a zonas residenciales con las siguientes características:

1. Vecindarios complejos con inmuebles bastante diferentes entre sí

- Viviendas antiguas
- Calidades variadas de renovación
- Muchas arquitecturas y diseños diferentes
- Variación en atributos del inmueble como la GLA (superficie habitable bruta), el tamaño del lote, la elevación y la vista.

2. Condiciones de mercado complejas

- Condiciones de mercado que cambian rápidamente
 - Muchos tipos diferentes de compradores
3. Buenas fuentes de datos: el protocolo RCA se basa en la minería de datos y requiere datos abundantes y de calidad. Si bien puede utilizarse en zonas con datos escasos, podría no justificar el tiempo y el costo requeridos.
 4. Se desean estimaciones exactas del valor de mercado. RCA debería, en la mayoría de los casos, poder proporcionar una exactitud de $\pm 1-2\%$. El informe del valuation engineer debe comentar las condiciones de exactitud en la conclusión de valor final. Si faltan datos o información, o si no se han descubierto patrones consistentes en el mercado mediante la minería de datos, la exactitud se verá afectada y se deberá informar al cliente de la situación. Específicamente, un R^2 (coeficiente de determinación) y un CVR2 (R^2 con validación cruzada, calculado mediante la partición repetida del conjunto de datos de la MLS en particiones aleatorias de entrenamiento y de prueba) elevados, un modelo MARS significativo y un patrón fuerte en las características CQA (estado, calidad y atractivo) del inmueble dentro de los residuos ordenados son los principales medios para evaluar la exactitud y la confiabilidad de la conclusión de valor.

El desarrollo de estos protocolos RCA a lo largo de los últimos más de 20 años se basa en la disponibilidad de MLSListings Inc. (Sunnyvale, CA) y de otras MLS, actualmente accesibles a través de MLSListings, lo cual comprende más del 95 % de la experiencia del autor en análisis de datos. La pericia del autor proviene de un extenso trabajo de valuación en 15 condados primarios del norte de California, con experiencia adicional en 10 condados secundarios.

La metodología RCA es más eficaz en mercados que cumplen dos criterios clave:

1. Amplia disponibilidad de datos de valuación de alta calidad
2. Zonas residenciales complejas y heterogéneas caracterizadas por:
 - Estilos arquitectónicos diversos
 - Grados variables de renovación de los inmuebles
 - Patrones de desarrollo que abarcan varias décadas

Experiencia en condados de California

Las afirmaciones realizadas en este artículo se basan en la experiencia de valuación de inmuebles en los siguientes condados del norte de California:

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

Las zonas anteriores abarcan esencialmente el Silicon Valley y sus condados circundantes, extendiéndose hacia zonas más rurales, incluidos el Valle Central y las estribaciones de la Sierra. Cabe mencionar que el autor siempre ha tenido acceso a datos abundantes y *relativamente*

exactos en comparación con lo que muchos valuadores de otras zonas menos desarrolladas de los EE. UU. tienen a su disposición.

Los datos de la MLS utilizados cuentan con certificación RESO y cumplen con el RCF (RESO Common Format). A pesar de la certificación RCF, los datos de importancia usualmente secundaria a veces faltan o son incorrectos para algunos inmuebles, ya sea por errores de captura originales del Tax Assessor (autoridad de valuación fiscal) o por errores posteriores de los agentes de ventas al ingresar datos en la MLS. Con frecuencia estos errores siguen un patrón, como simplemente no realizar una captura si la característica no existe. Por ejemplo, en lugar de ingresar 0 en “carport spaces” cuando no hay cochera abierta, no se ingresa nada. En zonas donde las cocheras abiertas son poco comunes, el analista podría sustituir la ausencia de datos por 0, a fin de obtener un ajuste útil para las cocheras abiertas. Si bien R/earth y otras implementaciones de MARS pueden manejar datos faltantes, corresponde al Valuation Engineer (VE) decidir cómo tratarlos.

Los datos de MLSListings también están disponibles a partir de las MLS de algunos otros estados, desde Florida hasta Hawái. Esto puede ser útil para probar qué tan bien funcionan los protocolos y el código en diferentes estados.

3. Valuation Engineer

Definición 1: Valuation Engineer

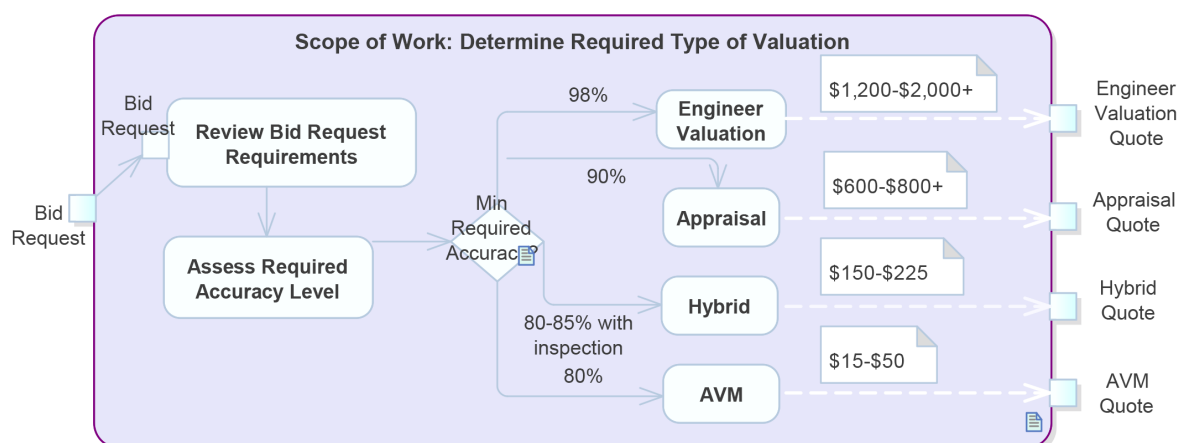
Para evitar ambigüedades, el término “Valuation Engineer” (o “VE”, en forma abreviada) se utiliza para denotar a un individuo que posee la pericia de un valuator con licencia, junto con conocimientos, habilidades y experiencia avanzados en la utilización de minería de datos, estadística, programación, protocolos detallados e inteligencia artificial para determinar el valor de mercado de inmuebles complejos. También es importante reconocer que la valuación es una tarea que requiere alguna forma de licencia debido al potencial comprobado de fraude.

Los prestamistas suelen ofrecer varias opciones de valuación:

1. Modelos de valuación automatizada (AVM, por sus siglas en inglés) - la opción de menor costo
2. Avalúos híbridos - que utilizan inspectores sin licencia bajo la supervisión de un valuator con licencia
3. Avalúos estándar - realizados por valuadores con licencia

Ausentes de la consideración habitual para el trabajo de valuación se encuentran los avalúos de nivel de ingeniería. Sencillamente no hay muchos valuadores con el conjunto de habilidades necesario para realizarlos. Debería existir una necesidad de ellos. Pueden proporcionar una exactitud significativamente mayor y casi garantizar una objetividad bastante buena mediante el empleo de la metodología RCA (MARS y restricciones).

Para inmuebles conformes en fraccionamientos más nuevos, los avalúos estándar generalmente son suficientes. Sin embargo, las valuaciones de nivel de ingeniería se recomiendan para viviendas antiguas en zonas con un historial significativo de renovaciones que abarca más de 30 años. La elección entre estos métodos debe seguir un proceso de decisión estructurado basado en las características del inmueble y las condiciones del mercado.



4. Supuestos

Supuesto 1: Área geográfica & tipo de inmueble

El alcance de este artículo se centra principalmente en viviendas unifamiliares y condominios en las áreas metropolitanas y urbanas de California. Si bien el Sales Comparison Approach (enfoque de comparación de ventas) se aplica a otros tipos de inmuebles y regiones geográficas, los enfoques de ingresos y de costos suelen desempeñar un papel más dominante al valorar terrenos agrícolas, propiedades multifamiliares y bienes raíces comerciales. En tales casos, el enfoque RCA puede requerir modificaciones o tener una aplicabilidad limitada.

Supuesto 2: Almacenamiento de datos

CSV y Excel son ampliamente utilizados por los sistemas MLS y se asumen como la vía principal para alimentar datos a los programas RCA. Aunque los feeds de brokers pueden utilizarse para alimentar datos de manera continua a bases de datos SQL. El respaldo y el almacenamiento intermedio a menudo emplean bases de datos SQL, y dos que desempeñan papeles importantes son las gratuitas y de código abierto SQLite y PostgreSQL. Los archivos de texto o RTF también se utilizan para fragmentos de informes, y los archivos PNG, JPG y TIFF se utilizan para gráficos. Nótese que el SQL Server de Microsoft también es utilizado por muchos, pero la edición gratuita “SQL Server Express” es muy limitada (p. ej., solo se puede usar 1 núcleo de CPU) y las demás versiones tienen cuotas de licencia considerables para uso comercial. Las bases de datos de bienes raíces en California pueden ser muy grandes y los joins pueden consumir mucho tiempo, por lo que una computadora con 8 o más núcleos y 128G de RAM puede ser útil para joins grandes. Las bases de datos SQL también son importantes cuando se necesita encontrar todos los inmuebles dentro de un área delimitada arbitrariamente. Así, si por ejemplo se utilizó el análisis de conglomerados (Cluster Analysis) para definir vecindarios, uno se inclinaría a usar alguna base de datos SQL. Sin embargo, las distintas bases de datos tienen distintas funciones integradas para manejar información geográfica y, por lo tanto, no son compatibles entre sí. Para esta serie de artículos sobre RCA, se asumirá PostgreSQL para las tareas SQL complejas y SQLite para las tareas simples.

Supuesto 3: Se utiliza “R/earth” para el procesamiento MARS.

Hasta 2017, el software MARS de Salford Systems era una solución asequible a \$270 por año. Sin embargo, tras la adquisición de Salford Systems por parte de Minitab, MARS fue incorporado a su paquete SPM a \$16,000 anuales — un precio prohibitivo para la mayoría de las pequeñas empresas. Afortunadamente, el paquete “earth” del lenguaje de programación R (Milborrow [2024a](#)), mantenido por Stephen Milborrow, ha evolucionado hasta igualar o superar las capacidades del SPM MARS de Minitab. El paquete earth ofrece ventajas adicionales: está disponible de forma gratuita, se integra sin problemas con el completo ecosistema estadístico de R y puede accederse desde Python mediante Pandas, lo que proporciona una excelente flexibilidad para diferentes flujos de trabajo. Las discusiones relacionadas con MARS asumen el uso específico de R/earth.

Supuesto 4: Notación

Los datos crudos provenientes de fuentes externas, como archivos CSV y hojas de cálculo de Excel, se representan mediante caracteres latinos estándar. Esto incluye los datos de inmuebles del Multiple Listing Service (MLS), los ajustes de configuración de MARS (Multivariate Adaptive Regression Splines) mediante el paquete R/earth y los datos específicos del proyecto. Esta notación se extiende a sus correspondientes representaciones como data frames en R o Python.

Al tratar las *características del inmueble* y sus diversos tipos y subconjuntos, la notación cambia a letras griegas para una mayor precisión y claridad analítica. Si bien esta notación matemática puede parecer más rigurosa, cumple un propósito crucial: distinguir las características abstractas del inmueble de sus representaciones en datos crudos. Por convención, los caracteres griegos denotan consistentemente características del inmueble, mientras que los caracteres latinos indican estructuras de datos de hoja de cálculo. Por ejemplo, el símbolo M representa la matriz de datos de la MLS, organizada con los inmuebles como filas y sus características correspondientes como columnas.

Supuesto 5: Lenguajes de programación

En este campo, R y Python con Pandas son los lenguajes dominantes, utilizándose C o C++ cuando la velocidad es importante. R/earth está escrito, a su vez, en R y C.

Supuesto 6: Indexación

La indexación puede variar entre los lenguajes de programación y las matemáticas, lo cual puede resultar confuso:

- Las matemáticas y R comienzan la indexación en 1.
- C, C#, C++ y Python comienzan en 0.

En este artículo, para mantener la consistencia y la simplicidad, toda indexación comienza en 0 (como en Python o C). Funciona de la siguiente manera:

- Para listas, matrices o data frames:
 - El índice 0 se reserva para el “inmueble a valorar” (el inmueble principal objeto de estudio).
 - Ejemplo: en una tabla donde las filas son inmuebles, la fila 0 es el inmueble a valorar y la columna 0 contiene los identificadores únicos de todos los inmuebles.
 - Un consejo práctico: ordene los inmuebles con el inmueble a valorar en primer lugar (fila 0) y luego liste los demás de la fecha de venta más reciente a la más antigua, numerándolos secuencialmente (0, 1, 2, etc.).
- Cuando no hay inmueble a valorar:
 - Si una estructura de datos no incluye inmuebles (p. ej., solo variables o funciones), el índice 0 se omite y la indexación comienza en 1.
 - Nota especial para usuarios de R: dado que R utiliza indexación basada en 1, al programar en R las matemáticas de este artículo, sume 1 a todos los índices para ajustar de la indexación basada en 0 a la basada en 1.

5. Workflow del RCA

Comprender el flujo de trabajo del RCA es esencial para un análisis eficaz. El proceso sigue una secuencia estructurada de tareas analíticas y puntos de decisión

5.1. Tipos de análisis MARS

El flujo de trabajo del análisis MARS depende de si se necesita únicamente un análisis de mercado o si se requiere la valuación o el avalúo de un inmueble determinado. Obsérvese que el avalúo de un inmueble específico normalmente requiere un análisis de su zona de mercado.

- **Análisis de mercado:** se centra en áreas amplias tales como condados, ciudades o regiones definidas. Este enfoque no involucra un inmueble a valorar específico y, por lo tanto, no requiere una conclusión de valor. El flujo de trabajo se concentra en el desarrollo del modelo, su validación y la generación de analítica de soporte. El proceso concluye una vez que el modelo óptimo ha sido identificado y documentado.
- **El análisis de un inmueble individual** comprende todos los elementos del análisis de mercado más la determinación del valor de un inmueble específico. Esto difiere de manera fundamental de los métodos tradicionales de valuación. En los enfoques tradicionales, el valor se deriva típicamente ajustando de 3-9 ventas comparables y calculando un promedio ponderado. En cambio, el RCA determina el valor principalmente mediante dos componentes:
 1. El valor predicho por el modelo MARS
 2. El residuo estimado específico del inmueble (con la puntuación CQA asociada)

El Sales Grid en el RCA sirve principalmente para explicar las diferencias de valor entre el inmueble a valorar y los comparables, ilustrando las contribuciones de las variables. Una ventaja clave del RCA es su estabilidad: si bien los clientes pueden debatir la selección de comparables, la conclusión de valor permanece consistente independientemente de cuáles comparables se elijan para la presentación, ya que todas las ventas (sean 15 o 600) se ajustan mediante el mismo modelo. Aunque las ventas más antiguas se ajustan automáticamente a las condiciones actuales del mercado, las ventas recientes son, en general, preferibles. La eficacia del modelo para capturar los factores de precio disminuye con la distancia temporal, pues las características de mercado no medidas cobran mayor importancia en períodos más largos.

5.2. Workflow de procesamiento

El flujo de trabajo general de procesamiento puede describirse mediante estos pasos:

1. Alcance del trabajo.
2. Configurar las carpetas y archivos iniciales del proyecto.
3. Descargar los datos del MLS. Consúltense los Protocolos 1.1–1.3 sobre el número de registros de inmuebles necesarios y el establecimiento de limitaciones en los rangos de fechas.
4. Completar las hojas de configuración: Regression Variables, Interactions, Aggregations, Calculations.
5. Ejecutar la regresión MARS (R/earth) para crear el modelo de precios.

6. Generar las contribuciones al valor de los inmuebles & los residuos (residuo puro y residuo/SF) a partir del modelo MARS.
7. Ordenar los inmuebles por residuo/SF.
8. Crear puntuaciones CQA para cada venta comparable.
9. Suponiendo que se ha producido una clasificación de calidad de los inmuebles que verdaderamente representa el atractivo general de los mismos, encontrar la mejor posición para el inmueble a valorar dentro de la clasificación y asignarle una puntuación CQA, estudiando las tendencias generales de los patrones en la clasificación.
10. Asignar el residuo/SF correspondiente al inmueble a valorar como residuo/SF · GLA (del inmueble a valorar).
11. **Sumar la estimación de precio de MARS y el residuo asignado para llegar a la conclusión de valor del inmueble a valorar.**
12. Descomponer el residuo del inmueble a valorar en componentes descriptivos y contribuciones al valor.
13. Calcular las contribuciones al valor & los ajustes producidos por el modelo para los comparables, además de todos los subtotales y los precios de venta ajustados.
14. Los precios de venta ajustados serán iguales a la conclusión de valor final del inmueble a valorar.
15. Seleccionar los 6-12 inmuebles más similares para el Sales Grid.
16. Descomponer los montos de residuo total de los comparables en contribuciones al valor de los componentes residuales, fijando el valor total del residuo en 0 una vez completada la descomposición.
17. Recalcular todos los subtotales y el precio de venta ajustado de los comparables del Sales Grid. Verificar los valores.
18. Cargar el Sales Grid al informe.

Obsérvese que, con el RCA, la conclusión de valor final NO se deriva directamente de los valores de los comparables en el Sales Grid. Con el método RCA, el Sales Grid se convierte en una herramienta destinada únicamente a explicar por qué inmuebles que parecen similares al inmueble a valorar se venden por más o por menos que la conclusión de valor final o el valor de avalúo del inmueble a valorar. Esto debe verse de otra manera: La conclusión de valor del inmueble a valorar depende de la estimación de precio del modelo MARS y del residuo estimado por el Valuation Engineer. Pero estas dos cantidades dependen en gran medida de todos los inmuebles que entraron en la regresión de R/earth, la cual creó la clasificación de residuos de los inmuebles que determinó la puntuación CQA del inmueble a valorar y, por consiguiente, determinó su residuo mediante la ubicación en la clasificación efectuada por el Valuation Engineer. Los ajustes en el Sales Grid están determinados, en realidad, por las restricciones matemáticas impuestas por el RCA; aunque la descomposición del residuo queda en cierta medida a merced del VE, los totales de las contribuciones al valor de la descomposición deben sumar siempre el residuo asociado. Sabemos por la Demostración 2 que, si bien los valores efectivos de los componentes residuales pueden variar según el juicio del VE, sus totales para un inmueble dado están completamente restringidos y, por lo tanto, no cambiarán el precio de venta ajustado del inmueble asociado, sin importar cómo altere el VE la descomposición del residuo.

5.3. Workflow de datos

Obsérvese que en este artículo se presenta únicamente un flujo de trabajo de alto nivel para el método RCA. Otros artículos profundizarán en las diversas etapas y subetapas del análisis RCA.

5.3.1. Requisitos de datos para MARS

Será necesario especificar una fecha inicial para las transacciones de venta y una fecha final basada en la fecha efectiva del avalúo o la valuación. Esto normalmente toma en consideración el número estimado de variables de regresión e interacciones que se necesitarán para cada grado de interacción superior a 1.

5.3.2. Protocolos: listings de ventas cerradas

Protocolo 1.1: Determinación del número requerido de listings de ventas cerradas del MLS

Este protocolo describe cómo calcular el número mínimo recomendado de listings de ventas cerradas que deben ingresarse a la regresión MARS. Será necesario hacer una estimación aproximada para comenzar, ejecutar MARS en ejecuciones de primer, segundo y posiblemente tercer grado para generar el modelo inicial, y luego revisar el número de filas conforme a las reglas siguientes para determinar si se necesitan más filas. Este protocolo supone que las variables independientes son 100% independientes, lo cual rara vez ocurre en bienes raíces; por ejemplo, la GLA (superficie habitable bruta) y los conteos de habitaciones suelen estar altamente correlacionados, y es probable que la GLA también muestre correlación con el tamaño del lote. Esto lleva al cálculo de los “Grados de Libertad (Degrees of Freedom, DOF).” Cada variable verdaderamente independiente recibe un grado de libertad. Pero, cuando existen dependencias entre las llamadas variables independientes, un cálculo exacto de los DOF requiere un análisis que está más allá del alcance de este artículo. Simplemente téngase presente que, debido a la probable colinealidad entre variables, las estimaciones siguientes probablemente queden de altas.

1. Para los términos de primer orden (sin interacciones), la regla de ~ 15 filas por variable funciona bien
2. Por cada interacción de dos vías permitida, deben añadirse aproximadamente 15-30 inmuebles
3. Para las interacciones de orden superior (3+ variables), cada una debería contar con 30+ inmuebles.

Sin embargo, esta no es una regla rígida. El número real necesario puede depender de:

- La relación señal-ruido de los datos
- La complejidad de las relaciones que se intenta modelar
- La variabilidad de la variable de respuesta
- La calidad de los datos

Ejemplo:

Así, si se tiene:

- 10 variables principales
- 5 interacciones de dos vías permitidas
- 2 interacciones de tres vías permitidas

El número mínimo de ventas podría calcularse como:

$$\begin{aligned}
 N &= (10 \text{ variables} * 15 \text{ filas/variable}) + \\
 &\quad (5 \text{ interacciones-de-dos-vías} * 22 \text{ filas/interacción-de-dos-vías}) + \\
 &\quad (2 \text{ interacciones-de-tres-vías} * 30 \text{ filas/interacción-de-tres-vías}) \\
 &= 320 \text{ filas}
 \end{aligned}$$

Como alternativa, puede procederse de la siguiente manera: Después de ejecutar R/earth sobre un conjunto de datos, puede usarse el número de funciones base del modelo, más uno, como estimación de los grados de libertad, y luego multiplicar este número por 20-30 para obtener el número de registros de entrada necesarios.

Protocolo 1.2: Los períodos de venta deben abarcar cambios mínimos en las preferencias de los compradores

Al recopilar datos históricos de inmuebles, no existe una limitación temporal estricta si los cambios en las preferencias del mercado pueden capturarse mediante variables apropiadas. Considérense estos puntos clave para manejar la evolución arquitectónica en el mercado: Los cambios en los estilos arquitectónicos dominantes (como un desplazamiento del estilo California Ranch al Mediterráneo) pueden modelarse mediante:

- Identificadores de subdivisión
- Códigos de área del MLS
- Rangos de fecha de construcción

Sin embargo, si estos cambios de estilo no pueden capturarse adecuadamente mediante variables específicas, se enfrenta una restricción importante: si bien la fecha de venta se utiliza típicamente para ajustar por condiciones de mercado, emplearla simultáneamente para controlar las preferencias arquitectónicas confundiría estos dos efectos distintos. En tales casos, el enfoque preferido consiste en:

1. Limitar la recopilación de datos al período posterior a la introducción del estilo arquitectónico más reciente
2. Restringir la selección de comparables a subdivisiones que compartan el mismo estilo arquitectónico

Esto garantiza que el análisis mantenga la consistencia en las características de los inmuebles, permitiendo a la vez un ajuste adecuado por las condiciones de mercado a lo largo del tiempo. Como alternativa, al menos en este caso, suponiendo que el estilo arquitectónico pueda extraerse de los datos, podría considerarse la interacción entre estas dos variables. Sin embargo, tal extracción no siempre es posible.

Protocolo 1.3: En una zona con pocas ventas, haga lo mejor que pueda

Al trabajar con modelado MARS sobre datos de ventas, se necesitan aproximadamente 15 transacciones para un análisis significativo. Si se dispone de menos ventas (por ejemplo, 5 inmuebles), puede duplicarse cada registro dos veces para alcanzar el umbral de 15 inmuebles. Consideraciones importantes para este enfoque:

1. Utilizar únicamente regresión de primer grado para reducir el sobreajuste
2. Ignorar u omitir la validación cruzada, ya que el conjunto de datos es demasiado limitado
3. Si bien MARS probablemente superará a los métodos de pares emparejados (matched pairs) o de regresión lineal estándar en este escenario, téngase presente que la duplicación de datos afecta la validez de las mediciones de incertidumbre y de la significancia estadística

La mejor solución a largo plazo es ampliar el conjunto de datos con ventas reales adicionales. Sin embargo, si debe procederse con datos limitados, la duplicación de datos ofrece una solución temporal viable, reconociendo a la vez sus limitaciones estadísticas.

5.4. Estructura de carpetas del proyecto

Conviene analizar la estructura de carpetas del proyecto. Es una buena idea tener una carpeta separada para los análisis de mercado y otra para los avalúos de inmuebles individuales. Ambas carpetas tienen una estructura de subcarpetas similar:

```

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
│   │       │           └── ...

```

El procesamiento de la Etapa 1 creará la carpeta de versión con sus subcarpetas. Copiará el archivo `MLS_Original.xlsx`, aun cuando esta hoja de cálculo *no debería* cambiar nunca, sino reflejar exactamente lo que se descargó del MLS en una fecha y hora determinadas; no obstante, en un proyecto más prolongado existe la posibilidad de que se necesite una nueva descarga del MLS. A continuación, esta etapa hará una copia del libro `MLS.xlsx` expandido, con sus parámetros de configuración, dentro de la carpeta de versión. Puede haber muchas carpetas de versión, cada una creada para una ejecución completa de MARS.

Luego, las etapas siguientes generarán adiciones de columnas, ordenamientos y otros cambios

en el libro de Excel, y escribirán una versión actualizada como `MLS_Stage(N)` en la carpeta de datos del proyecto cuando el avalúo se haya completado. Para reiterar: el `MLS.xlsx` raíz experimenta cambios constantes, de modo que la copia de versión es una mutación continua del libro `MLS` original, con eliminaciones y adiciones de variables, además de otros cambios. El código de procesamiento en R o Python muy probablemente se ejecutará con la entrada del libro `Data MLS.xlsx`. De este modo, siempre se cuenta con una copia de todos los datos que respaldan cada versión de la ejecución del RCA.

Se crean tantas versiones como sea necesario hasta que ya no sea posible lograr mejoras, momento en el cual se selecciona la mejor versión como versión final. Podría cambiarse el nombre de la carpeta añadiendo “_Final” al final, de modo que resulte evidente al examinar el directorio. Algunas versiones anteriores con modelos interesantes o competitivos deberían normalmente conservarse. Cada carpeta de versión contiene una serie de archivos de texto, diagramas y gráficos que pueden incorporarse a un informe. El libro final `MLS_Stage5.xlsx` contiene tantas hojas como sean necesarias para el informe. En particular, puede tener una hoja de cálculo para transferirla a un Sales Grid de Fannie Mae. También contendrá necesariamente los cálculos o la descomposición detrás de las agregaciones. Esta es potencialmente la parte que más tiempo consume del enfoque RCA: la lucha por encontrar el modelo perfecto en una zona de mercado difícil.

Todo lo necesario para recrear los sales grids debe copiarse a la carpeta de versión, incluido el código R o Python utilizado — si el código experimenta cambios frecuentes.

5.4.1. Protocolos: múltiples ejecuciones de MARS (earth)

Protocolo 2.1: Ejecuciones con sobreajuste para estudiar quirúrgicamente los factores de precio de la zona de mercado

Durante las ejecuciones iniciales de earth, el sobreajuste debe usarse estratégicamente para descubrir los factores de precio subyacentes. Por ejemplo, una zona de mercado representada por 800 ventas puede contener solo 3 inmuebles con talleres, y en esos casos aislados el taller puede tener un impacto importante en el precio de venta final. Una ejecución de solo pasada hacia adelante (forward pass) producirá un modelo sobreajustado que muestra la contribución al valor aproximada de estos talleres. Sin embargo, dado que solo hay 3 ventas de este tipo entre 800, los términos de taller serán eliminados en la pasada hacia atrás (backward pass), y con certeza bajo validación cruzada. El valor que la pasada solo hacia adelante asigna a los talleres puede anotarse y transferirse al análisis de variables latentes, para usarse con cualquier inmueble con taller que aparezca en el Sales Grid. Estos modelos sobreajustados también proporcionan estimaciones detalladas de muchas otras características con datos escasos que serían eliminadas por la poda de la pasada hacia atrás o por las ejecuciones con validación cruzada.

Guárdese cada una de estas ejecuciones exploratorias como su propia carpeta de versión: las contribuciones al valor registradas son evidencia que el VE querrá citar más adelante al descomponer los residuos. Entiéndase, por supuesto, que el VE debe apoyarse en última instancia en la poda y la validación cruzada para lograr que el modelo final sea robusto frente a inmuebles que la regresión nunca vio — como un inmueble a valorar que aparezca posteriormente. Las ejecuciones sobreajustadas son instrumentos de estudio; nunca constituyen el modelo final.

Protocolo 2.2: Múltiples ejecuciones con validación cruzada bajo diferentes parámetros de `earth()`

Las zonas de mercado difieren en estructura y tamaño, y el VE normalmente necesita experimentar con una serie de ejecuciones: para desarrollar una comprensión de cuáles variables impactan más el precio; para juzgar si el conjunto existente de variables medidas cumple adecuadamente con la estimación del valor, o si deben añadirse nuevos datos (distancia al agua, elevación, distancia a centros comerciales, etcétera); y para descubrir nuevos tipos de interacciones, a la vez que se descartan otras por ineficientes o engañosas. Al final, el VE busca un buen R2 sobre los datos de entrenamiento y el mejor CVR2 posible frente a los datos de prueba. Llegar a ello puede requerir recopilar nuevos datos, mejorar la exactitud de los datos existentes o afinar los parámetros de `earth()` según el tamaño y el carácter de la zona de mercado (p. ej., según los grados de libertad disponibles — véase el Protocolo 1.1).

Precisamente por esto la estructura de proyecto anterior admite muchas carpetas de versión de ejecución: cada modelo candidato se conserva con sus parámetros y su R2/CVR2, de modo que la trayectoria a lo largo de las ejecuciones pueda compararse. Una versión cuyo R2 sube mientras su CVR2 se estanca o cae está ajustando ruido en lugar de señal de mercado, y el historial de versiones es lo que hace esto visible.

Protocolo 2.3: Expansión mediante Prolog de las columnas de comentarios y de características descriptivas

Aunque esta facilidad se encuentra todavía en una fase un tanto experimental, `earthUI` incorpora ahora un motor Prolog integrado (`vProlog`) y utiliza gramáticas de cláusulas definidas (Definite Clause Grammars, DCGs) para analizar sintácticamente las columnas de texto libre del MLS — Public Remarks ante todo, pero también los comentarios de los agentes y las listas de características delimitadas por comas, como las características de las habitaciones o las estructuras existentes. La expansión convierte frases como “owned solar,” “3 car garage,” o “40x50 workshop” en columnas de características ordinarias que pueden ofrecerse a la regresión, y las características con datos escasos recuperadas de esta manera alimentan directamente el análisis de variables latentes (véase la subsección posterior sobre las calificaciones de estado y calidad como regresores aproximados; la funcionalidad en sí se documenta en el artículo complementario sobre `earthUI` en este mismo número).

La expansión de texto multiplica el conjunto de predictores candidatos, y ello hace necesarias múltiples ejecuciones de `earth/RCA` para pruebas: las columnas expandidas deben probarse dentro y fuera del modelo, consolidarse en conteos o booleanos derivados cuando sean colineales, y verificarse contra el CVR2 para confirmar que portan señal de mercado y no ruido. Trátese cada uno de estos experimentos como su propia carpeta de versión de ejecución, exactamente como en los Protocolos 2.1 y 2.2. Y dado que la propia gramática de análisis sintáctico está en desarrollo, el VE debe revisar las columnas extraídas contra el texto fuente antes de confiar en ellas dentro de un modelo.

5.5. El libro de Excel principal

Dado que la mayoría de los servicios de MLS descargan los datos de los inmuebles a hojas de cálculo CVS de Excel, Excel es probablemente la forma más conveniente de almacenar los datos y los diversos parámetros de configuración. Existe un libro de Excel principal para cada proyecto, llamado `MLS.xlsx`, que tiene una hoja llamada “`MLSData`”, para almacenar los listings de ventas cerradas de inmuebles que se descargaron del MLS, así como varias hojas de configuración y parámetros:

```
Libro: MLS.xlsx/  
├─ Hoja: Project Data  
├─ Hoja: MLSData  
├─ Hoja: Regression Variables  
├─ Hoja: Calculations  
├─ Hoja: Allowed Interactions  
├─ Hoja: Aggregations  
└─ Hoja: (parámetros de ejecución de MARS)
```

Los datos originales se almacenan en un libro con un nombre similar a `MLS_Original.xlsx`. Debe hacerse una copia con un nombre diferente, digamos “`MLS.xlsx`,” que se cargará en el programa RCA de R o Python. Ambos libros se almacenan bajo la carpeta `Data`. Ahora bien, el libro `Data/MLS.xlsx` probablemente pasará por muchas iteraciones de ejecuciones, probando distintas combinaciones de variables, parámetros de MARS y cambios en otras configuraciones. Cada iteración crea una nueva carpeta de entrada/salida nombrada con la fecha y la hora, con el formato `VER_YYYYMMDDHHMMSS`. Cada carpeta tendrá una copia del libro `MLS.xlsx` de entrada y contendrá los libros de Excel generados para cada una de las cuatro etapas de procesamiento (Etapas 2–5), así como los documentos de texto, diagramas y gráficos de salida. Obsérvese que la Etapa 5, la etapa final, puede contener varias hojas adicionales para el informe, según el tipo de cliente y el SOW (alcance del trabajo).

5.6. Copias de seguridad

Al escribir hojas de cálculo de Excel nuevas o modificadas en las carpetas de versión del proyecto, también pueden hacerse copias de seguridad en bases de datos SQLite o PostgreSQL dentro de la carpeta del proyecto. Los archivos de base de datos tienen menos probabilidades de corromperse involuntariamente en períodos posteriores, durante su uso para revisión o consulta.

6. Notación del libro de Excel

6.1. Dimensiones descriptivas

Existen aproximadamente seis dimensiones descriptivas para un libro de Excel “M”:

1. Avalúos “a”
2. Versiones de ejecución “r”
3. Etapas del flujo de trabajo “w”
4. Hojas “s”
5. Listing de propiedades “p”
6. Conjunto de variables “v”

Así, una celda individual de un libro de Excel particular M puede describirse como:

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

(Nota: “*” significa algún índice válido)

o bien:

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

donde:

- n_p = # de filas (depende de la hoja)
- n_v = # de columnas (depende de la hoja)
- n_r = # de versiones de ejecución
- n_m = # de etapas del flujo de trabajo
- n_g = # de proyectos de análisis
- n_h = # de hojas del libro de Excel

Entiéndase que “W” hace referencia al libro de Excel completo, compuesto por seis hojas de datos centrales, cada una con su propia tabla, las cuales, al ejecutarse el programa asociado en R o Python, se cargan en data frames internos separados. Tanto R como Python cuentan con funciones internas capaces de tomar un nombre y determinar a qué fila o columna del data frame pertenece. Por ejemplo, si en el subíndice v_j , $v = (\text{“GLA”, “LotSize”, “BathRms”, “BedRms”, ...})$ y $j=0$, entonces $v_j = \text{“GLA”}$. Tanto R como Python podrán resolver internamente “GLA” hacia la columna en la que está almacenada, que, por ejemplo, podría ser la columna 15 del data frame asociado. Así, en este artículo indexamos nuestros arreglos de nombres, los cuales pasan un nombre al programa en ejecución, que reindexa dicho nombre hacia el índice real de columna o fila del data frame interno. Es un tanto complicado. Pero, para ser precisos, esto es necesario y sumamente conveniente. Se trata de una “separación de responsabilidades” entre el rol del analista y el rol del desarrollador; aunque en la práctica el Valuation Engineer (VE) podría desempeñar ambos roles. Nótese que un “data frame” de R se denomina “dataframe” en Python.

Nótese también que la hoja de parámetros de ejecución de MARS no constituye una dimensión de W, ya que W se relaciona con la versión de los datos que se introducen en MARS, independientemente de los parámetros que determinan cómo MARS/earth procesa dichos datos.

En aras de la claridad, se presenta una notación multidimensional compleja para el libro de Excel, la cual se relaciona de manera más directa con los data frames internos de R o Python en los que se cargan las hojas de datos del libro. Es una notación “pesada”. Pero cuando lleguemos a las demostraciones y demás, descartaremos gran parte de ese lastre mediante supuestos sobre aquello con lo que estamos trabajando. No obstante, esta notación ayuda a obtener desde el inicio una mejor visión general del flujo de trabajo y de los datos.

Comencemos con la notación para el conjunto de libros de Excel que procesaremos.

6.1.1. La dimensión de proyecto (“a”)

Una vez aprobado el alcance del trabajo (SOW), es necesario configurar una carpeta de proyecto para almacenar los datos de entrada y salida, incluidos los parámetros de configuración, los gráficos de salida, los diagramas, las hojas de cálculo y los fragmentos de informe. El VE debe identificar todos los proyectos con un ID secuencial simple, como un número entero, de modo que si un informe llegara a extraviarse, resultaría evidente. Si un proyecto se cancela, debe existir un medio consistente para indicar dentro de la carpeta del proyecto que este ha sido cancelado, pero dicha carpeta nunca debe eliminarse. Suponemos que los ID de proyecto comienzan en 0 y se incrementan de 1 en 1, aunque existen otras posibilidades. Suponemos además que todos los ID de proyecto válidos se almacenan en un conjunto ordenado “a” creado con el programa de procesamiento.

Los datos del proyecto de valuación se encuentran en la hoja de cálculo de proyecto del libro de Excel, pero también podrían almacenarse en alguna base de datos SQL central,

A continuación se sugieren los siguientes campos de proyecto:

1. ID local: Un ID entero corto que se genera secuencialmente para la empresa. Constituiría una excelente clave primaria para el almacenamiento - y podría ser generado automáticamente por la base de datos con cada proyecto sucesivo.
2. GUID global: Si se desea, puede crearse un GUID (Globally Universal ID) mediante

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

3. Ubicación del inmueble (análisis de mercado) o dirección
4. Fecha efectiva del avalúo
5. Fecha del encargo
6. Fecha de inspección
7. Fecha del informe original
8. Fecha del primer informe actualizado
9. Fecha del segundo informe actualizado
10. Fecha del tercer informe actualizado
11. Nombre del cliente
12. Dirección del cliente
13. Nombre del prestamista
14. Dirección del prestamista
15. Nombre del propietario
16. Dirección del propietario
17.

Ejemplo de lista de ID de proyecto:

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

6.1.2. Dimensión de versión de ejecución ("r")

En segundo lugar dentro del flujo de trabajo, ejecutamos la regresión MARS sobre los datos de entrada. Esto probablemente implicará desde varias hasta muchas ejecuciones, a medida que modificamos diversos parámetros para mejorar el modelo. Cada ejecución creará una nueva carpeta de versión para la salida, con un nombre basado en la fecha y hora de ejecución. Por ejemplo,

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

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

Las versiones prometedoras, con modelos y salidas competitivos, deben conservarse hasta que se decida una versión final, o quizá más tiempo si tienen alguna utilidad. Puede resultar conveniente modificar la carpeta de versión para indicar que se trata de la versión final, colocando "Final" al final del nombre.

6.1.3. Etapas del flujo de trabajo ("w")

El flujo de trabajo del RCA se divide habitualmente en cinco etapas separadas, según se enumeran a continuación. Cada etapa debe completarse con éxito antes de iniciar la siguiente. El VE debe poder detener el procesamiento al final de una etapa para revisar los resultados antes de continuar. Esto es necesario porque las zonas de mercado son bastante complejas y, con frecuencia, muy singulares. Pueden presentarse problemas inusuales que exijan un enfoque nuevo. El VE debe poder intervenir en el código de R o Python para realizar los cambios necesarios e incluso probar nuevos enfoques. Y, en tales casos, es posible que desee experimentar y probar diferentes enfoques que requieran múltiples ejecuciones de la misma etapa en la que está trabajando. No obstante, el VE también debe tener la libertad de ejecutar todas las etapas con un solo comando, - con algunas salvedades (p. ej., la agregación de tercer grado debe haberse determinado previamente mediante una ejecución anterior de la Etapa 2). Esto significa que, al completarse cada etapa, esta debe almacenar todos sus datos internos en archivos o bases de datos, de modo que la siguiente etapa pueda iniciarse como si la etapa anterior aún estuviera en memoria, recargando en memoria los datos de la etapa previa.

Estas son las etapas:

```
w = ("stage1", -Configuración
      "stage2", -Procesamiento MARS
      "stage3", -Sales Grid final
      "stage4", -Fragmentos de informe
      "stage5") -Revisión/Limpieza
```

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

6.1.4. Hojas del libro de Excel (“s”)

En el libro de Excel principal (Core) del RCA, contamos con las siguientes hojas, cada una de las cuales se mapea a un data frame interno:

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

La razón por la que las agregaciones de segundo grado se colocan en una hoja separada es que pueden definirse con mayor facilidad mediante una tabla simétrica que muestra todos los pares posibles de interacciones, lo que permite especificar, para cada par, una variable de agregación a la cual se agregan (suman) las contribuciones al valor y los ajustes asociados. No podríamos hacer esto con las agregaciones de tercer grado, porque ello requeriría una tabla simétrica tridimensional, lo cual no es posible en Excel. Además, una tabla de 20x20x20 supondría manejar 8,000 celdas. Resulta más sencillo simplemente listar los nombres de los términos de tercer grado en la tabla junto con las agregaciones de primer grado, por ejemplo:

Variable	Agregar a
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. Listings de propiedades (“p”) - filas de MLSDData

Los listings de propiedades constituyen las filas de la hoja MLSDData. No se encuentran en las demás hojas hasta la Etapa 4, en la que se crean los fragmentos finales del informe con la selección final de inmuebles comparables. A continuación se muestra una lista de ejemplo de ID de propiedades:

```
p = ("MLS823211", "MLS823456", "MLS823542", "MLS823721", "...")
```

6.1.6. Variables MLS (“v”) Columnas de MLSDData

Más adelante se dirá más sobre el manejo de variables. Baste señalar por ahora que la hoja MLSDData contiene típicamente entre 30 y 40 variables de diversos tipos, aunque por lo general solo se seleccionarán de 10 a 20 como variables independientes para la regresión MARS. Estas variables también aparecen en otras hojas, aunque típicamente como una lista de todas las variables posibles que pueden utilizarse para su envío a MARS. Un ejemplo de lista de variables es:

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

6.1.7. Ejemplo final de notación

Reunamos la notación anterior en un ejemplo final. Utilizando las listas anteriores, sea

$${}_{w_1}^{r_1} W_{s_0}^{a_{141}} {}_{v_0}^{p_2} \quad (3)$$

Nótese que aquí utilizamos indexación desde 0, de modo que se supone que el primer elemento de las listas tiene índice 0. Python también utiliza indexación desde 0, pero R utiliza indexación desde 1.

Así, lo anterior significa que nos referimos al valor de la celda correspondiente a:

- La variable v_0 o columna “GLA”
- Para la propiedad p_2 o ID “MLS823542”
- De la hoja s_0 o “MLSDData”
- Para la etapa del flujo de trabajo w_1 o “stage2”
- De la versión r_1 o
“ver20241209_151130”
- Del proyecto a_{141} o “142”

La notación anterior puede entonces escribirse también como:

$${}_{\text{“stage2”}}^{\text{“ver20241209_151130”}} W_{\text{“MLSDData”}}^{\text{“142”}} {}_{\text{“GLA”}}^{\text{“MLS823542”}} \quad (4)$$

6.1.8. Justificación

“¿Por qué no usamos simplemente índices o números en la notación?” Tenemos dos requisitos para nuestra notación:

1. Subconjuntos; p. ej., el analista selecciona con frecuencia subconjuntos de una lista, tales como subconjuntos de variables. Por ejemplo, puede haber 35-40 variables en la hoja MLSDData, pero solo se usan 10-20 de ellas en una ejecución determinada. De una ejecución a la siguiente, el VE cambia con frecuencia las variables que se someten a MARS/earth.

2. Iteración; p. ej., $i = 1, 2, \dots, n_1$

Nuestros algoritmos, protocolos y demostraciones normalmente necesitan poder iterar sobre todos los elementos de una lista ordenada, del primero al último — o bien seleccionar algún elemento en particular.

Así pues, necesitamos una letra que represente el subconjunto o la lista de elementos, y esta debe llevar otro subíndice que podamos usar para iterar sobre los elementos de la lista, de modo que nos proporcione las cadenas de texto que se utilizan para indexar las filas o columnas del data frame.

6.1.9. ¿Matrices?

¿Y qué hay de las matrices del álgebra lineal? Las matrices funcionan únicamente con números, ya sean enteros o reales. Las matrices como estructura de datos están soportadas tanto en R como en Python. Sin embargo, muchas de las variables con las que trabajamos en los datos de la MLS son lo que se denomina “strings” o datos de tipo carácter. Incluso si algunas variables son solo números, a menudo no son números reales sino simplemente nombres. Por ejemplo, los códigos de área de la MLS para una ciudad pueden ser (“660”, “661”, “662”, “663”, $\dots$, “672”). La regresión de earth permite especificar variables como variables “factor”, en cuyo caso los números se tratan simplemente como nombres, sin secuencia alguna. En tales casos, el modelo producido se limitará a proporcionar una contribución al valor para cada número, en lugar de una función que recorra la secuencia. Otros ejemplos incluyen “C1”, $\dots$,”C6” para el estado (Condition) y “Q1”, $\dots$,”Q6” para la calidad (Quality).

6.1.10. Primera simplificación

La notación anterior para el libro de Excel central del proyecto resulta útil para vincular los elementos entre sí, pero es demasiado pesada para llevarla consigo al explicar algoritmos o demostraciones matemáticas. Cuando realizamos el trabajo tedioso, podemos suponer que este se desarrolla dentro del proyecto, la versión, la etapa y la hoja dados. Simplemente indicamos con qué estamos trabajando, si no está ya implícito. Todo lo que realmente necesitamos de lo anterior, una vez que indicamos con qué estamos trabajando, es:

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

donde: n_p = # de propiedades
 n_v = # de variables de entrada
 n_s = # de hojas de entrada

6.1.11. Segunda simplificación

Resulta más conveniente renombrar las hojas según los tipos específicos de data frames que soportan, en lugar de usar índices. En cada hoja, las filas y las columnas tienen funcionalidades distintas. Así, el conjunto W de data frames se descompone en los data frames M, V, A, I y C de la siguiente manera:

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

donde: n_p = # de propiedades
 n_v = # de variables de entrada

Nota: El inmueble a valorar es siempre, por defecto, el primero en las listas de propiedades

En la hoja de cálculo de Excel “MLSDData” y su data frame asociado, la primera fila está siempre reservada para el inmueble a valorar. Nunca se ordena junto con las demás propiedades de la MLS que aparecen debajo. Nunca se introduce en la regresión MARS como dato independiente; se introduce más bien como dato objetivo, o como la variable dependiente. Lo mismo se aplica a las listas de propiedades. En todas las listas de propiedades se asume siempre que el primer elemento, con índice 0, es el ID del inmueble a valorar, salvo que se indique lo contrario.

Incluso si la tarea consiste en crear un modelo para una zona de mercado sin ningún inmueble a valorar, la primera fila se carga de todos modos con datos ficticios de alguna propiedad. Esto hace mucho más conveniente escribir programas, especificar algoritmos o realizar demostraciones. Así, si usamos c para representar la lista de propiedades, c_0 es el inmueble a valorar y $c_i, i = 1, \dots, n_p$ son los listings comparables que se introducirán en la regresión MARS/earth.

◇ **Especificaciones de variables de regresión**

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

donde: n_v = # de variables disponibles
 n_s = # de especificaciones

Obsérvese que $n_s \leq n_v$, ya que normalmente seleccionamos un subconjunto de las variables disponibles como entrada a la regresión.

◇ **Cálculos**

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

donde: n_v = # de variables
 n_c = # de especificaciones de cálculo

◇ **Interacciones**

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

donde: n_f = # de variables de origen
 n_t = # de variables de destino

Obsérvese que el data frame I es un data frame simétrico, con las variables listadas alfabéticamente en las filas “from” (origen) y las columnas “to” (destino). Las celdas se marcan con un “1” para indicar que se permite una interacción entre el par de variables asociado. En caso contrario, un “0” indica que no se permite interacción alguna. Para una interacción de segundo grado, esto es relativamente simple, con una sola interacción posible. Con una regresión de tercer grado, tenemos tres interacciones por pares entre las tres variables, donde cada par de interacciones debe estar permitido para generar el término de interacción (o variable) de las tres variables.

◇ **Agregaciones**

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

donde: n_v = # de variables

7. El Sales Grid del RCA

Téngase presente que el protocolo RCA requiere un programa en R o Python capaz de acceder a R/earth y ejecutar todos los pasos detallados. Es un proceso intensivo en datos y el trabajo no podría realizarse manualmente de ninguna manera. Al mismo tiempo, suponiendo que el programa esté libre de errores, se necesita un programa de software para garantizar que el trabajo requerido se realice con precisión y sin error humano.

También es una necesidad que los VE que realizan este trabajo sean capaces de modificar dichos programas según se requiera, especialmente cuando se enfrentan a zonas de mercado, vecindarios, inmuebles u otros requisitos o problemas nuevos e imprevistos.

Dentro del procesamiento del RCA tenemos tres tipos de Sales Grids:

1. El Sales Grid de Procesamiento del RCA, denominado en este artículo “MLSDData,” que es un data frame de gran tamaño, con su hoja de cálculo asociada, que puede contener numerosos detalles intermedios que finalmente pasan por un proceso de agregación para llegar al Sales Grid del RCA que se menciona más abajo. Un ejemplo sería demasiado grande para mostrarlo aquí. Lo más probable es que los Valuation Engineers adapten esta hoja de cálculo a sus preferencias. Tengo la intención de escribir en el futuro un artículo aparte sobre el desarrollo de un Sales Grid de este tipo.
2. El Sales Grid Final del RCA, que se utiliza en informes no destinados a las GSE y que se asemeja a la [Table 2](#). Este imita, por ejemplo, el formulario de Fannie Mae, lo suficiente como para ofrecer al lector cierta base de familiaridad. Sin embargo, obsérvese que contiene más columnas por comparable de venta que el cuadro de Fannie Mae: cada comparable tiene una columna de “Contribución al valor” junto a sus columnas de “Medida” y “Ajuste,” y el inmueble a valuar también cuenta con una — columnas de las que carece el formulario de Fannie Mae. Un formato más sencillo es también el de la [Table 1](#).
3. El Sales Grid de Carga para las GSE, que se utiliza para cargar el Sales Grid en software como Alamode. Nótese, sin embargo, que no todos los paquetes de software de formularios de avalúo ofrecen la posibilidad de cargar los datos del Sales Grid desde una hoja de cálculo. Por necesidad, dichos grids no pueden incluir una columna separada para las contribuciones al valor, de las cuales las GSE como Fannie Mae no tienen comprensión alguna. La estrategia consiste en entregarles el formulario que exigen, pero respaldarlo con un Sales Grid del RCA, así como con otra documentación de apoyo, tal como los detalles de cualquier agregación de variables.

Los Sales Grids de Procesamiento del RCA no se parecen a los Sales Grids de Fannie Mae, sino que son, aproximadamente, una transposición del Sales Grid de Fannie Mae, con las filas y columnas intercambiadas: cada fila contiene los datos de algún inmueble y cada columna contiene los datos de alguna variable del inmueble. Fácilmente puede haber cientos o miles de filas de inmuebles y 200 o más columnas. La primera fila se reserva para los nombres de las variables, y la segunda para un inmueble a valuar, incluso si solo se está realizando un análisis de mercado. Cuando la hoja de cálculo se lee en un data frame, la primera fila de la hoja se convierte en el atributo names, y la segunda fila, correspondiente al inmueble a valuar, se convierte en la primera fila del data frame.

Las columnas del Sales Grid del RCA incluyen columnas de identificación y, a continuación, las variables primarias, externas y sintéticas, combinadas e introducidas como variables independientes en MARS. A partir de estas variables, MARS produce como salida aquellas que considera que tienen un impacto significativo sobre el precio de venta, también conocido como variable dependiente u objetivo. En la regresión de segundo grado, podemos obtener pares de

estas variables que actúan como nuevas variables. Con la regresión de tercer grado, también podemos obtener tripletes de variables que actúan como supuestas variables independientes, aunque, por supuesto, están altamente correlacionadas con las variables subyacentes.

Cuadro 1: Hoja de cálculo RCA de muestra: disposición vertical

Inmueble a valuar			Comparable 1			Comparable 2		
	Medida	Contribución al valor	Medida	Contribución al valor	Ajuste	Medida	Contribución al valor	Ajuste
Precio de venta				\$1,150,000			\$1,500,000	
Valor base		\$500,000						
Características medidas			Características medidas			Características medidas		
Superficie habitable	1,530	\$463,500	1,350	\$382,500	\$81,000	1,700	\$540,000	-\$76,500
Tamaño del lote	6,000	\$200,000	6,500	\$220,000	-\$20,000	7,000	\$240,000	-\$40,000
Antigüedad	53	\$63,600	57	\$68,400	-\$4,800	51	\$61,200	\$2,400
Baños	2	\$15,000	3	\$30,000	-\$15,000	3	\$30,000	-\$15,000
Dormitorios	3	\$6,000	4	\$9,000	-\$3,000	4	\$9,000	-\$3,000
Total medido		\$1,248,100.00		\$1,209,900.00	\$38,200		\$1,380,200.00	-\$132,100
Características residuales			Características residuales			Características residuales		
Residuo	CQA 5.8	\$30,000	CQA 2.5	-\$59,900	\$89,900	CQA 8.2	\$119,800	-\$89,800
Descomposición del residuo			Descomposición del residuo			Descomposición del residuo		
Diseño		7,000		-\$15,000	\$22,000		\$35,000	-\$28,000
Estado		5,000		-\$8,000	\$13,000		\$20,000	-\$15,000
Calidad		8,000		-\$10,000	\$18,000		\$30,000	-\$22,000
Utilidad funcional		5,000		-\$12,000	\$17,000		\$15,000	-\$10,000
Amenidades		5,000		-\$14,900	\$19,900		\$19,800	-\$14,800
Subtotal		30,000		-\$59,900	\$89,900		119,800	-\$89,800
Total			Total			Total		
Contribución total al valor		\$1,278,100.00			\$1,150,000.00			\$1,500,000.00
Precio de venta ajustado					\$1,278,100.00			\$1,278,100.00

8. Variables

En el lado de entrada de la ejecución de MARS, contamos únicamente con variables simples, que son:

1. Características del inmueble cargadas desde la hoja de cálculo del MLS.
2. Variables derivadas creadas por el VE para refinar las variables del MLS o para añadir transformaciones de variables existentes. Por ejemplo, el VE puede preferir transformar la Date of Sale en “DaysOffMarket”, que es igual a la fecha efectiva del avalúo menos la Date of Sale.
3. El VE puede importar datos de fuentes de terceros, como las tasas de interés hipotecarias. O bien, si faltan datos de latitud, longitud o elevación en los datos del MLS, puede importarlos de otras fuentes. Estas variables podrían denominarse “variables externas”.

En el lado de salida de MARS, trabajamos con un modelo de precios que es una combinación de términos que:

1. Contienen solo una de las variables de entrada y proporcionan la ecuación de la contribución al valor de dicha variable. Estos se denominan términos (o variables de salida) de primer grado. El término mismo se trata como un valor de contribución al valor y se le asigna un nombre.
2. Contienen dos de las variables de entrada y constituyen la ecuación de la contribución al valor de la interacción de las dos variables dadas. Estos términos pueden denominarse términos (o variables de salida) de segundo grado.
3. Contienen tres de las variables de entrada y constituyen la ecuación de la contribución al valor de la interacción entre las tres variables. Estos términos pueden denominarse en este artículo términos (o variables de salida) de tercer grado.

Las variables de salida de MARS son, entonces, contribuciones al valor, incluida la contribución al valor residual. A su vez, se tratan como variables en los cálculos del Sales Grid, tales como el cálculo de ajustes o la creación de gráficos.

Podemos tener distintos tipos de variables que se encuadran en dos ontologías básicas: (1) el origen de la variable y (2) la aplicación de la variable, es decir, no residual frente a residual

8.1. Ontología de origen de las variables

1. Variables de primer grado
 - 1.1 Variables de regresión del MLS
 - 1.2 Variables de regresión externas
 - 1.3 Variables de regresión derivadas

2. Variables factor:

Para manejar las variables factor, se necesita una función capaz de identificarlas. Dado que se especifican en el libro de Excel de configuración del proyecto, se crea una función que devuelve verdadero si la variable que se le pasa ha sido clasificada como de tipo factor.

3. Variables de interacción: Estas son creadas por MARS al generar el modelo, si el grado se establece en 2 o superior. Sin embargo, 3 variables en interacción es lo máximo que podría recomendarse debido a la dificultad de interpretación, la probabilidad de sobreajuste y el impacto sobre los grados de libertad, con el consiguiente aumento de los registros de inmuebles necesarios para garantizar un modelo robusto.
 - 3.1 Variables de segundo grado: Estas son comunes y buenos ejemplos serían GLA:BathRms, GLA:LotSize, DateOfSale:BathRms, Latitude _Longitude.
 - 3.2 Variables de tercer grado: Interacciones menos comunes que sugieren cautela y posible sobreajuste.
4. Variables de descomposición del residuo

8.2. Ontología de aplicación de las variables

1. Contribuciones al valor generadas por MARS
 - 1.1 Variables de primer grado
 - 1.1.1. Variables de regresión del MLS
 - 1.1.2. Variables de regresión externas
 - 1.1.3. Variables de regresión derivadas
 - 1.2 Variables de interacción
 - 1.2.1. Variables de segundo grado
 - 1.2.2. Variables de tercer grado
 - 1.3 Variables de descomposición del residuo
 - 1.4 Variables de agregación
2. Variables & valores generados por el ingeniero
 - 2.1 Variables de descomposición del residuo
 - 2.1.1. Diseño/Estética
 - 2.1.2. Paisajismo
 - 2.1.3. Utilidad funcional
 - 2.1.4. Vista
 - 2.1.5. Estado
 - 2.1.6. Calidad
 - 2.1.7. ...

A estos grupos de variables se les asignará su notación y se describirán con mayor detalle en la siguiente sección

Protocolo 3.1: Variables factor

Las variables factor, o simplemente factores, pueden aparecer en todos los términos, incluidas las interacciones. Son importantes porque deben recibir un tratamiento especial en la regresión MARS y en la elaboración de gráficos.

Si bien las variables con valores de tipo carácter son necesariamente factores, las variables numéricas deben marcarse como variables factor para que MARS las reconozca como tales. Por ejemplo, las zonas del MLS suelen identificarse mediante un número. En particular, las zonas de vecindario en el San Francisco Bay Area se identifican a menudo con enteros únicos. Si estos valores se pasan a MARS como factores, MARS intentará descubrir si ciertas zonas tienen un valor contributivo promedio sobre las ventas de inmuebles dentro de sus límites, sin considerar relación alguna con otras zonas. En cambio, con las variables normales, asumirá que existe algún tipo de continuidad en las contribuciones al valor a medida que cambia el valor de la variable. Esta es una distinción importante. Además, estas variables factor pueden aparecer en términos del modelo junto con variables no factor, lo que dificulta la elaboración de gráficos. En concreto, al graficar términos de una sola variable con factores, agrupamos todos esos factores y simplemente dibujamos un gráfico de barras que muestra sus contribuciones, positivas o negativas.

Si los factores aparecen con otras variables en los términos, siempre aparecerán con un único valor posible, porque las variables factor solo pueden tener un valor para un inmueble dado: si una zona tiene el valor 662, por definición no puede tener ningún otro valor para esa zona. Si esto le resulta confuso, para complicar más las cosas, en el modelo generado por MARS no existe un único nombre de variable para el factor; MARS tomará el nombre raíz de la variable factor que recibió como entrada y luego creará nuevas columnas con nombres de variable formados al anexar los valores posibles. Así, una sola columna en la hoja de cálculo de Excel, como `MlsArea`, tendrá múltiples columnas del data frame creadas para cada valor posible, como `MlsArea660`, `MlsArea661`, etc. Una variable como `MlsArea661` tendrá el valor 1 para los inmuebles en la zona 661; de lo contrario, 0. Después, MARS efectuará la regresión sobre esas columnas como si representaran variables booleanas. Estos nombres modificados se utilizarán en el modelo proporcionado por MARS.

Probablemente querrá convertir esos nombres del modelo de vuelta a su nombre original para que los gráficos resulten más comprensibles para los clientes, a quienes les gusta lo claro y sencillo (al menos tan sencillo como sea posible). En lugar de ver `MlsArea661=1`, veremos `MlsArea=661`. Entonces podemos graficar los términos indicando simplemente en el encabezado del gráfico algo como “Para MLS Area = 662”. De este modo, los términos de grado N con M variables factor se grafican como términos de grado $(N-M)$, y la dimensión del factor desaparece en el título.

Obsérvese que dos o más variables factor podrían existir en un mismo término, por lo que probablemente se necesitaría un gráfico para cada par posible de valores. Si la variable x , con tres valores distintos, y la variable y , con dos valores distintos, aparecen en el mismo término con la variable no factor z , entonces se necesitarían $2 \cdot 3 = 6$ gráficos. Un ejemplo sería “`MlsArea`” + “`Architectural Style`”; p. ej., las casas victorianas en San Francisco podrían valer más o menos según la zona de San Francisco en la que se encuentren. Afortunadamente, en la mayoría de las zonas suele haber solo 1-3 estilos arquitectónicos lo suficientemente importantes como contribuyentes al valor para ser utilizados en el modelo (aunque puede haber 6+ estilos arquitectónicos diferentes).

8.3. Términos de segundo y tercer grado

Con la regresión de segundo o tercer grado, algunos de los términos del modelo de salida pueden incluir dos o, respectivamente, tres variables. Por ejemplo:

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

Obsérvese que estos términos se nombran ordenando alfabéticamente los nombres de las variables y concatenándolos, quizás con un guion bajo. Hacemos lo mismo con los términos de tercer grado sobre las tripletas de nombres de variables producidas en el modelo. Por ejemplo, “BathRms_GLA_LotSz.”

Estas variables se añaden a las variables de entrada primarias en la Etapa 3, cuando el programa RCA calcula las contribuciones de todas las variables para cada inmueble, incluidas las variables de interacción.

8.3.1. Variables de ajuste

El programa RCA también calculará los ajustes para todas las variables e inmuebles en la Etapa 3, después de calcular las contribuciones al valor, restando la contribución al valor del inmueble comparable correspondiente a una característica de la contribución del inmueble a valuar para esa característica.

8.3.2. Variables de agregación

Las variables de agregación se definen para reemplazar subconjuntos de las variables, sumando todas sus contribuciones al valor y ajustes en una variable común que se utilizará en el informe. Por ejemplo, podríamos haber dividido tanto el área habitable terminada como la no terminada en variables de mayor detalle:

1. Área terminada legal sobre rasante
2. Área terminada legal bajo rasante
3. Área terminada ilegal sobre rasante
4. Área terminada ilegal bajo rasante
5. Área no terminada sobre rasante
6. Área no terminada bajo rasante

Sí, hay zonas donde esto tiene sentido.

La regresión MARS efectuará la regresión sobre todas las variables y generará un modelo que proporcionará contribuciones y ajustes para ellas. Pero para un informe en formulario, estas distintas contribuciones probablemente tendrán que combinarse o agregarse en menos variables. Los detalles de la agregación deben proporcionarse en un anexo. Un ejemplo más frecuente sería agregar los ajustes de interacción en uno de los ajustes de las variables primarias. Por ejemplo, podríamos agregar el ajuste de GLA_LotSize a GLA o a LotSize, o incluso dividirlo 50:50 entre ambos, según lo determinado en la hoja de cálculo Aggregation.

8.3.3. Variables especiales & notación

π_i	Precio de venta neto real del inmueble $i = 1, 2, \dots, n_p$; para el inmueble a valuar, π_0 denota la conclusión de valor (véase la Demostración I)
α_i	Precio de venta neto ajustado del inmueble $i = 1, 2, \dots, n_p$
ρ_i	Precio de venta neto estimado por MARS para el inmueble $i = 0, 1, \dots, n_p$
β	Valor base del modelo
$\varphi_{i,j}$	Contribución al valor de la variable de regresión $j = 1, 2, \dots, n_r$ para el inmueble $i = 0, 1, \dots, n_p$
$\delta_{i,j}$	Ajuste de la variable de regresión $j = 1, 2, \dots, n_r$ para el inmueble $i = 1, 2, \dots, n_p$
ϵ_i	Residuo total del inmueble $i = 0, 1, \dots, n_p$
$\epsilon_{i,c}$	Contribución al valor del componente residual c , $c = 1, 2, \dots, n_c$ para el inmueble $i = 0, 1, \dots, n_p$
$\xi_{i,c}$	Ajuste del componente residual c , $c = 1, 2, \dots, n_c$ para el inmueble $i = 0, 1, \dots, n_p$

donde

$n_p = \#$ of comparable properties

$n_r = \#$ of regression variables

$n_c = \#$ of residual breakdown variables

8.3.4. Descomposición de los residuos

Una vez que la regresión MARS ha creado un modelo, este genera los precios de venta estimados a partir de las variables de entrada seleccionadas por MARS. Resta ese valor del Net Sale Price correspondiente para obtener un valor residual para cada inmueble. Ahora bien, el protocolo RCA estipula que, una vez que el VE ha decidido los comparables definitivos, por lo general 6-12 listings de inmuebles, debe revisar cada listing y descomponer el residuo calculado en las características más probables (variables latentes) que, a su juicio, impactan el valor, ignorando la posibilidad de errores de datos aleatorios. Podría, por ejemplo, tener estas variables de

descomposición para un inmueble dado:

1. Diseño
2. Utilidad funcional
3. Venta por encima/por debajo del mercado
4. Estado
5. Calidad

Es probable que distintos inmuebles tengan algunas variables de descomposición del residuo diferentes, las cuales deben reunirse en un único conjunto ordenado.

Estimación de las contribuciones de características infrecuentes. Algunas características portadoras de valor son demasiado infrecuentes para el canal de regresión: un taller presente en 3 de 780 ventas no puede obtener un coeficiente defendible, y el pasadizo hacia adelante normalmente omitirá por completo dicha variable. Estas son características residuales, y sus contribuciones pueden estimarse *en el espacio de los residuos*: se ajusta el modelo sin la característica y luego se efectúa la regresión de los residuos del modelo sobre la(s) columna(s) excluida(s) mediante mínimos cuadrados ordinarios. Para un indicador binario, el coeficiente es una contribución de suma global (equivalentemente, el desplazamiento medio del residuo de las ventas marcadas); para una columna numérica, es una contribución por unidad. Cuando varias amenidades correlacionadas se estiman *conjuntamente* (p. ej., la superficie de las construcciones auxiliares junto con un indicador de taller), la regresión reparte el crédito entre ellas en lugar de contar dos veces los mismos dólares en dos filas de la descomposición. Las estimaciones respaldadas por menos de unas diez observaciones son solo indicativas y deben corroborarse mediante el costo depreciado o una muestra ampliada (p. ej., indicadores recuperados de los comentarios públicos); la propia restricción residual acota cualquier valor que finalmente se asigne. earthUI (v0.11+) implementa este método directamente como la pestaña *Residual Contribution*, incluidas la estimación conjunta, la evidencia de residuos por venta y una nota de documentación para el expediente de trabajo.

Notación: Utilizamos la notación de la tabla anterior para llegar a:

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

donde n_c es el número de componentes de la descomposición del residuo.

El VE debe emplear su juicio para descomponer el residuo en los componentes dados. Podría pensarse que esto permitiría que el sesgo se introdujera en la conclusión de valor final. Sin embargo, no es así, porque contamos con una demostración matemática de que, si los ajustes de la descomposición suman el residuo, no afectarán la conclusión de valor. El propósito de la descomposición no es establecer el valor, sino explicar por qué el residuo es el que es en comparación con los residuos de venta de otros inmuebles. El único punto donde el sesgo del ingeniero puede afectar la conclusión de valor final es la estimación del residuo del inmueble a valuar, dado que, por supuesto, no existe un precio de venta real del inmueble a valuar con el cual calcular su residuo.

Cada inmueble tiene su conjunto de contribuciones al valor residuales, pero estas deben fusionarse en un único conjunto C sobre todos los inmuebles:

Definiendo: $A_i = \{\{\epsilon_{i,c}\}_{c=1}^{n_c}\}$

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

donde $[\epsilon_{i,c}]$ es el nombre de variable para el valor $\epsilon_{i,c}$, A_i es el conjunto de dichas variables y n_c es el número de valores en cada ϵ_i .

Así, con los datos del inmueble a valorar y de los comparables en columnas separadas, debajo de todas las contribuciones al valor de las variables del modelo de regresión para cada inmueble, deberíamos ver el conjunto C de n_c contribuciones al valor residuales correspondientes.

9. Demostraciones fundamentales del RCA

Sea:

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

Demostración I: Todos y cada uno de los precios de venta ajustados de los comparables son iguales al precio de venta estimado del inmueble a valorar.

Para cualquier comparable i , el precio de venta ajustado α_i es la suma del precio de venta neto π_i más los ajustes: las diferencias entre las contribuciones al valor del inmueble a valorar y las del comparable para el valor base, las variables de regresión $j = 1, 2, \dots, n_r$ y el residuo. Para el inmueble a valorar, que carece de una venta real, el precio de venta neto π_0 se define como la conclusión de valor $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$. Entonces:

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

Demostración II: Alterar la descomposición del residuo, bajo la restricción de que todas las contribuciones al valor de los componentes residuales sumen el residuo del comparable, no tiene ningún impacto en la conclusión de valor final del protocolo RCA.

Esto es casi trivial. Sin embargo, si no se reflexiona sobre la matemática, resulta fácil pasarlo por alto, porque los valuadores tradicionales se concentran en los ajustes en lugar de en las contribuciones al valor. Son las contribuciones al valor las que se cancelan aquí — una omisión de décadas. Para algún comparable k , $k > 0$, el ajuste del residuo total es igual a la suma de los n_c componentes residuales de la descomposición, que a su vez es igual a la suma de las diferencias de las contribuciones al valor, la cual es igual al residuo total del inmueble a valorar menos el residuo total del comparable. Y esto es independiente del número de componentes residuales

creados o de lo que representen. Mientras el valor de los componentes residuales sea igual al residuo total tanto para el inmueble a valorar como para el comparable elegido, ξ_k no cambiará.

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

10. Modelo MARS

10.1. Funciones “ g ” de Variables del Modelo y de Interacción

Un principio fundamental del método RCA es que el “razonamiento explicativo” reviste la máxima importancia. Lamentablemente, proporcionar gráficas significativas para las diversas ecuaciones de contribución al valor, en especial cuando se emplean interacciones y variables factor, puede resultar difícil. Por ello se considera imperativo establecer un protocolo claro y estricto para la creación de estas gráficas. Eso se hace aquí, pero debe entenderse que el protocolo está dirigido al desarrollador que escribe el código en R o Python para generar las gráficas, y podría ser de interés solo superficial para quienes no participan en la programación.

El modelo de precios generado por MARS es la suma de una constante base más términos adicionales, cada uno de los cuales involucra de 1 a 3 variables. Para analizarlo, primero se divide el modelo en términos individuales. Luego, los múltiples términos que comparten el mismo conjunto de variables, y que representan una sola interacción, se combinan en expresiones de interacción únicas. Supóngase que el grado máximo de regresión es 3. Considérese un modelo de regresión de tercer grado compuesto por una constante base, n_1 expresiones de primer grado, n_2 expresiones de interacción de segundo grado y n_3 expresiones de interacción de tercer grado. Esto nos da cuatro grupos distintos de expresiones, que son de una sola variable o interacciones de múltiples variables, cada uno graficado de manera diferente según dos factores: (1) si incluyen una variable factor (una variable categórica), y (2) el número de variables únicas que contienen. Específicamente:

- Las expresiones de primer grado (1 variable) se grafican en 2D.
- Las expresiones de segundo grado (2 variables) se grafican en 3D.
- Las expresiones de tercer grado (3 variables) se grafican como un *conjunto* de gráficas 3D.
- Si una expresión contiene una variable factor, cada factor único reduce en uno la dimensionalidad de la gráfica (p. ej., una gráfica 3D se convierte en 2D si una de las variables es un factor).

La indexación de cada función g proporciona la información necesaria para determinar cómo debe graficarse, reflejando el número de variables únicas y la presencia de variables factor.

10.2. Indexación de las Funciones g

Cada función g se indexa como $f_{j,k}g_k^j$, donde:

- $f_{j,k}$ es el primer índice (un superíndice antes de g):
 - j va de 0 a 3 e identifica el grupo:
 - $j = 0$: la constante base (una excepción a la indexación habitual).
 - $j = 1$: expresiones de primer grado.
 - $j = 2$: expresiones de segundo grado.
 - $j = 3$: expresiones de tercer grado.
 - k es el índice de una función específica dentro de su grupo (p. ej., $k = 1$ para la primera función del grupo j).
 - El valor de $f_{j,k}$ indica el número de variables factor únicas en la función:

- $f_{j,k} = 0$: sin variables factor.
- $f_{j,k} > 0$: el número de variables factor (p. ej., 1, 2 o 3).
- Índices posteriores en g :
 - El superíndice j después de g (p. ej., g^j) repite el número de grupo para mayor claridad.
 - El subíndice k después de g (p. ej., g_k) coincide con el k de $f_{j,k}$, identificando la función dentro de su grupo.

Si bien los índices de f y g coinciden numéricamente, sus funciones difieren: $f_{j,k}$ cuenta las variables factor, lo cual es fundamental para las decisiones de graficación, mientras que j y k en g definen su grupo y su posición.

10.3. Dimensiones de Graficación

La dimensión d de la gráfica para una función $f_{j,k} g_k^j$ se calcula como:

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

donde:

- j es el número de variables únicas (o número de grupo, sin considerar el caso de la constante).
- $f_{j,k}$ es el número de variables factor.
- d es el número de variables continuas (no factor) por graficar, con un máximo de 3 (dado que el grado máximo es 3). Una función con d variables continuas se dibuja en $d + 1$ dimensiones: $d = 1$ da una gráfica 2D, $d = 2$ una gráfica 3D (o un mapa de calor 2D), y $d = 3$ un *conjunto* de gráficas 3D.

Por ejemplo:

- Si $j = 3$ (3 variables) y $f_{j,k} = 0$ (sin variables factor), entonces $d = 3 - 0 = 3$, lo que requiere un conjunto de gráficas 3D, cada una fijando una de las variables.
- Si $j = 3$ y $f_{j,k} = 1$ (una variable factor), entonces $d = 3 - 1 = 2$, lo que requiere una gráfica 3D (o un mapa de calor 2D) para cada valor de la variable factor que aparece en el modelo.
- Si $j = 0$ (la constante), normalmente no se necesita graficación, pues se trata de un solo valor.

Cuando $d = 3$ (p. ej., una expresión de tercer grado sin variables factor), visualizar la función completa puede requerir múltiples gráficas 3D, cada una fijando una variable en distintos valores para mostrar un subconjunto del comportamiento de la función.

10.4. La Ecuación Final del Modelo

El modelo MARS se descompone en cuatro grupos de expresiones (constante, de primer, segundo y tercer grado). La indexación $f_{j,k} g_k^j$ nos indica:

- j : el grupo y el número de variables.
- $f_{j,k}$: el número de variables factor.

- $d = j - f_{j,k}$: la dimensión de la gráfica.

Esta estructura guía la forma en que se visualiza cada función g , garantizando que el método de graficación corresponda a la complejidad de la expresión y a los tipos de variables.

La estimación del precio de venta $\hat{P}$ proporcionada por el modelo MARS puede expresarse como:

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

$n_1 = \#$ of first degree terms
 $n_2 = \#$ of second degree terms
 $n_3 = \#$ of third degree terms

10.5. Número de Gráficas Necesarias para Graficar un Modelo

¿Cuántas gráficas se requieren para visualizar todas las funciones de un modelo? La respuesta depende de las variables, de sus interacciones y de si son variables factor o no factor.

El VE puede limitar las variables y las interacciones en el archivo de configuración de Excel para R/earth, lo que influye en qué interacciones se permiten. Un enfoque práctico consiste en dejar que R/earth evalúe todas las interacciones posibles y seleccione aquellas con contribuciones significativas a la estimación del precio de venta, verificando al mismo tiempo problemas como el sobreajuste o la colinealidad. Algunas interacciones pueden quedar excluidas, lo cual afecta a las interacciones de grado superior (p. ej., si se excluye un par en una interacción de tercer grado, se excluye la combinación completa).

Para las variables factor (p. ej., MLS Area con 12 valores interactuando con GLA), la graficación se vuelve más compleja. Esto podría requerir 12 gráficas—una por zona. Si otro factor como Architectural Style (6 valores) interactúa con MLS Area y GLA, podrían necesitarse hasta $12 \cdot 6 = 60$ gráficas. Sin embargo, R/earth rara vez encuentra significativas todas las combinaciones; por lo general, solo unas pocas interacciones clave (p. ej., zonas y estilos específicos) ameritan graficarse, según lo identifica el propio modelo.

He aquí una guía simplificada del número de gráficas necesarias:

1. **Funciones de una sola variable (no factor):** 1 gráfica.
2. **Funciones de una sola variable (factor):** 1 gráfica de barras que combina todos los valores del factor.
3. **Función base:** No se necesita gráfica.
4. **Interacciones de 2.º grado (sin factores):** 1 gráfica (p. ej., mapa de calor 2D con colores para el precio de venta).
5. **Interacciones de 2.º grado (1 factor):** 1 gráfica por cada valor del factor en el modelo (p. ej., 2 valores = 2 gráficas).
6. **Interacciones de 2.º grado (2 factores):** 1 gráfica con una matriz de combinaciones de factores (p. ej., 3 zonas $\times$ 2 estilos = 6 celdas).
7. **Interacciones de 3.º grado (sin factores):** Múltiples gráficas (p. ej., 4 gráficas para el número de baños = 1, 2, 3, 4).

8. **Interacciones de 3.º grado (1 factor):** 1 gráfica por cada valor del factor en el modelo.
9. **Interacciones de 3.º grado (2 factores):** 1 gráfica por cada combinación de valores de los factores en el modelo.
10. **Interacciones de 3.º grado (3 factores):** 1 gráfica por cada valor del factor con el menor número de valores distintos, usando una matriz de colores para los otros dos.

Escenarios más complejos pueden requerir técnicas de graficación avanzadas.

Existen otras posibilidades con gráficas probablemente más sofisticadas.

Un ejemplo de un modelo completo generado por MARS se encuentra en [Figure 1](#).

11. Características de los datos

La siguiente sección analiza MARS. MARS depende por completo de los datos que recibe, y la mayor parte de estos datos provendrá de algún MLS (Multiple Listing Service) o de los Tax Assessors. Sin embargo, pueden pasar por otros proveedores de datos que los mantienen y mejoran para luego venderlos con cierta ganancia. En algunas zonas son bastante confiables; en otras, no tanto. En casos críticos, el inspector del inmueble debe realizar sus propias observaciones y mediciones, cuya exactitud puede cuestionarse.

El Valuation Engineer necesita comprender qué representan los datos y qué tan bien cumplen la función de representar lo que se supone que representan. Esto plantea toda una lista de cuestiones:

- ¿Cómo se clasifica una variable como una medición?
- Si una variable no es una medición, ¿qué otra cosa puede ser?
- ¿Cómo puede ser útil la regresión si los datos que recibe no son exactos?

La mayor parte de los datos que obtenemos de los servicios de listings de la MLS y de los datos del Tax Assessor para la regresión son datos numéricos que implican mediciones o conteos. Una parte es lógica, verdadero o falso.

11.1. Propósito del residuo

El residuo sirve como un indicador indirecto del valor aportado por las variables no incluidas en el modelo de regresión. En el lenguaje de la literatura hedónica, estas características no medidas son *variables latentes*: variables que influyen en el precio de venta pero que no se miden directamente, en la mayoría de los casos porque no existe una forma práctica de medirlas objetivamente. Sus valores deben, en cambio, inferirse y, en la práctica, el valor de una variable latente se estima a menudo mediante juicio subjetivo — en RCA, un juicio anclado en el ranking de residuos. Al ordenar los inmuebles según su residuo o su residuo por pie cuadrado, obtenemos un ranking objetivo que va desde los inmuebles con menor atractivo hasta los de mayor atractivo. Este método de clasificación es eficaz, siempre que un analista competente con un conocimiento profundo del mercado local construya el modelo MARS (Multivariate Adaptive Regression Splines). Los siguientes son algunos requisitos adicionales:

- Calidad de los datos y del modelo: Un modelo de regresión bien construido y datos de alta calidad son requisitos previos para obtener residuos confiables. En la San Francisco Bay Area, mi objetivo es desarrollar un modelo con un R^2 de aproximadamente 75-80 % y un R^2 de validación cruzada (CV R^2) de 55-70 %.
- Evitar el sobreajuste, limitar el R^2 : El Valuation Engineer debe aspirar a un R^2 máximo de alrededor del 80 % en la mayoría de los mercados maduros de zonas de precios más altos, como la SF Bay Area, porque gran parte del valor competitivo de un inmueble residencial reside en su atractivo subjetivo, que no es cuantificable y, por lo tanto, no puede incluirse directamente en los modelos de regresión. Un R^2 significativamente superior al 80 % podría indicar un modelo sobreajustado, lo que exige un examen más detenido. No obstante, una advertencia: existen zonas de mercado con valores de R^2 superiores al 90 % sin sobreajuste, y hay zonas difíciles de valorar donde lo mejor que puede lograrse es un 50-60 % o menos.
- R^2 alto en zonas específicas: Sin embargo, hay regiones, como fraccionamientos u otras zonas residenciales uniformes, donde valores más altos de R^2 son sorprendentemente comunes debido a la homogeneidad de los inmuebles.

11.2. Las calificaciones de estado y calidad como regresores aproximados

Las escalas de calificación de estado y calidad de Fannie Mae (C1–C6 y Q1–Q6) merecen un comentario especial. Estas escalas son demasiado burdas para tratarse como verdaderas características medidas y, dado que las asigna el valuador, son en la práctica calificaciones subjetivas — variables latentes por derecho propio. Resulta útil distinguir explícitamente dos niveles: la calificación burda del valuador utilizada como regresor, *estado de grano grueso* (condition-1) y *calidad de grano grueso* (quality-1), frente a las variables latentes de grano fino recuperadas del residuo, *estado de grano fino* (condition-2) y *calidad de grano fino* (quality-2). El VE puede aceptar las calificaciones de grano grueso del valuador como aproximaciones y efectuar la regresión sobre ellas, obteniendo contribuciones al valor aproximadas que resultan útiles como asignaciones de valor iniciales para las variables latentes de grano fino. Cuando así se procede, el protocolo consiste en restar de la estimación del modelo las contribuciones al valor estimadas por la regresión para estas dos variables de grano grueso — retirándolas de los atributos medidos en la sección de regresión del Sales Grid — y sumarlas de nuevo al residuo. A partir de ahí, se retiran posteriormente del residuo y se asignan a las variables latentes de grano fino correspondientes, `res_condition1` y `res_quality1`. La información adicional extraída de los comentarios públicos de la MLS (`public_remarks`) y de listas de cadenas relacionadas se utiliza luego para realizar microrrefinamientos de estos valores de las variables latentes. El resultado final es que las contribuciones al valor de grano fino reemplazan a las contribuciones al valor de grano grueso: las calificaciones burdas sirven únicamente como andamiaje para el modelo aproximado, y los valores que finalmente se reportan para estado y calidad son los de grano fino, derivados del residuo.

11.3. Errores de datos

Según comentarios en redes sociales, los datos de la Multiple Listing Service (MLS) de California son, en general, de mayor calidad que los de muchas otras regiones de Estados Unidos. Típicamente, cabe esperar que los datos de la MLS sean de buena calidad en la mayoría de las áreas metropolitanas maduras, pero de menor calidad en las zonas rurales.

Si bien ocasionalmente hay errores en los datos por diversas razones, estos errores en general deberían cancelarse entre sí y tener un impacto mínimo en los promedios utilizados en los modelos de precios si existen datos suficientes.

Sin embargo, pueden producirse sesgos sistemáticos en zonas específicas. Por ejemplo, considérese un vecindario donde muchas viviendas se construyeron inicialmente con un segundo piso completo, pero a algunos compradores se les dio la opción de eliminar habitaciones del segundo piso para crear techos abovedados en una o más habitaciones de la planta baja, lo que reduce significativamente la superficie habitable bruta (GLA). Por una u otra razón, los cambios no se actualizaron posteriormente en los registros del Tax Assessor. En consecuencia, la MLS podría mostrar 2,400 pies cuadrados cuando el valor medido es de solo 2,000 pies cuadrados. En tales escenarios, los datos de la MLS deben corregirse antes de ejecutar la regresión MARS (Multivariate Adaptive Regression Splines). Es importante señalar que obtener esta información para los comparables puede requerir cierto esfuerzo de investigación. Las discrepancias deben anotarse para referencia futura.

Dado lo anterior, es razonable suponer que, en zonas con MLS de alta calidad, los errores que se filtran al residuo son mínimos en comparación con el impacto de características subjetivas y no medidas como el diseño, el estado y la calidad. Esta es una consideración importante al estimar la exactitud del método RCA. Por ejemplo, si el valor de R^2 para una zona de mercado dada es del 80 %, entonces aproximadamente el 20 % del precio se atribuye al residuo CQA. Un error de $\pm 10\%$ al ubicar el inmueble a valuar dentro del ranking de residuos se propaga, por lo tanto,

como un error de $\pm 10\% \cdot 20\% = \pm 2\%$ en la conclusión de valor final. ¡Así es como se mejora la exactitud! En muchos casos son posibles estimaciones de valor exactas muy por debajo de $\pm 1\%$.

Por lo tanto, tratar los residuos como indicadores indirectos del valor colectivo de las variables no modeladas en la regresión MARS es razonable, siempre que el Valuation Engineer cuente con suficiente experiencia en el mercado local para identificar dónde y por qué existen sesgos sistémicos de medición. Este enfoque asegura que el análisis refleje con exactitud los valores reales de los inmuebles, reconociendo las limitaciones y los matices de los datos inmobiliarios.

Ejemplo desarrollado: medir determinantes del valor no medidos en Sea Ranch

Cuando el R^2 alcanzable en una zona de mercado es más bajo de lo que la aparente homogeneidad del mercado haría esperar, la respuesta del marco consiste en identificar y medir la variable faltante en lugar de aceptar la limitación. La comunidad de Sea Ranch, en el condado de Sonoma, lo ilustra directamente.

Un caso concreto de la respuesta de “medir, no aceptar”. La comunidad de Sea Ranch, en el condado de Sonoma, es un desarrollo costero donde la distancia al océano es un determinante del valor primario: una vivienda a varios cientos de pies del acantilado alcanza un precio sustancialmente distinto del de una situada a un cuarto de milla tierra adentro, manteniendo constantes las demás características. Los datos de la MLS para la comunidad no registran la distancia al océano como variable. Un modelo MARS inicial ajustado con las variables disponibles de la MLS produjo un poder explicativo menor del que la aparente homogeneidad del mercado debería haber sustentado, con una estructura de residuos que no ordenaba suficientemente los inmuebles según ninguna característica visible en los datos — una señal de que faltaba un determinante importante en el modelo, y no de que el mercado fuera genuinamente ruidoso.

El remedio fue la medición directa. El autor obtuvo un mapa de la comunidad que mostraba la disposición de los inmuebles respecto de la línea de costa y midió la distancia al océano de varios cientos de viviendas. Las distancias medidas se agregaron como variable al conjunto de datos de la MLS y se volvió a ejecutar la regresión MARS. La nueva variable fue seleccionada por el modelo como significativa, el poder explicativo aumentó sustancialmente y el ranking de residuos pasó a ordenar los inmuebles según patrones consistentes con las características subjetivas restantes — vista, estado, calidad — en lugar de la señal latente de distancia al océano que previamente lo había contaminado. La elevación se ha abordado de manera similar en vecindarios de ladera donde la variable de la MLS falta o no es confiable.

El principio metodológico es general. Cuando el R^2 alcanzable cae por debajo de lo que la aparente homogeneidad del mercado debería sustentar, la primera hipótesis es que falta una variable importante en el modelo, y la respuesta apropiada es identificar cuál es y recolectar las mediciones directamente. Las fuentes incluyen mapas comunitarios, datos GIS, registros del Tax Assessor, registros del departamento de construcción, imágenes satelitales e inspección de campo. El costo de este trabajo de medición es real pero acotado; una vez recolectada la variable para una zona de mercado dada, puede trasladarse a encargos futuros en esa zona. RCA absorbe nuevas variables limpiamente porque el marco es indiferente a qué variables entran en la regresión MARS — la estructura de residuos indica al VE si la nueva variable captura señal real, y las contribuciones al valor se ajustan en consecuencia.

La respuesta alternativa — aceptar un R^2 bajo y reportar bandas de error más amplias — es defendible cuando los determinantes del valor faltantes son genuinamente inmedibles (preferencias específicas del comprador, sentimiento de mercado transitorio, circunstancias idiosincrásicas de la transacción), pero no cuando simplemente no se han recolectado. El primer paso del VE ante un modelo débil debe ser preguntarse si la señal faltante corresponde a lo segundo.

11.4. ¿Qué es la exactitud?

La exactitud perfecta de una valuación requiere:

1. Conformidad con el valor de mercado: La transacción de venta hipotética del inmueble a valuar se ajusta al 100 % a la definición dada de valor de mercado
2. Desarrollo de un modelo exacto y robusto: Se desarrolla un modelo MARS de alta calidad con un R^2 de alrededor de 75-80 % y un CVR2 de aproximadamente 55-70 %, según la ubicación.
3. Buen análisis de errores de datos: Todos los errores de datos probables de las variables medidas se comprenden y están contenidos, con poca filtración hacia los residuos.
4. Contabilización del sesgo sistémico: Cualquier sesgo sistémico en los datos debe contabilizarse y ajustarse.
5. Análisis del ranking de residuos: Asegurarse de que los inmuebles del ranking de residuos sigan un patrón razonable de características desde el rango más bajo hasta el más alto. Las anomalías, como ventas testamentarias, ventas cortas (short sales) y ventas en subasta, deben investigarse, explicarse y, potencialmente, eliminarse.
6. Ubicación del inmueble a valuar: Con una comprensión profunda del ranking de residuos, el inmueble a valuar debe posicionarse con exactitud. Debe identificarse un inmueble con un atractivo significativamente mayor y otro con un atractivo significativamente menor, y el inmueble a valuar debe ubicarse entre ellos y puntuarse. Luego se intenta estrechar el ranking entre el inmueble superior y el inferior tanto como sea posible.
7. Razonamiento explicativo: Nunca debe sacrificarse el razonamiento explicativo en aras de la exactitud. El valuador debe poder explicar las diferencias de valor entre el inmueble a valuar y los inmuebles comparables, al menos a una persona de inteligencia razonable. Por ejemplo, un buen modelo MARS con un R^2 y un CVR2 óptimos puede utilizarse para explicar las contribuciones al valor de las características medidas, mientras que la descomposición del residuo por parte del Valuation Engineer puede utilizarse para explicar las contribuciones al valor de otras características.
8. Sobreajuste: Deben evitarse los modelos sobreajustados, no utilizando características del inmueble que puedan identificar, ya sea individualmente o en combinación, uno o varios inmuebles. En particular, debe evitarse el uso de variables factor cuyos valores puedan especificar un número reducido de inmuebles. Una variable factor que identifica vecindarios en conjuntos de datos de tamaño considerable es apropiada, pero una que identifica inmuebles con piscina en una zona donde existen muy pocas piscinas debe usarse con cautela. Una característica que identifica inmuebles únicos, como coordenadas GIS exactas, números de parcela o direcciones, está absolutamente prohibida.

Si el R^2 de la regresión es del 80 %, esperamos que alrededor del 20 % del valor del inmueble a valuar se deba al residuo. Ahora bien, esto es solo aproximado. Sabemos por la curva CQA-Residuo que la mayor parte de la distribución de los residuos se encuentra en los extremos inferior y superior del CQA. Con base en la curva CQA-Residuo y la pendiente extrema para los valores bajos y altos de CQA, podríamos esperar una estimación menos exacta del residuo en estas zonas. Sin embargo, esto se ve contrarrestado por la relativa escasez de inmuebles en estas zonas y por la mayor diferencia de atractivo que observamos entre inmuebles adyacentes. Así, mientras que en el rango medio podríamos hablar del impacto de un rango de $\pm 5\%$, en los extremos nos preocuparía más un rango algo más estrecho de $\pm 2 - \pm 3\%$

Calcular la exactitud del RCA es, en general, algo difícil debido a la no linealidad de la curva CQA-Residuo. Pero es posible caso por caso, siempre que se disponga de un ranking de residuos razonable. El rango de las puntuaciones CQA es de 0.0 a 10.0. Con un buen modelo MARS, el ranking de residuos debería mostrar un patrón claro de menor a mayor atractivo, y no debería resultar difícil calificar dentro de un $\pm 5\%$ del rango total, es decir, ± 0.5 para la puntuación CQA. Si hay 100 inmuebles en la regresión, eso equivaldría a clasificar el inmueble a valorar con exactitud dentro de un grupo de 10 inmuebles. Así, con referencia a los datos de ventas de Burlingame de la Tabla 3, calculemos la exactitud (o el error) aproximada para un CQA estimado por el VE de 7.5, 5.1 y 1.9, donde su conclusión de valor para el inmueble a valorar es de \$3,000,000. Nótese que, dado que la Tabla 3 solo contiene una muestra de los varios cientos de inmuebles utilizados, será necesario interpolar el residuo cuando corresponda (suponiendo linealidad). Para ser claros, suponemos que el VE está muy seguro, en cada caso, de que su clasificación del inmueble a valorar se encuentra dentro de un rango de ± 5 inmuebles en el ranking de residuos de solo 100 inmuebles.

CQA= 7.5 ± 0.5					
CQA	Residuo	Dif. abs.	Error esp.	Precio de venta hip.	Error pct.
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	Residuo	Dif. abs.	Error esp.	Precio de venta hip.	Error pct.
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	Residuo	Dif. abs.	Error esp.	Precio de venta hip.	Error pct.
2.40	-\$228,411				
1.90	-\$276,381	\$47,970			
1.40	-\$326,938	\$50,557	$\pm \$49,264$	\$3,000,000	$\pm 1.64\%$

Finalmente, la única manera objetiva de verificar la exactitud es establecer un protocolo estricto (que este artículo solo insinúa parcialmente a grandes rasgos) y verificar su cumplimiento. Esto es necesario porque los precios de venta reales pueden desviarse del valor de mercado debido a imperfecciones del mercado. Comparar avalúos con ventas posteriores puede proporcionar cierta indicación de la exactitud. Sin embargo, las complejidades y la inestabilidad del mercado inmobiliario exigen la pericia de valuadores o de Valuation Engineers altamente calificados para proporcionar estimaciones de valor basadas en estándares y protocolos establecidos. Estos deben ser lo suficientemente rigurosos como para asegurar que Valuation Engineers competentes lleguen de manera independiente a casi la misma conclusión de valor. Este debería ser un objetivo alcanzable en la mayoría de los casos en que se dispone de buenos datos.

11.5. El inmueble a valorar como anomalía

Dado que estimar el valor de mercado de un inmueble requiere el análisis de transacciones de venta recientes, es justo preguntarse si las transacciones de venta recientes representan adecuadamente las diversas características de un inmueble a valorar más antiguo que carece de

modernización y reparaciones. Esto es particularmente cierto en la San Francisco Bay Area, donde las transacciones de venta recientes suelen corresponder a viviendas que han pasado por actualizaciones para mejorar su precio de venta. Así, el estado y la calidad promedio de los comparables de venta son generalmente superiores a los de viviendas más antiguas que no han estado en el mercado durante bastante tiempo. Esta diferencia de valor debería ser captada por la estimación del residuo para el inmueble a valorar o, en otras palabras, por su puntuación CQA. Esta es una consideración importante con respecto a la calidad de los datos y la exactitud al valorar el inmueble a valorar.

11.6. La importancia del CVR2 frente a datos inesperados

La regresión MARS se basa en múltiples iteraciones de R/earth aplicadas a un subconjunto de inmuebles seleccionado aleatoriamente, que típicamente comprende del 80 % al 90 % del conjunto de datos de la MLS, denominado conjunto de entrenamiento. El 10 % - 20 % restante de los inmuebles sirve como conjunto de prueba. En esta fase, el modelo MARS se crea a partir del conjunto de datos de entrenamiento y luego se utiliza para predecir los precios de venta de los inmuebles del conjunto de datos de prueba, que después se comparan con los precios de venta reales. Esta metodología ha sido reconocida por los especialistas en minería de datos como la estrategia más eficaz para desarrollar un modelo robusto capaz de predecir los precios de venta de inmuebles aún no observados.

Sin embargo, la utilidad primaria de la métrica CVR2 no radica en reflejar la exactitud fáctica del modelo de regresión respecto de las ventas reales de la zona de mercado, sino en optimizar las capacidades predictivas del modelo respecto de ventas hipotéticas de inmuebles a una fecha efectiva especificada. Estos inmuebles no están incluidos en el conjunto de datos de entrenamiento y pueden exhibir combinaciones de características nuevas o inesperadas, ausentes tanto en los datos de los inmuebles del conjunto de entrenamiento como en los del conjunto de prueba.

Dado este contexto, ¿cómo podemos estimar el precio de venta de un inmueble que no ha sido actualizado en los últimos 15 años, suponiendo que fue puesto en venta, por ejemplo, un mes antes de la fecha efectiva del avalúo? Un modelo que demuestra un CVR2 sólido tiene más probabilidades de proporcionar la estimación más exacta que un modelo que no ha pasado por esta validación cruzada repetitiva. También es vital que el conjunto de datos incluya un número suficiente de inmuebles comparables con distintos grados de actualizaciones, reparaciones y modernización.

En última instancia, si poseemos un modelo robusto que predice con exactitud las contribuciones al valor de las diversas características de los inmuebles en nuestro conjunto de datos de inmuebles comparables, podemos ubicar eficazmente el nuevo inmueble a valorar dentro del ranking por residuo que el modelo establece para estos inmuebles. Esta ubicación nos permite asignar al inmueble una puntuación CQA y un residuo apropiados.

12. Residuos

Como lo indica el nombre “Residual Constraint Approach,” los residuos son centrales en esta metodología. Aquí, residuo se refiere a la diferencia entre el precio de venta neto de un inmueble y el precio de venta predicho por el análisis de regresión.

Un residuo positivo sugiere que el inmueble se vendió por más que su precio estimado, con base en atributos medibles típicos como la superficie habitable bruta (GLA) y el tamaño del lote, lo que implica que podría tener un atractivo mayor que el promedio. Un residuo negativo indica que el inmueble se vendió por menos, lo que potencialmente sugiere un menor atractivo u otros atributos negativos no capturados.

Dado que las viviendas más grandes probablemente tienen residuos más grandes, podríamos utilizar el residuo por pie cuadrado con base en la GLA para una comparación más normalizada. Una discusión más profunda de este tema se reserva para otro artículo.

12.1. Ranking & Scoring (clasificación y puntuación)

Después de crear el modelo MARS (Multivariate Adaptive Regression Splines) para las variables medidas, se calculan los residuos para cada inmueble. Estos residuos representan la diferencia entre los precios de venta netos observados y los precios de venta predichos por el modelo. Debe crearse una columna adicional para el Residual/SF, es decir, el residuo dividido por la superficie habitable de la residencia, en el entendido de que, en general, el residuo de inmuebles específicos es más significativo con el tamaño de la superficie habitable. Otras posibles estadísticas pueden funcionar mejor en algunos vecindarios, pero son tema para otra ocasión.

Supuesto 7: “Clasificación por residuo”

En este artículo, las referencias a “clasificación por residuo” deben entenderse como clasificación por residuo, por residuo/SF, o clasificación por alguna otra función de los residuos, según el Valuation Engineer lo considere más apropiado para la zona de mercado. La experiencia del autor en el SF Bay Area indica que el residuo/SF es generalmente el más confiable, especialmente cuando se trata de un inmueble a valorar con una GLA (Gross Living Area) inusualmente pequeña o grande.

A continuación, los inmuebles se ordenan por sus residuos, del más negativo al más positivo. Luego se genera una puntuación de estado, calidad y atractivo (Condition-Quality-Appeal, CQA) para cada inmueble, en un rango de 0.0 a 10.0. Esta puntuación se calcula de la siguiente manera:

La puntuación CQA corresponde al rango percentil del residuo de un inmueble entre todos los inmuebles. Específicamente, la puntuación se calcula determinando el porcentaje de inmuebles que tienen residuos menores que el del inmueble dado y dividiendo ese porcentaje entre 10. Por ejemplo, si el residuo de un inmueble es mayor que los residuos del 50 % de los inmuebles, entonces su puntuación sería 5.0.

Debido al redondeo, un inmueble puede recibir una puntuación de 10.0, que representa al inmueble con el residuo más alto. Sin embargo, una puntuación perfecta de 10.0 es teóricamente imposible porque, lógicamente, ningún inmueble puede tener un residuo mayor que el del 100 % de los inmuebles, incluido él mismo. Por otro lado, la puntuación más baja posible de 0.0 sí es posible, ya que al menos un inmueble no tiene otros inmuebles con residuos menores. Nótese que más de un inmueble podría tener una puntuación de 0.0 debido al redondeo si se comparan muchos inmuebles.

Por lo tanto, la puntuación CQA proporciona una medida intuitiva de qué tan bien los atributos

de un inmueble se alinean con las expectativas del modelo, donde las puntuaciones más altas indican una mejor alineación con las predicciones del modelo.

Es esencial reconocer que la puntuación CQA solo puede considerarse un indicador indirecto del atractivo de un inmueble comparable cuando las condiciones de venta son acordes con su valor de mercado. El Valuation Engineer es responsable de determinar si el valor residual incorpora elementos por encima o por debajo del mercado, mediante una explicación exhaustiva y una investigación minuciosa de la transacción de venta. Si bien las ventas significativamente por encima o por debajo del mercado no son frecuentes, pueden ocurrir en circunstancias tales como ventas forzadas, sucesiones testamentarias, incompetencia, colusión, etc.

La Tabla 3 presenta un extracto semirrealista de datos de la MLS, junto con los residuos y las puntuaciones CQA, tanto en total como por pie cuadrado. El número total de inmuebles analizados en este estudio fue de aproximadamente 600. Es digno de mención que varias variables que fueron incluidas en la regresión no se muestran. En otras palabras, los residuos presentados son posteriores a que el impacto de otras variables, como la fecha de venta, haya sido considerado por la regresión MARS.

Cuadro 3: Comparables de venta ordenados por residuo

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. Curva CQA-Residuo

Un gráfico de la CQA frente al Residual/SF para un vecindario invariablemente dará como resultado la curva característica CQA-Residuo que se muestra en [Figure 2](#). Este gráfico en particular representa datos de Burlingame, CA, hacia 2022. Generalmente se observan curvas similares al graficar las puntuaciones CQA frente a los valores de los residuos; sin embargo, prefiero utilizar el Residual/SF por su relevancia para contextualizar la CQA con respecto a la superficie habitable bruta (GLA) del inmueble a valorar. Al multiplicar el Residual/SF por la GLA, se puede convertir la CQA de un inmueble en un valor residual correspondiente.

El Valuation Engineer debe evaluar cuidadosamente si el tamaño del inmueble, específicamente la superficie habitable bruta (GLA), influye significativamente en la magnitud de los residuos. Esta evaluación es crítica porque puede determinar el método más exacto para calcular los valores de los inmuebles. Por ejemplo, los electrodomésticos de cocina de alta gama y los accesorios de baño lujosos pueden aumentar sustancialmente el valor residual de una vivienda. Sin embargo, en vecindarios donde el tamaño y la distribución de las cocinas y los baños son generalmente uniformes, estas mejoras podrían no correlacionarse fuertemente con la GLA. Su impacto en el valor residual puede depender menos del espacio que ocupan y más de la calidad que representan.

Por el contrario, las características y los materiales exteriores costosos, que mejoran notablemente el atractivo exterior de un inmueble, pueden tener una relación más directa con el tamaño del inmueble. Las viviendas más grandes pueden dar cabida a exteriores más elaborados, como un paisajismo extenso, materiales de techado superiores o patios amplios, que naturalmente realzan el atractivo general del inmueble y, en consecuencia, su valor residual.

Por lo tanto, es esencial que los Valuation Engineers consideren estas dinámicas al analizar los residuos. No solo deben evaluar el valor intrínseco añadido por accesorios y características de alta calidad, sino también comprender cómo estas mejoras interactúan con las dimensiones físicas del inmueble. Este enfoque integral asegura que las valuaciones reflejen con exactitud tanto los atributos cualitativos como los cuantitativos de un inmueble.

En consecuencia, se recomienda graficar tanto la curva CQA-Residuo como la curva CQA-Residual/SF y estudiar las diferencias, las cuales pueden indicar patrones en una zona de mercado. El Valuation Engineer puede hacer que su programa calcule ambos residuos y tome algún promedio ponderado para calcular el residuo que se asignará al inmueble a valorar. Puede que ello marque una gran diferencia, o puede que no.

12.3. Características de la curva CQA

La curva de estado, calidad y atractivo (CQA) ofrece un modelo sofisticado para comprender cómo fluctúan los valores de los inmuebles en función del atractivo, el estado y la dinámica del mercado dentro de un área determinada. Este modelo captura la interacción de los factores económicos y físicos que influyen en el mercado inmobiliario, reflejando las realidades matizadas de la valuación inmobiliaria en los distintos segmentos del mercado.

12.3.1. Análisis de los componentes de la curva CQA

1. Decaimiento exponencial en el lado izquierdo de la curva:

A medida que la puntuación CQA disminuye de alrededor de 2 a 0, los inmuebles exhiben características de deterioro rápido. Esta parte de la curva se representa bien mediante una función de decaimiento exponencial, lo que enfatiza la rapidez con que un inmueble puede perder su atractivo y su valor sin el mantenimiento y las actualizaciones adecuados. La

degradación física, como la madera podrida, los colores desvanecidos y el metal oxidado, acelera este declive, particularmente en viviendas que no reciben mantenimiento regular.

2. Incremento lineal en el rango medio:

La mayor parte del mercado típicamente encaja en esta categoría, donde el atractivo aumenta de manera casi lineal desde puntuaciones de alrededor de 2 hasta 8. Esta sección refleja inmuebles que en general se mantienen conforme al estándar, pero que no necesariamente presentan cualidades o ubicaciones excepcionales. Representa la amplia clase media de la vivienda, abarcando la mayoría del mercado habitacional, donde las viviendas tienen precios competitivos basados en sus características estándar y su estado general.

3. Crecimiento exponencial en el lado derecho de la curva:

Aquí, la curva se eleva exponencialmente desde las puntuaciones 8 a 10, reflejando el extremo de mayor atractivo del mercado. Los inmuebles de este segmento a menudo pertenecen a personas más adineradas que no solo invierten en un mantenimiento superior, sino también en mejoras que aumentan significativamente el atractivo y el valor del inmueble. El crecimiento exponencial refleja la escasez de inmuebles de tan alta calidad y la alta demanda entre los compradores acaudalados. Y entiéndase: una vivienda de menor precio puede calificar muy alto en términos de atractivo, mientras que una vivienda grande y costosa puede calificar bajo. Estas dinámicas se descubren mediante el uso experto de MARS y usualmente pasan desapercibidas para la mayoría de los valuadores.

12.3.2. Dinámica del mercado e implicaciones sociales

La curva CQA narra implícitamente la historia de la riqueza relativa y la disparidad económica dentro de las distintas regiones. En lugares como el condado de San Mateo, la variación en los valores de los inmuebles entre áreas como Burlingame y vecindarios más adinerados como Hillsborough, Woodside o Atherton ilustra divisiones socioeconómicas significativas. Esta disparidad afecta todo, desde el mantenimiento de los inmuebles hasta las valuaciones de mercado, y en última instancia influye en la demografía de los compradores y en las tendencias del mercado inmobiliario.

Además, la anécdota sobre una persona relativamente adinerada que invierte en una vivienda de alta calidad dentro de un vecindario de nivel medio con el que se siente cómoda y donde, por diversas razones, preferiría vivir, subraya otro aspecto crítico de la economía inmobiliaria — el concepto de inmuebles “sobreconstruidos” (overbuilt). Tales inmuebles, aunque muy atractivos, pueden no encontrar compradores locales capaces de pagar el precio que podrían haber alcanzado en una zona de mercado de precios más altos.

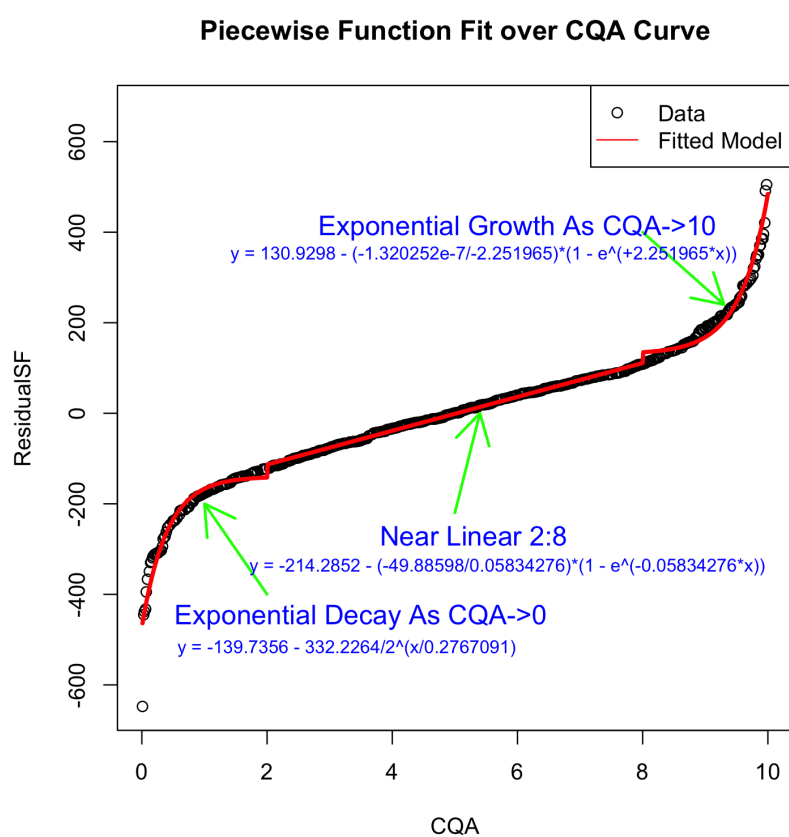
La curva CQA-Residuo es característica de cada zona de mercado, y no debe pasarse por alto al realizar el análisis.

12.4. Residuo del inmueble a valorar

A diferencia de los comparables, el inmueble a valorar no posee un residuo determinado objetivamente; este debe ser estimado por el Valuation Engineer (VE). Esta estimación introduce un grado de subjetividad en el protocolo del Residual Constraint Approach (RCA), marcando una coyuntura crítica en la que el juicio personal desempeña un papel fundamental.

Para estimar la puntuación de estado, calidad y atractivo (CQA), el VE analiza meticulosamente los inmuebles clasificados, buscando patrones discernibles dentro de las clasificaciones. Este análisis a menudo implica una revisión detallada de las fotos de la MLS asociadas con cada

Figura 2: Función característica CQA-Residuo



inmueble. Una revisión exhaustiva es impracticable para un conjunto de datos grande, como uno de 600 inmuebles. En su lugar, el VE aplica un razonamiento práctico para aproximar la posición probable del inmueble a valorar dentro de la clasificación general. Los inmuebles que requieren reparaciones extensas, o “fixers,” generalmente se posicionan cerca del fondo, mientras que aquellos con actualizaciones y atractivo superiores tienden a clasificarse cerca de la cima.

El VE adopta entonces un enfoque más focalizado, ‘saltando’ dentro de la clasificación para identificar comparables que coincidan estrechamente con el atractivo del inmueble a valorar. A medida que el rango se estrecha, el VE presta mayor atención a las similitudes entre los inmuebles, particularmente en aspectos como la calidad, el estado, la estética del diseño (incluidos elementos como la pintura, la carpintería, la cantería y las ventanas), la utilidad funcional y otras características relevantes

Es crucial comprender que los juicios del VE están fundamentalmente guiados por las clasificaciones establecidas por el mercado. Esto asegura que los elementos subjetivos del proceso estén anclados en las realidades del mercado, en procura de una valuación justa y exacta del inmueble a valorar.

12.5. Conclusión de valor del RCA

12.5.1. Consideraciones generales

Determinar una puntuación CQA es fundamental para establecer el valor residual estimado de un inmueble, ya que estos elementos están intrínsecamente vinculados. Si bien es posible evaluar un inmueble sin una CQA, esta métrica proporciona un contexto valioso al ilustrar la posición del inmueble en comparación con otros en la zona de mercado. Por ejemplo, una puntuación CQA de 3.2 indica que el inmueble tiene mayor atractivo que el 32 % de los inmuebles comparables en las cercanías. Si existen objeciones a la puntuación asignada, los objetores deben consultar las clasificaciones y justificar cualquier ajuste propuesto con base en evidencia fotográfica u otros datos sustanciales. Si el sistema de clasificación es robusto y el Valuation Engineer (VE) ha posicionado con exactitud el inmueble dentro de este marco, es poco probable que surjan discrepancias significativas.

En la fase de elaboración del informe, típicamente se seleccionan de 6 a 12 comparables de venta recientes y relevantes para respaldar la conclusión de la valuación. Se analizan las diferencias entre estos comparables y el inmueble a valorar, ajustándose cada variable en consecuencia. Estos ajustes se aplican luego a los precios de venta netos de los comparables para derivar los precios de venta ajustados. Como se demostró matemáticamente, estos precios ajustados deben converger con el precio de venta estimado del inmueble a valorar.

12.5.2. Revisión, auditoría, quejas

Una vez determinado el residuo estimado del inmueble a valorar, pueden calcularse los ajustes residuales para todos los comparables. Estos ajustes sirven como restricciones para desgloses manuales adicionales por parte del VE, adaptados a las variables CQA específicas que se consideren pertinentes. Es importante enfatizar que estos desgloses detallados son explicativos y no influyen en la valuación final, como se afirma en la Demostración II.

En cuanto a las quejas presentadas por usuarios no satisfechos con la conclusión de valor del avalúo, un valuator tradicional típicamente tiene que perder tiempo explicando por qué tal o cual inmueble con un precio más alto o más bajo no tiene el impacto en el valor que el usuario esperaba. Con el RCA, todas las ventas comparables posibles (usualmente 100-600) se evalúan típicamente exactamente al mismo valor que la conclusión de precio, por lo que se requiere

poco esfuerzo para mostrar los ajustes de algún inmueble arbitrario. Todo el proceso del RCA es automatizado y objetivo, excepto por la puntuación del residuo del inmueble a valorar. Y así, otra ventaja del RCA es que la crítica por imperfecciones en el juicio subjetivo del VE queda principalmente restringida a esa única área - y puede gestionarse de manera efectiva y eficiente. Además, suponiendo un R^2 de alrededor del 80 %, el impacto de un error de $\pm 10\%$ en la estimación del residuo del inmueble a valorar equivale a un error de $\pm 2\%$ en la estimación final del valor.

En muchos casos, el inmueble a valorar encaja tan perfectamente en la clasificación residual de los inmuebles, que puede esperarse que la exactitud esté dentro de $\pm 1\%$.

La suma del residuo estimado y el precio de venta del inmueble a valorar estimado por MARS produce la conclusión de valor final, conocida como la Conclusión de Valor del Residual Constraint Approach (RCA). Este proceso asegura un enfoque integral y transparente de la valuación inmobiliaria, fomentando la confianza en la exactitud y la equidad de la evaluación.

13. Conclusión

El Residual Constraint Approach es un marco con tres contribuciones sustantivas: una propiedad estructural anti-anclaje derivada de ajustar el modelo sobre el mercado completo en lugar de sobre un conjunto reducido de comparables (Demostración I); un resultado de invariancia que permite que la descomposición del residuo sirva como capa explicativa sin alterar la conclusión de valor (Demostración II y la observación que la sigue); y una regularidad empírica — la curva característica CQA-Residuo — que parece repetirse en distintas zonas de mercado con la forma característica exponencial-lineal-exponencial. Las limitaciones del marco son igualmente sustantivas: la descomposición del residuo es invariante pero no identificadora, el paso de asignación de CQA depende del juicio del VE, que aún no ha sido validado mediante un estudio formal de concordancia entre evaluadores, y las afirmaciones de exactitud están pendientes de una evaluación comparativa frente a alternativas.

Lo que este artículo presenta es un marco y un protocolo. Lo que aún no presenta, pero sí lo hace el programa de investigación más amplio actualmente en curso, es una cadena de procesamiento ejecutable y auditable que operacionaliza el protocolo en software con valores predeterminados defendibles. Esa cadena de procesamiento — el paquete earthUI en CRAN para la etapa de funciones base de MARS, el paquete glmnetUI para el modelo estructural regularizado sobre dichas bases, y el paquete mgcvUI para el suavizado espacial — es el tema de artículos complementarios en este número y constituye el puente del RCA-como-método al RCA-como-entregable. La brecha entre doctrina y práctica en la regulación de la valuación (Craytor, 2026, este número) aporta la razón por la que ese puente importa: los productos de la valuación se consumen económicamente como insumos de riesgo prospectivos, y un marco que hace que la opinión de valor presente sea auditable y reproducible es una condición previa para cualquier tratamiento honesto de ese uso prospectivo. El RCA es un movimiento dentro de ese proyecto más amplio. No es el proyecto completo, y este artículo no pretende que lo sea.

Referencias

- Ambrose, Brent W. et al. (2025). «Do Appraiser and Borrower Race Affect Mortgage Collateral Valuation?» En: *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». En: *Valuation Engineer Journal* 1.1.
- Friedman, Jerome H. (1991). «Multivariate Adaptive Regression Splines». En: *The Annals of Statistics* 19.1, págs. 1-141.
- Hastie, Trevor, Robert Tibshirani y 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 y Michael LaCour-Little (2020). «AVM versus Appraisal-Based Underwriting in Refinance Mortgages: The Trade-off Between Noise and Bias». En: *SSRN Electronic Journal*. DOI: [10.2139/ssrn.3603327](https://doi.org/10.2139/ssrn.3603327).
- Kuhn, Max y 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*. Inf. téc. URL: <http://www.milbo.org/doc/earth-notes.pdf>.

- Milborrow, Stephen (2024c). *Variance Models in Earth*. Inf. téc. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.
- Perry, Andre M., Jonathan Rothwell y David Harshbarger (2018). *The Devaluation of Assets in Black Neighborhoods*. Inf. téc. 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*. Inf. téc. Brookings Institution.
- Property Appraisal and Valuation Equity Task Force (2022). *Action Plan to Advance Property Appraisal and Valuation Equity*. Inf. téc. U.S. Department of Housing y Urban Development.
- Tzioumis, Konstantinos (2016). «Appraisers and Valuation Bias: An Empirical Analysis». En: *Real Estate Economics* 45.3, págs. 679-712. DOI: [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- Williamson, Jake y Mark Palim (2022). *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Inf. téc. 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: una interfaz interactiva para el paquete earth (MARS)

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

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

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

Este artículo de acceso abierto se distribuye bajo los términos de la [licencia Creative Commons Atribución 4.0](#).

Resumen. earthUI es una interfaz gráfica de usuario para el paquete de R **earth**, que implementa Multivariate Adaptive Regression Splines (MARS). Ofrece tres modos de uso — modelado predictivo general, valuación inmobiliaria y análisis de zona de mercado — y guía al usuario a través de la importación de datos, la configuración del modelo, el ajuste y la interpretación de los diagnósticos y salidas gráficas de earth. Este artículo documenta los requisitos de formato de datos de earthUI, el flujo de trabajo de modelado, las pantallas de salida y la referencia completa de funciones.

Palabras clave: MARS, multivariate adaptive regression splines, earth, R, interfaz gráfica de usuario, valuación inmobiliaria

Nota del editor (julio de 2026): Este artículo describe earthUI 0.10.0. Poco antes de la publicación, earthUI 0.11.0 absorbió las aplicaciones complementarias glmnetUI y mgcvUI (descritas en los dos artículos siguientes), consolidando los flujos de trabajo de MARS, elastic-net y GAM en una sola aplicación. La interfaz y el flujo de trabajo descritos aquí no se ven afectados por la consolidación.

Índice

1. Introducción	4
1.1. ¿Qué es earthUI?	4
1.2. Un paquete, tres rutinas	4
1.3. Tres modos de propósito	4
1.4. Funciones específicas para bienes raíces	5
1.5. Instalación de earthUI y vProlog	5
1.6. Primeros pasos	6
2. Requisitos de los datos de entrada del MLS	7
2.1. Formato y estructura del archivo	7
2.2. Columnas requeridas para el modo Appraisal	7
2.3. Convenciones especiales de nomenclatura de columnas	8
2.4. Calidad e integridad de los datos	8
2.5. Ubicación de la fila del inmueble a valorar	9
3. Modo de Propósito General (General Purpose Mode)	10
3.1. Panorama general	10
3.2. El flujo de trabajo de la barra lateral	10
3.3. Pestañas del panel principal	11

3.4.	Persistencia de la configuración	11
3.5.	Modo oscuro (Dark Mode)	11
3.6.	Configuración regional y de idioma (Locale)	12
3.6.1.	Países admitidos	12
3.6.2.	Anulación del tamaño de papel	12
3.6.3.	Guardar valores predeterminados	13
4.	Modo de Valuación (Appraisal Mode)	14
4.1.	Manejo de la Fila del Inmueble a Valuar	14
4.2.	Effective Date y Sale Age	14
4.3.	Designaciones de Columnas Special	14
4.4.	Residual Contribution (Estimación de Características Poco Frecuentes)	15
4.5.	Panorama de los Ajustes RCA	16
5.	Modo de Análisis de Zona de Mercado (Market Area Analysis Mode)	17
5.1.	Diferencias con el Modo de Valuación	17
5.2.	Cuándo Usar el Modo de Mercado	17
6.	Selección de variables	18
6.1.	Variable(s) objetivo	18
6.2.	La tabla de predictores	18
6.3.	Detección de tipos de datos y anulaciones	19
6.4.	Indicadores Factor y Linear	19
6.5.	Referencia de tipos especiales de columna	19
6.6.	Objetivos múltiples	20
7.	Selección de Parámetros	21
7.1.	Parámetros Básicos	21
7.2.	Pasada Hacia Adelante (Forward Pass)	21
7.3.	Interacciones Permitidas (Allowed Interactions)	22
7.4.	Poda	22
7.5.	Validación Cruzada	23
7.6.	Filtrado por Subconjunto (Subset Filtering)	23
7.7.	Valores de Parámetros Recomendados	23
7.7.1.	Parámetros de la Pasada Hacia Adelante	23
7.7.2.	Parámetros de Poda	25
7.7.3.	Parámetros de Validación Cruzada	25
7.7.4.	Modelo de Varianza	26
7.7.5.	Archivo de Salida de Earth	26
7.8.	Valores Predeterminados de Configuración	26
8.	Ajuste del modelo Earth	27
8.1.	El botón Fit	27
8.2.	Ventana modal de ajuste y salida de traza	27
8.3.	Registro del ajuste	27
9.	Descarga de datos	28
9.1.	Columnas de salida	28
9.2.	Orden de columnas y formato de Excel	28
9.3.	Puntuaciones CQA	29
9.4.	Ordenamiento y fila del inmueble a valuar	29
9.5.	Exportación automática para mgcvUI	29

9.6. Importación en glmnetUI	30
9.6.1. Beneficios clave	30
9.6.2. Importante: la columna de pesos debe coincidir	30
9.6.3. Flujo de trabajo	30
10.Cálculos RCA y descarga	31
10.1. Apertura del diálogo RCA	31
10.2. Interpolación de la puntuación CQA	31
10.3. Columnas de salida	31
10.4. La pestaña RCA Adjustments	32
11.Sales Comparison Grid	33
11.1. Descripción general y propósito	33
11.2. Modal de selección de comparables	33
11.3. Estructura de la cuadrícula	33
11.4. Precio de venta, concesiones y Net SP	34
11.5. Filas agrupadas (Location, Site, Age)	34
11.6. Filas de variables del modelo	34
11.7. CQA Residual y filas de características residuales	35
11.8. Fórmula del Adjusted Sale Price	35
11.9. Protección de hojas y celdas editables	35
11.10.Trabajo con la cuadrícula en Excel	35
12.Descarga de informes	37
12.1. Generate Quarto Report	37
12.2. Convert Quarto Report	37
12.3. Contenido del informe	37
13.Los módulos de posprocesamiento: Elastic Net y GAM	39
13.1. ¿Por qué ajustes de apoyo?	39
13.2. Ejecución y estado compartido	39
13.3. La rutina de Elastic Net	39
13.4. La rutina GAM	39
13.5. Un flujo de trabajo práctico	40
14.Conjunto de datos de demostración: Appraisal_1.csv	41
14.1. Descripción	41
14.2. Columnas	42
14.3. Inicio rápido sugerido	42
15.Experimental: extracción de texto con Prolog & reglas de derivación	44
15.1. Qué hace	44
15.2. Requisitos: vProlog (Trealla), no SWI-Prolog	45
15.3. Cómo habilitarla y ejecutarla	45
15.4. Reglas de derivación	45
15.5. Archivo de datos activo versionado	46
15.6. Interfaz programática & archivos incluidos	46
15.7. Experimento relacionado: priores por región climática	46
15.7.1. Más allá de California: paquetes jurisdiccionales y módulos de zona de mercado	46

1 Introducción

1.1 ¿Qué es earthUI?

earthUI es una interfaz gráfica de usuario para el paquete **earth** de R, que implementa Multivariate Adaptive Regression Splines (MARS) (Friedman 1991). El motor de modelado es el paquete **earth** de Stephen Milborrow (Milborrow 2024), documentado en detalle por Milborrow (s.f.[a]) y Milborrow (s.f.[b]). Se ejecuta como una aplicación Shiny local — no hay inicio de sesión, ni servidor, ni cuentas. Se inicia desde R, se importa un conjunto de datos (CSV o Excel), se configura el modelo y se ajusta de manera interactiva.

La aplicación proporciona un flujo de trabajo completo: importación de datos, configuración de variables, ajuste del modelo, gráficos de diagnóstico, importancia de las variables, ecuaciones del modelo e informes descargables en formato HTML, Word o PDF.

1.2 Un paquete, tres rutinas

A partir de la versión actual, el paquete earthUI también *contiene* sus dos Post Processing Modules (módulos de posprocesamiento), que en versiones anteriores eran los paquetes separados glmnetUI y mgcvUI. Instalar earthUI instala las tres rutinas; cada una se ejecuta como su propia aplicación Shiny local:

- `earthUI::launch()` — la aplicación earth (MARS) descrita en este artículo, en el puerto 7878. `earth()` es el motor central de valuación.
- `earthUI::launch_glmnet()` — el Post Processing Module de elastic net (puerto 7879), que importa las funciones base del modelo earth y las reestima bajo regularización lasso/elastic-net (véase “glmnetUI Import” más adelante).
- `earthUI::launch_mgcv()` — el Post Processing Module de GAM (puerto 7880), que utiliza las ubicaciones de los nodos del modelo earth como puntos de partida para los términos suavizados de `mgcv` (véase “mgcvUI Auto-Export” más adelante).

Las tres rutinas comparten un mismo árbol de proyectos, una misma base de datos de configuración y una misma copia de la infraestructura de soporte, de modo que un proyecto abierto en una aplicación queda inmediatamente disponible para las demás. Las rutinas glmnet y mgcv constituyen evidencia de apoyo para la conclusión de valor del modelo earth, no flujos de trabajo en competencia: el patrón previsto consiste en ajustar y refinar primero el modelo earth, y luego ejecutar los módulos sobre el mismo proyecto cuando se desee corroboración.

1.3 Tres modos de propósito

earthUI organiza el trabajo en **proyectos** — entidades con nombre que mantienen juntos los archivos de entrada, los resultados y la configuración en una ubicación fija. Cada proyecto lleva un **Purpose** (propósito), elegido en el asistente New Project cuando se crea el proyecto. Existen tres modos:

- **General** — Regresión Earth para cualquier tipo de población o conjunto de datos. Este es el modo predeterminado. Proporciona el flujo de trabajo completo de modelado MARS sin adiciones específicas de dominio.
- **For Appraisal** (para valuación) — Regresión Earth adaptada a la valuación inmobiliaria. Agrega funciones específicas para la valuación de un inmueble individual, incluidos el manejo del inmueble a valuar, las designaciones especiales de columnas, el Residual Constraint Approach (RCA) (es.: enfoque por restricciones residuales) y la generación del Sales Comparison Grid.

- **Market Area Analysis** (análisis de zona de mercado) — Regresión Earth adaptada a estudios de zonas de mercado. Agrega funciones para analizar grupos de inmuebles en un mercado definido, en lugar de valorar un único inmueble a valorar.

En los tres modos, el motor central de modelado es idéntico — siempre se ajusta un modelo Earth (MARS). El ajuste de propósito controla qué herramientas y elementos de interfaz adicionales están disponibles.

El propósito es fijo por proyecto. El propósito de un proyecto — junto con su país, estado, condado y ciudad — queda bloqueado cuando se crea el proyecto. Cambiar a un proyecto diferente restablece earthUI a un estado predeterminado limpio (los datos importados, los resultados del modelo, las pestañas, la configuración de variables y los parámetros de earth se borran en su totalidad) y luego restaura la configuración guardada de ese proyecto según su propósito.

1.4 Funciones específicas para bienes raíces

Cuando se selecciona **For Appraisal** o **Market Area Analysis**, earthUI activa varias funciones diseñadas para el análisis de bienes raíces:

- **Designaciones especiales de columnas** — Cada predictor puede etiquetarse con un rol especial como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` o `display_only`. Estas designaciones controlan cómo se maneja la columna durante el ajuste, la salida y la generación del Sales Grid (cuadro de comparación). Véase el Capítulo 6 para la referencia completa.
- **Redondeo de latitud y longitud** — Las columnas designadas como latitud o longitud se redondean automáticamente a 3 decimales para prevenir el sobreajuste.
- **Column Sale Age** — Cuando una columna se designa como `contract_date` y se proporciona una Effective Date (fecha efectiva), earthUI calcula una columna `sale_age` (días entre la fecha de venta y la fecha efectiva) y la sustituye como predictor.
- **Cálculos RCA (solo valuación)** — En el modo de valuación, después de ajustar el modelo, earthUI puede calcular la salida del Residual Constraint Approach (RCA), que produce ajustes por comparable, resúmenes de ajustes netos/brutos y un precio de venta ajustado para el inmueble a valorar.
- **Sales Comparison Grid (solo valuación)** — Genera un libro de Excel de varias hojas con fórmulas para los ajustes de los comparables, desgloses residuales por característica y precios de venta ajustados. Véase el Capítulo 11.

***Sugerencia:** El modo de propósito General funciona con cualquier conjunto de datos — financiero, científico, de ingeniería o inmobiliario. Los modos de valuación y de mercado simplemente agregan funciones de conveniencia para los profesionales de bienes raíces.*

1.5 Instalación de earthUI y vProlog

earthUI está publicado en CRAN, por lo que la instalación estándar es una línea en R:

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

Este artículo describe la versión actual. Si CRAN aún ofrece una versión anterior cuando usted lea esto, la versión de desarrollo se instala desde GitHub (nótese el `subdir` — el paquete reside en la carpeta `pkg/` del repositorio):

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

vProlog (opcional). La funcionalidad experimental de procesamiento de observaciones (véase “Prolog Text Extraction” cerca del final de este artículo) utiliza el paquete vProlog — un motor Prolog embebido y autocontenido. Deliberadamente *no* es una dependencia obligatoria: earthUI se instala y se ejecuta sin él, y las funciones de Prolog simplemente permanecen deshabilitadas hasta que esté presente. vProlog se distribuye a través del repositorio r-universe del autor, que ofrece binarios precompilados para Windows y macOS (no se necesita ninguna cadena de herramientas de compilación):

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

No se requiere ninguna instalación de Prolog en el sistema (SWI-Prolog u otra) — el motor Trealla está compilado dentro del propio paquete. Todo lo demás en este artículo funciona sin vProlog.

1.6 Primeros pasos

Para usar earthUI:

1. **Instale el paquete** desde CRAN como se describió anteriormente.
2. **Inicie la aplicación:** ejecute `earthUI::launch()` en R, o desde la línea de comandos ejecute `Rscript -e 'earthUI::launch()'`. La aplicación se abre en su navegador web en el puerto 7878. También puede acceder a la aplicación directamente navegando a `http://localhost:7878` en su navegador. (Los Post Processing Modules se inician de la misma manera: `launch_glmnet()` en el puerto 7879, `launch_mgcv()` en el puerto 7880.)
3. **Cree o abra un proyecto** en la Sección 1 de la barra lateral. El asistente New Project (nuevo proyecto) captura el nombre del proyecto, el Purpose (General, For Appraisal o Market Area Analysis), la ubicación (país, estado, condado, ciudad) y, opcionalmente, un archivo de datos inicial, que se copia en la carpeta `in/` del proyecto.
4. **Importe sus datos** en la Sección 2, que lista los archivos CSV y Excel de la carpeta `in/` del proyecto activo. Los archivos que usted agregue manualmente mediante Finder o Explorer aparecen tras hacer clic en Refresh.
5. **Configure las variables** — elija su(s) variable(s) objetivo y los predictores, establezca los tipos de datos y asigne cualquier rol especial de columna.
6. **Establezca los parámetros de Earth** — degree, nprune, filtros de subset y otros argumentos de `earth()`.
7. **Ajuste el modelo** — haga clic en “Fit Earth Model” y revise los resultados en el panel principal.
8. **Exporte** — descargue las predicciones como Excel, genere paquetes de informes Quarto o (en el modo de valuación) calcule los ajustes RCA y los Sales Comparison Grids.

La configuración se guarda del lado del servidor en una base de datos del proyecto, indexada por proyecto y propósito, mediante los botones “Save current as default” de las Secciones 3 y 4 (véase el Capítulo 3, “Settings Persistence”).

2 Requisitos de los datos de entrada del MLS

Para los flujos de trabajo de valuación inmobiliaria y análisis de mercado, los datos de entrada provienen típicamente de una exportación de un Multiple Listing Service (MLS). Este capítulo describe la estructura de archivo esperada y las columnas que earthUI puede utilizar.

2.1 Formato y estructura del archivo

earthUI acepta archivos **CSV** y **Excel** (.xlsx, .xls). Durante la importación, los nombres de las columnas se convierten automáticamente a **snake_case** — por ejemplo, “Living SqFt” se convierte en **living_sqft**, “Contract Date” se convierte en **contract_date** y “Sale Price” se convierte en **sale_price**. Esta normalización garantiza referencias de columnas consistentes a lo largo de todo el flujo de trabajo. El separador del CSV y la marca decimal utilizados durante la importación se determinan según la configuración regional (véase el Capítulo 3, “Locale & Regional Settings”).

El archivo de datos debe ser una tabla plana con una fila por inmueble y una columna por atributo. La primera fila del archivo debe contener los encabezados de las columnas.

2.2 Columnas requeridas para el modo Appraisal

Si bien earthUI funciona con cualquier conjunto de columnas, el flujo de trabajo completo de valuación (ajustes RCA + Sales Comparison Grid, cuadro de comparación de ventas) se beneficia de contar con las siguientes columnas en la exportación del MLS:

Columna	Tipo especial	Propósito
Sale Price	(target)	Variable de respuesta del modelo
Contract Date	contract_date	Se utiliza para calcular sale_age (días desde la fecha efectiva)
Listing Date	listing_date	Se utiliza junto con la fecha de contrato para calcular el DOM si no existe una columna DOM
Days on Market	dom	Días en el mercado; se muestra en el Sales Grid
Concessions	concessions	Concesiones de la venta; Net SP = Sale Price – Concessions en el Sales Grid
Living Area (SF)	living_area	Habilita los residuos por pie cuadrado (residual_sf , cqa_sf)
Lot Size	lot_size	Se agrupa en la fila "Site Size Dimensions" del Sales Grid
Site Dimensions	site_dimensions	Se agrupa con el tamaño del lote en el Sales Grid
Latitude	latitude	Se redondea a 3 decimales; se utiliza para el cálculo de proximidad y el grupo Location
Longitude	longitude	Se redondea a 3 decimales; se utiliza para el cálculo de proximidad y el grupo Location
Area ID	area	Se agrupa en la fila "Loc: Long Lat Area" del Sales Grid
Actual Age	actual_age	Se agrupa en la fila "Actual Age Effective Age" del Sales Grid
Effective Age	effective_age	Se agrupa con la edad real en el Sales Grid
Parcel Number	(none)	Se muestra en la fila APN del Sales Grid
Listing ID	(none)	Se muestra en la fila APN del Sales Grid
Address	display_only	Se muestra en la fila Address del Sales Grid; se excluye del modelo

Los nombres de las columnas de la hoja de cálculo pueden estar en un idioma extranjero — los nombres “especiales” están en inglés para que el programa de R pueda darles un tratamiento especial. En los demás casos, los nombres de columna proporcionados aparecen en los modelos de regresión, en los gráficos y (si se realizan avalúos) en el Intermediate Sales Grid.

No todas las columnas son requeridas. earthUI se adapta — si falta una columna, la característica correspondiente simplemente se omite. Por ejemplo, si no se designa una columna **concessions**, la fila Net SP del Sales Grid muestra el Sale Price sin deducción de concesiones. Sin embargo, para los modelos de precios inmobiliarios, ciertas columnas son altamente recomendadas para lograr un ajuste aceptable:

1. **Sale Age** — el número de días entre la fecha de venta del contrato y la fecha efectiva del avalúo o análisis. Si se utiliza un historial de ventas de varios años, especialmente para períodos superiores a 5 años, **sale_age** suele desempeñar un papel central en la estimación del precio de venta. De hecho, con frecuencia es tan importante que, sin él, earthUI no logra proporcionar modelo alguno.
2. **Living Area** — también se conoce con nombres como “Living Sqft,” “GLA” (gross living area, superficie habitable bruta), entre otros. Es otro determinante principal del precio de venta.
3. **Total Bath Count** — el número total de baños completos, de un cuarto, medios y de 3/4. Por ejemplo, dos baños completos y un medio baño corresponderían a un valor de 2.5.
4. **Garage Bays** o **Garage Area** — el número de espacios de estacionamiento en el garaje o la superficie del garaje en pies cuadrados.
5. **Lot Size** — la superficie del terreno del inmueble, típicamente en pies cuadrados o acres.
6. **Longitude**, **Latitude**, y, si está disponible, **Area ID**. Los ajustes correspondientes se combinarán bajo un único ajuste de Location en el Sales Grid.

2.3 Convenciones especiales de nomenclatura de columnas

earthUI identifica las columnas por su **designación de tipo especial**, no por su nombre de columna. Puede asignar a sus columnas cualquier nombre en la exportación del MLS — lo que importa es que asigne el tipo especial correcto en la tabla Variable Configuration (configuración de variables) del Capítulo 6.

Por ejemplo, su MLS podría exportar la superficie habitable como “GLA”, “Living SqFt”, “liv_area” o “gross_living_area”. Después de la importación (donde se convierte a snake_case), simplemente la designa como **living_area** en el menú desplegable Special. earthUI la utilizará entonces para los cálculos de residuos por pie cuadrado y para la agrupación en el Sales Grid, independientemente de su nombre original.

2.4 Calidad e integridad de los datos

- **Valores faltantes (NA):** Las filas con valores NA en cualquier columna predictora o de destino se eliminan automáticamente antes del ajuste. El **na.action** siempre se establece internamente en **na.fail**, y la eliminación de NA se realiza antes de la llamada a **earth()**.
- **Columnas de fecha:** Las fechas deben estar en un formato que R pueda interpretar (p. ej., 2025-06-15, 06/15/2025, June 15, 2025). earthUI detecta automáticamente las columnas de fecha cuando al menos el 80 % de los valores se interpretan correctamente. Los años de dos dígitos (p. ej., 06/15/25) se leen como el año coincidente más cercano, de modo que las exportaciones recientes del MLS se interpretan correctamente; los años de cuatro dígitos siempre se toman de forma literal.
- **Columnas numéricas:** El precio de venta, la superficie habitable, el tamaño del lote, las concesiones y campos similares deben ser numéricos. Si su MLS exporta los precios con

símbolos de moneda o separadores de miles (p. ej., “\$350,000” o “350.000,00”), puede ser necesario limpiarlos antes de la importación.

- **Columnas factor:** Las variables categóricas como el identificador de zona, el estilo o el estado deben contener un número manejable de valores únicos. earthUI detecta automáticamente las variables factor, pero usted puede anular dicha detección.

***Sugerencia:** Revise la columna T/F/NA de la tabla de predictores después de la importación — muestra los conteos de valores true/false/NA de cada columna. Las columnas con más del 30 % de valores faltantes se marcan en rojo. Considere excluir las columnas con alto porcentaje de NA o limpiar los datos antes de la importación.*

2.5 Ubicación de la fila del inmueble a valuar

En el modo **Appraisal**, la fila 1 debe corresponder al inmueble a valuar. Todas las filas restantes son ventas comparables. El precio de venta del inmueble a valuar puede dejarse en blanco (NA) o establecerse en cualquier valor — earthUI lo trata como NA durante el ajuste en cualquier caso.

En el modo **Market Area Analysis**, colocar un inmueble a valuar en la fila 1 es opcional, y ninguna fila se excluye automáticamente. Para excluir del ajuste la fila del inmueble a valuar (o cualquier otra fila), asígnele un valor de cero en una columna de ponderaciones designada o utilice un filtro de subconjunto.

En el modo **General**, no existe un tratamiento especial de filas — todas las filas se tratan por igual.

3 Modo de Propósito General (General Purpose Mode)

3.1 Panorama general

El modo General Purpose (propósito general) es el modo predeterminado al iniciar earthUI. Proporciona el flujo de trabajo completo de modelado MARS para cualquier conjunto de datos — no solo para bienes raíces. Puede utilizarse earthUI para datos científicos, análisis financiero, estudios de ingeniería o cualquier problema de regresión en el que se desee explorar relaciones no lineales e interacciones entre variables.

En el modo General, la interfaz omite las funciones específicas de bienes raíces (columnas especiales, antigüedad de la venta, redondeo de coordenadas, RCA). La barra lateral se simplifica para centrarse en la selección de variables, la configuración de parámetros, el ajuste del modelo y la exportación.

3.2 El flujo de trabajo de la barra lateral

La barra lateral está organizada en secciones numeradas y plegables que guían al usuario desde la selección del proyecto hasta la exportación:

1. Project (proyecto) — Permite elegir un proyecto existente o crear uno nuevo mediante el asistente New Project, que registra el propósito, el país y la ubicación administrativa (todos quedan bloqueados una vez creado el proyecto). El proyecto activo determina de dónde se leen los archivos de entrada y dónde se escriben todos los resultados — no existe un campo separado de carpeta de salida.

2. Import Data (importar datos) — Permite elegir un archivo CSV o Excel de la carpeta in/ del proyecto activo; un enlace Refresh vuelve a explorar la carpeta en busca de archivos agregados manualmente. Para archivos de Excel con varias hojas, aparece un selector de hoja. Los nombres de columna se convierten automáticamente a snake_case.

3. Variable Configuration (configuración de variables) — Selector de la variable objetivo (admite múltiples objetivos), Effective Date (fecha efectiva) y la tabla de predictores con casillas de verificación para Include, Factor, Linear y Latent. Véase el Capítulo 6 para los detalles completos.

4. Earth Call Parameters (parámetros de la llamada a earth) — Todos los argumentos de la función `earth()`: degree, penalty, nk, método de poda, validación cruzada y más. Véase el Capítulo 7 para la referencia completa de parámetros.

Execute Prolog Processing (ejecutar procesamiento Prolog) — Una sección opcional, sin numerar, entre las Secciones 4 y 5, deshabilitada a menos que la función experimental de Prolog esté habilitada en Settings. Véase el Capítulo 15.

5. Fit Earth Model (ajustar el modelo earth) — Un campo de semilla aleatoria y el botón que ejecuta el modelo de forma asíncrona. Véase el Capítulo 8 para los detalles del ajuste.

6. Download Estimated Target Variable(s) & Residuals (descargar variables objetivo estimadas y residuos) — (titulada “Download Estimated Sale Prices & Residuals” en los modos de valuación y de mercado) Exporta las predicciones, los residuos, las puntuaciones CQA y las contribuciones por función g como un archivo de Excel. Véase el Capítulo 9.

7–8. RCA Adjustments y Sales Grid — Solo en modo de valuación. Véanse los Capítulos 10 y 11.

9. Generate Quarto Report (generar informe Quarto) — Escribe un paquete de informe Quarto autocontenido (fuente .qmd más recursos) en la carpeta de salida del proyecto. Numerada como 7 fuera del modo de valuación. Véase el Capítulo 12.

10. Convert Quarto Report (convertir informe Quarto) — Renderiza cualquier archivo .qmd a HTML, Word o PDF. Véase el Capítulo 12.

3.3 Pestañas del panel principal

Después del ajuste, el panel principal ofrece diez pestañas:

Pestaña	Contenido
Data	Vista previa de los datos importados. En modo de valuación, se divide en las tablas Subject Property (inmueble a valuar) y Comparable Sales (ventas comparables).
Equation	La ecuación del modelo ajustado mostrada en LaTeX/MathJax, con cada función base y su coeficiente.
Summary	Métricas clave (R^2 , GRSq, GCV, RSS, $CV R^2$) y una tabla de coeficientes.
Variable Importance	Gráfico de barras y tabla clasificada de las puntuaciones de importancia de los predictores.
Contribution	Gráficos de contribución de las funciones g. Seleccione una función g del menú desplegable; los predictores univariados se muestran como gráficos de líneas, y las interacciones bivariadas ($\text{degree} \geq 2$) como gráficos 3D de perspectiva y de contorno.
Correlation	Mapa de calor de las correlaciones entre predictores.
Diagnostics	Tres gráficos: Residuals vs Fitted, Normal Q-Q y Actual vs Predicted.
RCA Adjustments	(solo en modo de valuación) Histogramas de Residual Adj %, Net Adj % y Gross Adj % entre los comparables, que se completan una vez finalizado el Paso 7 (Calculate RCA Adjustments).
Residual Contribution	Estima las contribuciones de mercado de características raras o excluidas mediante la regresión conjunta de los residuos del modelo sobre columnas que el modelo nunca utilizó (véase subsection 4.4).
ANOVA	Tabla de descomposición ANOVA que muestra la contribución de cada función base al RSS.
Earth Output	Salida de texto sin procesar de 'earth::summary()', incluidos los detalles del paso de poda.

3.4 Persistencia de la configuración

earthUI guarda su configuración del lado del servidor en una base de datos de proyectos (`projects.sqlite` en la raíz de `regProj`), indexada por proyecto y propósito. El guardado es **accionado por botones**: el botón “Save current as default” (guardar la configuración actual como predeterminada) de la Sección 3 guarda el objetivo, la fecha efectiva y la configuración de los predictores, mientras que el botón de la Sección 4 guarda los parámetros de la llamada a earth y la matriz Allowed Interactions. Cada botón actualiza únicamente su propia mitad, de modo que nunca se sobrescriben entre sí, y **el ajuste del modelo no guarda automáticamente** — puede experimentarse libremente sin sobrescribir los valores predeterminados guardados. Los diversos archivos de entrada de un proyecto (un pequeño extracto de prueba, el conjunto de datos completo) comparten una sola configuración y, dado que la configuración reside en el árbol del proyecto, viaja con él cuando se sincroniza, se respalda o se comparte la carpeta `regProj`.

3.5 Modo oscuro (Dark Mode)

Haga clic en el ícono de luna/sol en la esquina superior derecha para alternar entre los temas claro y oscuro. La preferencia de tema se guarda en el almacenamiento local y persiste entre sesiones.

3.6 Configuración regional y de idioma (Locale)

earthUI admite convenciones internacionales de formato de números, fechas y CSV mediante un sistema de configuración regional basado en países. El menú desplegable **Country** (país) del menú **Settings** (configuración; ícono de engranaje) de la barra de menú superior selecciona un preajuste para los 31 países admitidos. Cada preajuste configura:

- **Separador CSV** — coma (,) para EE. UU./Reino Unido/Japón o punto y coma (;) para la mayor parte de Europa, donde la coma se usa como marca decimal.
- **Marca decimal** — punto (.) o coma (,).
- **Separador de miles** — coma (EE. UU./Reino Unido/Japón), punto (Alemania/Italia/España), espacio (Finlandia/Francia/Polonia/países bálticos/Ucrania/Rusia) o apóstrofo (Suiza).
- **Formato de fecha** — MM/DD/YYYY (EE. UU.), DD/MM/YYYY (la mayor parte de Europa) o YYYY-MM-DD (Suecia/Lituania/Japón/Canadá).
- **Tamaño de papel** — Letter (EE. UU./Canadá/México) o A4 (el resto del mundo).

3.6.1. Países admitidos

País	Sep. CSV	Decimal	Miles	Fecha	Papel	
EE. UU., Canadá (csv)	,	.	,	MDY/YMD	Letter	
Reino Unido, Irlanda, Australia, Nueva Zelanda	,	.	,	DMY	A4	
Alemania, Austria, Bélgica, Países Bajos, Italia, España, Portugal, Dinamarca, Brasil	;	,	.	DMY	A4	
Francia, Polonia, Rep. Checa, Noruega, Letonia, Estonia	;	,	(espacio)	DMY	A4	
Finlandia, Suecia	;	,	(espacio)	DMY/YMD	A4	
Suiza	;	,	'	DMY	A4	
Ucrania, Rusia	;	,	(espacio)	DMY	A4	
Turquía	;	,	.	DMY	A4	
Japón, Corea del Sur	,	.	,	YMD	A4	
México	,	.	,	DMY	Letter	
Lituania	;	,	(espacio)	YMD	A4	

3.6.2. Anulación del tamaño de papel

Debajo del selector de país, un menú desplegable **Paper** (papel; Letter o A4) permite anular el tamaño de página del informe sin cambiar de país. El separador CSV, la marca decimal y el orden del formato de fecha siguen el preajuste del país seleccionado; al cambiar de país, la anulación del papel se restablece al valor predeterminado de ese país.

3.6.3. Guardar valores predeterminados

Haga clic en **Save** (guardar) en el menú Settings para almacenar sus preferencias regionales de forma global. Estos valores predeterminados se aplican a todas las sesiones futuras, independientemente del proyecto que se abra. La configuración por proyecto (objetivo, predictores, parámetros) se guarda por separado en la base de datos del proyecto; los valores regionales predeterminados persisten en todos los proyectos mediante una base de datos SQLite de nivel de usuario. Este enfoque de dos niveles está diseñado para organizaciones, como las firmas de auditoría, que trabajan con datos de múltiples países — establezca su país más frecuente como predeterminado y ajústelo cuando sea necesario.

***Sugerencia:** el formato de los números en los ejes de los gráficos y en las etiquetas de pendiente se adapta automáticamente a su configuración regional. La configuración alemana usa puntos para los miles (200.000), la finlandesa usa espacios (200 000) y la suiza usa apóstrofes (200'000). No se muestran símbolos de moneda — earthUI es agnóstico respecto de la moneda.*

4 Modo de Valuación (Appraisal Mode)

Cuando se selecciona **For Appraisal** (para avalúo) como Purpose (propósito), earthUI se configura para la valuación de un solo inmueble. Todas las funciones descritas en el Capítulo 3 permanecen disponibles; este capítulo cubre únicamente las adiciones específicas del modo de valuación.

4.1 Manejo de la Fila del Inmueble a Valuar

En el modo de valuación, **la fila 1 del conjunto de datos es el inmueble a valuar** y todas las filas restantes son ventas comparables. El archivo de entrada debe organizarse en consecuencia (véase el Capítulo 2). El precio de venta del inmueble a valuar puede dejarse en blanco o fijarse en cualquier valor — earthUI lo trata automáticamente como NA durante el ajuste.

Después de la importación, la pestaña Data se divide en dos secciones: **Subject Property** (inmueble a valuar; fila 1) y **Comparable Sales** (ventas comparables; filas 2+). La fila 1 se excluye siempre del ajuste del modelo — la notificación “Skipping row 1 (subject). Fitting on N rows.” así lo confirma. Tras el ajuste, el modelo genera de todos modos predicciones para la fila del inmueble a valuar, mostradas como `est_<target>` en la salida.

4.2 Effective Date y Sale Age

Un campo **Effective Date** (fecha efectiva) aparece en la sección Variable Configuration (configuración de variables) en todos los modos de propósito (con la fecha de hoy como valor predeterminado) — los análisis generales y de mercado también tienen una fecha efectiva, no solo los avalúos — pero solo se *utiliza* en los modos de valuación y de mercado. En ellos, si se designa una columna como `contract_date` en el menú desplegable Special, earthUI calcula una columna `sale_age` — el número de días enteros entre la fecha de contrato de cada venta y la fecha efectiva. Esta columna reemplaza a la columna de fecha original como predictor.

La primera vez que se hace clic en Fit (ajustar) después de designar una fecha de contrato, earthUI crea la columna `sale_age` y notifica que debe hacerse clic en Fit nuevamente para incluirla.

4.3 Designaciones de Columnas Special

En los modos de valuación y de mercado, aparece un menú desplegable **Special** (especial) para cada predictor en la tabla Variable Configuration. La lista completa de los tipos especiales y sus efectos:

Tipo Special	Efecto
no	Predeterminado — sin manejo especial
contract_date	Activa el cálculo automático de <code>sale_age</code> a partir del Effective Date
listing_date	Se usa junto con la fecha de contrato para calcular DOM cuando no existe una columna <code>dom</code>
dom	Identifica la columna de Days on Market (días en el mercado); se muestra en el Sales Grid (cuadro de comparación)
concessions	Identifica las concesiones de venta; se usa en la fórmula de Net SP (Sale Price – Concessions) en el Sales Grid
latitude	Valores redondeados a 3 decimales; se usan para la proximidad de Haversine y el grupo Location en el Sales Grid
longitude	Valores redondeados a 3 decimales; se usan para la proximidad de Haversine y el grupo Location en el Sales Grid
area	Se agrupa en la fila "Loc: Long Lat Area" del Sales Grid
living_area	Habilita los residuos por pie cuadrado (<code>residual_sf</code> , <code>cqa_sf</code>) en la salida de descarga
lot_size	Se agrupa en la fila "Site Size Dimensions" del Sales Grid
site_dimensions	Se agrupa con el tamaño del lote en el Sales Grid
actual_age	Se agrupa en la fila "Actual Age Effective Age" del Sales Grid
effective_age	Se agrupa con la edad real en el Sales Grid
display_only	Columna incluida en las exportaciones pero excluida del ajuste del modelo (se permiten varias)
remarks	(Experimental) Columna de comentarios de texto libre que el paso opcional de Prolog analiza y convierte en columnas de características (se permiten varias); véase el Capítulo 15
list	(Experimental) Columna delimitada por comas que el paso opcional de Prolog divide en columnas one-hot TRUE/FALSE (se permiten varias); véase el Capítulo 15

Solo se permite una columna por tipo especial (excepto `display_only`, `remarks` y `list`). Asignar un tipo especial a una segunda columna lo elimina automáticamente de la primera. Una pequeña insignia azul aparece junto al nombre de la variable, mostrando el tipo especial asignado.

4.4 Residual Contribution (Estimación de Características Poco Frecuentes)

Las características demasiado poco frecuentes para el canal de regresión — un taller presente en un puñado de ventas — pertenecen a la capa residual del RCA. La pestaña **Residual Contribution** (contribución residual) estima su contribución de mercado mediante la regresión de los residuos del modelo ajustado sobre columnas que el modelo nunca utilizó, por mínimos cuadrados ordinarios. Seleccionar varias columnas las estima de manera *conjunta*, de modo que las amenidades correlacionadas (p. ej., `outbuilding_sqft` junto con `pr_has_workshop`) se reparten el crédito en lugar de contabilizarse dos veces. Las banderas binarias producen estimaciones de suma global y listan como evidencia el residuo de cada venta marcada; las columnas numéricas producen contribuciones por unidad; el error estándar reportado es de un SE ($t \geq 2$ como criterio habitual de significancia). Los términos respaldados por menos de diez observaciones reciben una advertencia automática CAUTION en la nota generada para el expediente de trabajo, la cual documenta el método, las estimaciones y las salvedades en un solo párrafo para el informe. Solo los predictores que el modelo efectivamente *utilizó* se excluyen de la lista de candidatos — una bandera poco frecuente ofrecida al ajuste pero nunca seleccionada por la pasada hacia adelante sigue siendo elegible, dado que el modelo no absorbió nada de su valor.

4.5 Panorama de los Ajustes RCA

El botón **Calculate RCA Adjustments & Download** (calcular ajustes RCA y descargar; sección 7 de la barra lateral, visible solo en el modo de valuación después del ajuste) calcula ajustes derivados del mercado para cada comparable en relación con el inmueble a valorar. El flujo de trabajo completo del RCA se describe en el Capítulo 10.

Después de calcular los ajustes RCA, el botón **Generate Sales Grid & Download** (generar el Sales Grid y descargar; sección 8 de la barra lateral) queda disponible. El flujo de trabajo del Sales Grid se describe en el Capítulo 11.

***Sugerencia:** La puntuación CQA (estado, calidad y atractivo) que se asigna al inmueble a valorar controla qué proporción de la distribución de los residuos se atribuye al inmueble a valorar. Una puntuación de 5.00 sitúa al inmueble a valorar en la mediana de los comparables.*

5 Modo de Análisis de Zona de Mercado (Market Area Analysis Mode)

Cuando se selecciona **Market Area Analysis** (análisis de zona de mercado) como Purpose, earthUI ofrece las mismas funciones específicas de bienes raíces que el modo de valuación (columnas especiales, antigüedad de la venta, redondeo de coordenadas), pero está orientado al análisis de un grupo de inmuebles en lugar de la valuación de un solo inmueble.

5.1 Diferencias con el Modo de Valuación

- **Sin exclusión automática del inmueble a valorar** — a diferencia del modo de valuación, el modo de mercado no excluye automáticamente la fila 1 del ajuste. Se incluyen todas las filas; para dejar la fila del inmueble a valorar fuera del ajuste, asígnele un valor de cero en una columna de pesos designada (como en el conjunto de datos de demostración) o utilice un filtro de subconjunto.
- **Sin secciones de RCA ni de Sales Grid** — los pasos “Calculate RCA Adjustments & Download” y “Generate Sales Grid & Download” no están disponibles. El modo de mercado se centra en el ajuste del modelo y su salida, no en los ajustes por comparable.
- **Numeración de la barra lateral** — sin los pasos de RCA y Sales Grid, la sección Generate Quarto Report lleva el número 7 en lugar del 9 (la sección Convert Quarto Report conserva el 10).

5.2 Cuándo Usar el Modo de Mercado

El modo Market Area Analysis es apropiado cuando:

- Se construye un modelo de regresión para un vecindario o zona de mercado con el fin de comprender los determinantes del valor
- Se analiza cómo variables como la superficie en pies cuadrados, la edad, el tamaño del lote y la ubicación afectan los precios de venta en un grupo de inmuebles
- Se prepara el soporte para un análisis de condiciones de mercado o una delimitación de vecindario
- Se trabaja con un conjunto de datos que incluye un inmueble a valorar en la fila 1, pero se desea la opción de incluirlo en el ajuste o excluirlo de él

***Sugerencia:** El modo de mercado también es útil para la regresión inmobiliaria general cuando se desean las funciones de columnas especiales (antigüedad de la venta, redondeo de coordenadas) pero no se necesita el flujo de trabajo de ajustes RCA.*

6 Selección de variables

La Sección 3 de la barra lateral — **Variable Configuration** (configuración de variables) — es donde se eligen las columnas que participan en el modelo y cómo se tratan.

6.1 Variable(s) objetivo

El menú desplegable **Target (response) variable(s)** (variable(s) objetivo o de respuesta), situado en la parte superior de la Sección 3, enumera todas las columnas del conjunto de datos. Seleccione una columna para un modelo estándar de respuesta única, o varias columnas para un modelo de respuesta múltiple. Cuando se seleccionan varios objetivos, el modelo ajusta todas las respuestas simultáneamente mediante `earth(cbind(y1, y2, ...) ~ .)`.

Las columnas seleccionadas como objetivos se excluyen automáticamente de la lista de predictores.

6.2 La tabla de predictores

Debajo del selector de objetivos, una tabla enumera todas las columnas restantes con los siguientes campos:

Columna	Descripción
Variable	Nombre de la columna (el nombre completo se muestra en una descripción emergente si está truncado), precedido por un conmutador $\times/+$ (véase más abajo). En los modos de valuación/mercado, una insignia azul muestra el tipo especial asignado.
Type	Menú desplegable de tipo de datos: numeric , integer , character , logical , factor , Date , POSIXct
Include	Casilla de verificación — incluir esta columna como predictor en el modelo
Factor	Casilla de verificación — tratar esta columna como variable categórica
Linear	Casilla de verificación — forzar la entrada lineal únicamente (sin funciones bisagra (<i>hinge</i>))
Latent	Casilla de verificación — mantener esta columna completamente fuera del modelo (excluida de <i>earth</i> y de los datos exportados a <i>glmnetUI</i> / <i>mgcvUI</i>), reservándola para el trabajo con variables latentes
Special	Menú desplegable (solo valuación/mercado) — véase más abajo la referencia de tipos especiales de columna
T/F/NA	Tres conteos por columna: para las lógicas, TRUE/FALSE/NA; para las numéricas, $\neq 0/=0/NA$. Se muestran en rojo cuando falta más del 30 % de los valores

Una línea de ayuda encima de la tabla explica las abreviaturas: “Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`).”

Desactivación suave en lugar de eliminación. El conmutador $\times$ junto al nombre de cada variable atenúa la fila y mantiene la columna fuera del modelo sin eliminarla de los datos; cambia a un $+$ verde que vuelve a habilitar la columna. Las columnas nunca se destruyen.

6.3 Detección de tipos de datos y anulaciones

earthUI detecta automáticamente los tipos de datos al importar. Se reconocen las columnas numéricas, enteras, lógicas, de tipo factor y de fecha. Las columnas de caracteres que parecen fechas (al menos el 80 % de los valores se interpretan según formatos de fecha comunes) se clasifican como **Date**.

Puede anular cualquier detección cambiando el menú desplegable **Type**. Cuando se cambia una columna a **character** o **factor**, la casilla Factor se marca automáticamente. El cambio de tipos afecta la forma en que la columna se pasa a la función **earth()**.

6.4 Indicadores Factor y Linear

- **Factor** — Cuando está marcada, la columna se trata como variable categórica. Esto es apropiado para columnas como **style**, **area_id** o **grade**, que representan grupos discretos en lugar de mediciones continuas. Las columnas de tipo factor entran en el modelo como variables indicadoras (dummy).
- **Linear** — Cuando está marcada, la columna se fuerza a entrar en el modelo de forma lineal — no se crean funciones bisagra (hinge) (lineales por tramos) para ella. Esto es útil cuando se sabe que una variable tiene una relación estrictamente lineal con el objetivo.

6.5 Referencia de tipos especiales de columna

En los modos de valuación y mercado, el menú desplegable **Special** ofrece las siguientes opciones. Cada tipo puede asignarse como máximo a una columna (excepto **display_only**, **remarks** y **list**, que admiten varias):

Tipos de fecha y hora:

- **contract_date** — Activa el cálculo automático de **sale_age**. La columna de fecha original se reemplaza por una columna de enteros que mide los días entre la fecha de venta y la Effective Date (fecha efectiva de valuación).
- **listing_date** — Se utiliza como alternativa para calcular los Days on Market (días en el mercado; DOM = fecha de contrato – fecha de listing) cuando no se designa una columna **dom** explícita.
- **dom** — Identifica la columna de Days on Market. Se muestra en la fila APN y en la fila Date of Sale del Sales Grid (cuadro de comparación).

Tipos monetarios:

- **concessions** — Identifica las concesiones de venta (créditos del vendedor, incentivos al comprador, etc.). Se utiliza en el Sales Grid para calcular el precio de venta neto: Net SP = Sale Price – Concessions.

Tipos de tamaño y ubicación:

- **latitude** — Los valores se redondean automáticamente a 3 decimales para evitar el sobreajuste. Se utiliza para los cálculos de proximidad de Haversine (distancia del inmueble a valuar a cada comparable) y se agrupa en la fila Location del Sales Grid.
- **longitude** — Recibe el mismo tratamiento de redondeo que la latitud. Se utiliza para la proximidad y el grupo Location.
- **area** — Normalmente un identificador de vecindario o de zona. Se agrupa con la latitud y la longitud en la fila “Loc: Long | Lat | Area” del Sales Grid.
- **living_area** — Habilita los cálculos de residuos por pie cuadrado (**residual_sf** y **cqa_sf**) en el archivo de descarga.
- **lot_size** — Se agrupa en la fila “Site Size | Dimensions” del Sales Grid.
- **site_dimensions** — Se agrupa con el tamaño del lote en el Sales Grid (p. ej., “75x120”).

Tipos de antigüedad:

- **actual_age** — Se agrupa en la fila “Actual Age | Effective Age” del Sales Grid.
- **effective_age** — Se agrupa con la antigüedad real en el Sales Grid.

Tipos de visualización:

- **display_only** — La columna se incluye en las exportaciones a Excel pero se excluye por completo del ajuste del modelo. Utilícelo para campos de dirección, números de MLS, identificadores de parcela u otros datos de referencia que no deban ser predictores. Varias columnas pueden tener esta designación.

Tipos de extracción de texto (experimental):

- **remarks** — Una columna de texto libre (p. ej. **public_remarks**) que será analizada en columnas de características por el paso opcional de Prolog. No hace nada hasta que esa funcionalidad se habilita en Settings. Se admiten varias columnas. Véase el Capítulo 15.
- **list** — Una columna delimitada por comas (p. ej. **existing_structures**) que será dividida en columnas one-hot **<col>_<value>** de tipo TRUE/FALSE por el mismo paso. Se admiten varias columnas. Véase el Capítulo 15.

6.6 Objetivos múltiples

Al seleccionar más de una variable objetivo se ajusta un modelo Earth de respuesta múltiple. Cuando se seleccionan varios objetivos:

- El modelo predice todas las respuestas simultáneamente, compartiendo las funciones base entre los objetivos
- El botón **wp (response weights)** (pesos de respuesta) se activa, lo que permite asignar un peso numérico a cada objetivo
- Los modelos de varianza (**varmod.method**) se desactivan (no se admiten para respuesta múltiple)
- Las pestañas de resultados muestran métricas, ecuaciones y gráficos de diagnóstico por respuesta

***Consejo:** Las columnas designadas como **display_only** permanecen en el conjunto de datos y aparecen en la exportación a Excel, pero se excluyen por completo de la tabla de predictores. Utilice esta opción para columnas de identificadores, direcciones u otros campos de referencia.*

7 Selección de Parámetros

La Sección 4 de la barra lateral — **Earth Call Parameters** (parámetros de la llamada a `earth`) — brinda acceso a todos los argumentos aceptados por la función `earth()`. Cada parámetro cuenta con un icono azul de ayuda (?) con una explicación emergente (tooltip). Los parámetros están organizados en subsecciones plegables.

7.1 Parámetros Básicos

Parámetro	Valor predeterminado	Descripción
subset	(vacío)	Expresión de filtrado de filas. Véase “Filtrado por Subconjunto (Subset Filtering)” más abajo.
weights	NULL	Selector de columna para ponderaciones por caso (fila). Solo se listan las columnas numéricas.
wp	NULL	Ponderaciones de respuesta para modelos con múltiples variables objetivo. El botón abre un diálogo con una entrada numérica por objetivo (valor predeterminado 1.0 cada una).
keepxy	off	Conserva x, y, subset y weights en el objeto del modelo.
trace	0	Nivel de traza para la salida del ajuste (0 a 5).
glm	none	Familia GLM opcional: gaussian , binomial o poisson .
degree	1	Orden máximo de interacción. Establecer $\text{degree} \geq 2$ activa automáticamente la validación cruzada y muestra la matriz Allowed Interactions.
penalty	2	Penalización GCV por nodo. Valores más altos producen modelos más simples.

7.2 Pasada Hacia Adelante (Forward Pass)

Parámetro	Valor predeterminado	Descripción
nk	auto	Número máximo de términos antes de la poda. Se completa automáticamente con el valor recomendado para los datos cargados (véase “Valores de Parámetros Recomendados” más abajo).

Parámetro	Valor predeterminado	Descripción
thresh	0.001	Umbral del pasada hacia adelante. Valores más pequeños permiten más términos.
minspan	auto	Separación mínima entre nodos. Se completa automáticamente con el valor recomendado. Los valores negativos fijan el número máximo de nodos por predictor.
endspan	auto	Separación en los extremos — distancia mínima entre un nodo y el borde de los datos. Se completa automáticamente con el valor recomendado.
newvar.penalty	0	Penalización por introducir una nueva variable (favorece la reutilización de los predictores existentes).
fast.k	20	Número de términos padre a considerar en el algoritmo fast MARS.
fast.beta	1	Controla el factor de envejecimiento de fast MARS.

7.3 Interacciones Permitidas (Allowed Interactions)

Cuando **degree** se establece en 2 o más, aparece una matriz de interacciones debajo de los parámetros básicos. Se trata de una cuadrícula triangular superior de casillas de verificación, una por cada par de predictores. Una casilla marcada indica que se permite la interacción entre los dos predictores; desmarcarla prohíbe esa interacción específica.

Al hacer clic en el nombre de un predictor (etiqueta de fila o de columna) se alternan todas las interacciones de esa variable. Las casillas **Allow All** (permitir todas) y **Clear All** (borrar todas) en la parte superior ofrecen control en bloque. La matriz utiliza encabezados fijos, de modo que las etiquetas de columnas y filas permanecen visibles al desplazarse.

Una alerta informativa le recuerda: “Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).” (Los términos de interacción aumentan el riesgo de sobreajuste. Se ha activado la validación cruzada, de 10 pliegues.)

7.4 Poda

Parámetro	Valor predeterminado	Descripción
pmethod	backward	Método de poda: backward , none , exhaustive , forward , seqrep o cv .
nprune	NULL	Número máximo de términos después de la poda. Déjelo vacío para la selección automática.

7.5 Validación Cruzada

Parámetro	Valor predeterminado	Descripción
nfold	10	Número de pliegues de la validación cruzada. Establézcalo en 0 para desactivar la CV.
ncross	20	Número de repeticiones de la validación cruzada.
stratify	on	Estratifica las muestras de CV de modo que cada pliegue tenga una distribución similar de la respuesta.

Cuando $\text{degree} \geq 2$, la validación cruzada se activa automáticamente (nfold se establece en 10) para ayudar a protegerse del sobreajuste debido a las interacciones.

7.6 Filtrado por Subconjunto (Subset Filtering)

La entrada de texto **subset** acepta una expresión de R que filtra qué filas se utilizan para el ajuste del modelo. Puede escribir una expresión directamente (p. ej., `sale_age < 365 & area_id == 460`) o utilizar el botón **Build filter...** (construir filtro) para construirla visualmente.

Build Filter Dialog (diálogo de construcción de filtros) — Haga clic en “Build filter...” para abrir un diálogo guiado. Cada fila de condición tiene una lista desplegable de columnas, un operador (<, >, <=, >=, ==, !=) y una entrada de valor que se adapta al tipo de columna: entrada numérica para números, selector de fechas para fechas y lista desplegable de valores únicos para columnas de tipo carácter/factor. Las condiciones se unen con conectores AND (&) u OR (|). Una vista previa en la parte inferior muestra la expresión y cuántas filas coinciden. Haga clic en **Apply** (aplicar) para insertar la expresión en la entrada de texto.

Las columnas de fecha deben usar los envoltorios `as.Date("...")` o `as.POSIXct("...")` en las expresiones manuales. El diálogo Build Filter lo maneja automáticamente.

El filtrado por subconjunto no es destructivo — las filas excluidas permanecen en el conjunto de datos y reciben predicciones en la exportación a Excel.

***Consejo:** Utilice paréntesis para agrupar las condiciones OR cuando las combine con AND. Sin paréntesis, $A \ \& \ B \ | \ C$ se evalúa como $(A \ \& \ B) \ | \ C$, lo cual puede no ser lo que usted pretende.*

7.7 Valores de Parámetros Recomendados

earthUI muestra un valor recomendado debajo de los parámetros clave de la Sección 4. Estas recomendaciones se actualizan de forma reactiva en función del número de filas de ajuste (n) y de los predictores seleccionados (p). Las fórmulas se derivan del artículo de MARS de Friedman, de los algoritmos internos de earth y de pruebas empíricas. Los campos **nk**, **minspan** y **endspan** se completan previamente con sus valores recomendados al cargarse los datos; una edición manual se conserva.

7.7.1. Parámetros de la Pasada Hacia Adelante

nk (máximo de términos antes de la poda):

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

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

El valor predeterminado de `earth`, $\min(200, \max(20, 2p)) + 1$, no tiene en cuenta el tamaño del conjunto de datos y puede resultar demasiado pequeño para conjuntos grandes, restringiendo la pasada hacia adelante antes de que explore adecuadamente todos los predictores. El término $\lfloor n/10 \rfloor$ permite aproximadamente un término por cada 10 observaciones — una regla práctica estándar para evitar la sobreparametrización. El tope de 100 evita rendimientos decrecientes.

minspan (mínimo de observaciones entre nodos):

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

endspan (mínimo de observaciones desde los límites de los datos):

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

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

El `minspan` calculado automáticamente por `earth` utiliza la ecuación 43 de Friedman: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, que escala como $\ln(np)$ y crece demasiado lentamente para conjuntos de datos grandes. Con $n=1500$ produce solo ≈ 7 , lo que da aproximadamente 245 ubicaciones candidatas de nodo por predictor continuo — demasiadas, lo que permite que la pasada hacia adelante ajuste el ruido. La fórmula recomendada apunta a aproximadamente 100 candidatos por predictor para n grande, mientras que se remite al cálculo automático de `earth` para n pequeño (donde está bien probado). `Endspan` se fija ligeramente por encima de `minspan` para proporcionar protección adicional en los límites, lo cual ayuda cuando los datos tienen colas delgadas (algo común en bienes raíces).

penalty (penalización GCV por nodo):

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

Este es el propio valor predeterminado de `earth` — 2 para modelos aditivos, 3 para modelos con interacciones (la penalización más alta compensa el espacio de búsqueda más grande).

newvar.penalty (penalización por introducir un nuevo predictor):

Recomendado: **0.1** cuando los predictores están correlacionados (p. ej., la superficie habitable con el número de recámaras, el número de baños o el tamaño del lote). Esto sesga la pasada

hacia adelante hacia la reutilización de los predictores ya presentes en el modelo en lugar de introducir alternativas correlacionadas. El mecanismo: la mejora en RSS de cada nuevo predictor se multiplica por $1/(1 + \text{penalty})$. Con `newvar.penalty = 0.1`, una nueva variable debe mejorar el RSS al menos un 10 % más que otro nodo sobre una variable existente. Esto produce modelos más simples con menos predictores sin afectar las estimaciones finales de los coeficientes (la penalización se elimina después de la selección). Déjelo en 0 si los predictores no están correlacionados. Tenga en cuenta que `earth` no admite un `newvar.penalty` distinto de cero junto con ponderaciones por caso — `earthUI` desactiva la entrada (forzando 0) siempre que se designa una columna de ponderación.

7.7.2. Parámetros de Poda

pmethod: Recomendado: **backward**. La poda hacia atrás basada en GCV, la predeterminada, es determinista — los mismos datos producen siempre el mismo modelo, sin dependencia de semillas aleatorias. Esto es importante para la reproducibilidad, especialmente en el trabajo de valuación. El método `cv` utiliza validación cruzada para la poda, lo que introduce dependencia de la semilla.

nprune: Recomendado: **dejar vacío** (NULL). Deje que GCV seleccione el número óptimo de términos. Establecer `nprune` impone un tope rígido que anula el criterio de GCV.

7.7.3. Parámetros de Validación Cruzada

nfold (pliegues de CV): recomendado por banda de tamaño muestral:

n (filas de ajuste)	nfold
0–299	5
300–399	6
400–499	7
500–599	8
600–699	9
700+	10

Más pliegues entrenan cada pliegue con más datos (una estimación menos sesgada), pero reducen el pliegue de prueba reservado. La tabla escala gradualmente hasta el estándar de 10 pliegues a medida que crece la muestra, manteniendo cada pliegue de prueba lo suficientemente grande (aproximadamente 50+ filas) para una métrica estable por pliegue, y recurre a 5 pliegues para muestras pequeñas — donde la estimación de CVR^2 sigue siendo ruidosa, por lo que conviene aumentar `ncross` (p. ej. 15–20) para promediar la varianza. El tope de 10 refleja que cada pliegue adicional supone un ajuste completo de `earth` a cambio de una ganancia diagnóstica insignificante. Con `pmethod = "backward"`, la validación cruzada no afecta al modelo — solo calcula el CVR^2 diagnóstico y proporciona residuos para el modelo de varianza.

ncross (repeticiones de CV): recomendado por banda de tamaño muestral:

n (filas de ajuste)	ncross
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

Las repeticiones promedian la suerte en la partición de pliegues para estabilizar el CVR²; las muestras más grandes requieren menos repeticiones porque una sola validación cruzada ya es más estable. Los residuos de la CV repetida también alimentan el modelo de varianza.

7.7.4. Modelo de Varianza

varmod.method: Recomendado: **lm** — el valor predeterminado de Milborrow: un modelo de varianza lineal robusto para intervalos de predicción (la dispersión se estima a partir de residuos ruidosos, por lo que un modelo simple generaliza mejor). Si **lm** no converge con sus datos, pruebe **rlm** o **const** (homocedástico, siempre converge), o recurra a **earth** (que también maneja una dispersión no lineal). Requiere **nfold** > 1 y **ncross** > 1.

7.7.5. Archivo de Salida de Earth

Después de cada ajuste exitoso, earthUI escribe <filename>_earth_output_<timestamp>.txt en la carpeta de salida. Este archivo contiene los términos del modelo, estadísticas de resumen (R^2 , GRSq, CVRSq), detalles del modelo de varianza y el registro de traza. Se crea un archivo por ajuste, lo que proporciona un registro acumulativo para comparar configuraciones de parámetros.

7.8 Valores Predeterminados de Configuración

Un botón de opción en la parte superior de la Sección 4 controla qué valores predeterminados se cargan:

- **Use last settings for input file** (usar la última configuración para el archivo de entrada) — restaura la configuración que usted utilizó la última vez con este archivo en particular
- **Use default settings** (usar la configuración predeterminada) — aplica sus valores predeterminados personalizados guardados
- **Earth defaults** (valores predeterminados de earth) — restablece todos los parámetros a los valores de fábrica

Haga clic en **Save current as default** (guardar la configuración actual como predeterminada) en la parte inferior de la Sección 4 para almacenar los parámetros actuales de earth y la matriz Allowed Interactions; el botón correspondiente en la Sección 3 almacena la configuración de la variable objetivo, la fecha efectiva y los predictores. Ambos se guardan por proyecto y propósito en la base de datos del proyecto.

8 Ajuste del modelo Earth

8.1 El botón Fit

La sección 5 de la barra lateral contiene el botón **Fit Earth Model** (ajustar el modelo earth) junto con un campo **Random seed** (semilla aleatoria). La semilla hace reproducible la asignación de los pliegues de validación cruzada — un menú desplegable recuerda las últimas 10 semillas utilizadas, y la semilla empleada queda registrada en la pestaña **Earth Output** después del ajuste. Antes de ajustar, earthUI valida su configuración — debe estar seleccionada una variable objetivo y debe incluirse al menos un predictor. En los modos de valuación y de mercado, las columnas de latitud y longitud se redondean, y la antigüedad de la venta se calcula a partir de la fecha efectiva y la fecha de contrato (si están designadas).

8.2 Ventana modal de ajuste y salida de traza

Al hacer clic en **Fit**, aparece una superposición modal oscura con:

- **Encabezado** — “Fitting Earth Model” con un cronómetro que cuenta los segundos transcurridos y un botón **Abort** (abortar) que cancela el ajuste en segundo plano
- **Trace log** (registro de traza) — un área de texto monoespaciada y desplazable que muestra en tiempo real la salida de la función `earth()`: información del conjunto de datos, progreso del pasada hacia adelante, pliegues de validación cruzada y tiempo de finalización
- **Codificación por colores** — líneas normales en verde, líneas de validación cruzada en amarillo, errores en rojo

earthUI ajusta los modelos de forma asíncrona mediante `callr::r_bg()`, que ejecuta el cálculo en un proceso de R separado. Esto mantiene la aplicación con capacidad de respuesta durante ajustes de larga duración. Un observador con sondeo cada 300ms lee la salida del proceso en segundo plano y la transmite al Trace log.

Cuando el ajuste concluye, aparece un mensaje “Done in X.Xs”, se agrega un botón de cierre (X) a la ventana modal y se retira el fondo oscurecido para que las pestañas de resultados puedan renderizarse bajo demanda. La ventana modal se cierra luego automáticamente una vez que las salidas visibles terminan de renderizarse — o bien usted puede cerrarla antes.

En caso de éxito, se añade una marca de verificación verde al botón **Fit**. Si ocurre un error, el mensaje de error se muestra en el Trace log y aparece una notificación.

8.3 Registro del ajuste

Cada ajuste (exitoso o fallido) escribe un archivo de registro en su carpeta de salida:

- **Nombre del archivo:** `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Contenido:** marca de tiempo, toda la salida de traza del proceso de ajuste y cualquier mensaje de error
- **Ubicación:** la carpeta de salida del proyecto activo

***Sugerencia:** Si callr no está instalado, earthUI recurre al ajuste síncrono con una barra de progreso simple. Instale callr para disfrutar de la experiencia completa de salida de traza.*

9 Descarga de datos

Después del ajuste, la sección 6 de la barra lateral — titulada “Download Estimated Target Variable(s) & Residuals” en el modo general y “Download Estimated Sale Prices & Residuals” en los modos de valuación y de mercado — descarga un archivo de Excel con predicciones y diagnósticos. Esta salida se utiliza en el Paso 7 (RCA) para asignar una calificación CQA (estado, calidad y atractivo; Condition-Quality-Appeal) al inmueble a valorar. La salida se ordena por **residual_sf** y **cqa_sf** para ayudarlo a evaluar dónde se ubica el inmueble a valorar dentro del ranking. Si el modelo es de buena calidad, los inmuebles deberían quedar ordenados desde el menos atractivo hasta el más atractivo según las características residuales que no entraron en la regresión. El valor central debería ser aproximadamente 0, la mitad inferior valores negativos y la mitad superior valores positivos. Debería encontrar las viviendas de peor calidad, o “fixers” (para remodelar), cerca del final del ranking y las viviendas más atractivas en la parte superior. Habitualmente habrá excepciones por anomalías tales como ejecuciones hipotecarias, ventas cortas (short sales), ventas sucesorias (relacionadas con herencias) y ventas rápidas motivadas por un cambio de empleo u otras razones. La investigación de las anomalías suele revelar una razón pertinente para la anomalía en el precio.

La etiqueta del botón se adapta al modo de propósito:

- General/Mercado: **Download Output (Excel)**
- Valuación: **Download Intermediate Output (Excel)**

El formato del nombre de archivo es `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`.

9.1 Columnas de salida

Para cada variable objetivo se anexan las siguientes columnas:

Columna	Descripción
est_<target>	Predicción del modelo, p. ej. est_sale_price (1 decimal)
residual	Valor real menos el predicho (1 decimal)
cqa	Puntuación Condition-Quality-Appeal (2 decimales, rango 0–10)
residual_sf	Residuo dividido entre la superficie habitable (si está designada, 1 decimal)
cqa_sf	CQA calculado a partir del ranking mediante residual_sf (2 decimales)
<variable>_contribution	Valor de la contribución por función g (1 decimal, una columna por función g)
basis	Contribución al valor del intercepto, igual para todos los inmuebles (1 decimal)
calc_residual	Valor real menos (basis + todas las contribuciones) — columna de verificación (1 decimal)

Para modelos con múltiples variables objetivo, se agrega un sufijo **_<i>** para distinguir las columnas de cada objetivo.

9.2 Orden de columnas y formato de Excel

Las columnas de ranking se colocan en la **posición más a la izquierda** del archivo de salida, en este orden: **residual_sf**, **cqa_sf**, **residual**, **cqa**. Esto facilita recorrer la lista ordenada y evaluar dónde se ubica el inmueble a valorar dentro de la distribución de CQA.

A estas columnas se les aplica formato numérico de Excel:

Columna	Formato	Ejemplo
<code>residual_sf</code>	Numérico, 2 decimales	12.34
<code>cqa_sf</code>	Numérico, 2 decimales	7.25
<code>residual</code>	Numérico, 0 decimales	15,200
<code>cqa</code>	Numérico, 2 decimales	6.80

9.3 Puntuaciones CQA

La puntuación CQA (Condition-Quality-Appeal) clasifica el residuo de cada fila frente a todos los demás residuos. Para una fila dada, el CQA es el porcentaje de filas con un residuo con signo menor, multiplicado por 10. Esto produce una escala de 0–10 en la que:

- **CQA alto** (≈ 9 –10) — el inmueble se vendió por mucho más de lo que el modelo predijo (residuo positivo grande)
- **CQA bajo** (≈ 0 –1) — el inmueble se vendió por mucho menos de lo predicho (residuo negativo grande)
- **CQA ≈ 5** — el residuo se encuentra cerca de la mediana de todos los residuos

Cuando se designa una columna `living_area`, `cqa_sf` proporciona el mismo ranking con base en los residuos por pie cuadrado.

9.4 Ordenamiento y fila del inmueble a valuar

En los modos de valuación y de mercado, las filas de comparables se ordenan por `residual_sf` de forma descendente (o por `residual` si no se designó superficie habitable). La fila del inmueble a valuar (fila 1) permanece en la posición 1 y no se ordena. En el modo general, las filas se exportan en su orden original.

Este ordenamiento, combinado con las columnas de ranking situadas más a la izquierda, permite al valuator recorrer rápidamente los comparables desde el más sobrepredicho hasta el más subpredicho, y evaluar dónde cae la puntuación CQA asignada al inmueble a valuar dentro de esa distribución.

Los niveles de las variables factor en los datos de predicción se alinean con los datos de entrenamiento antes de llamar a `predict()`. Las filas con niveles de factor no vistos producirán predicciones NA.

***Sugerencia:** La columna `calc_residual` siempre debería ser igual a `residual` (dentro del margen de redondeo). Si difieren de manera significativa, ello indica un problema de cálculo que conviene investigar.*

9.5 Exportación automática para mgcvUI

En cada ajuste exitoso con `degree` ≤ 2 , earthUI guarda automáticamente el objeto de resultado completo como un archivo `.rds` en la carpeta de salida del proyecto. El nombre del archivo sigue el patrón `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

Este archivo puede cargarse con `readRDS()` en el módulo de posprocesamiento mgcv (mgcv Post Processing Module) — incluido en el paquete earthUI y que se inicia con `earthUI::launch_mgcv()`. La rutina GAM utiliza las ubicaciones de los nodos y las funciones base del modelo earth como puntos de partida para los términos suaves del GAM, lo que permite una transición fluida del modelado MARS al modelado GAM.

Los modelos con `degree` > 2 se omiten porque mgcvUI solo admite interacciones por pares. Cuando earthUI se inicia en modo Trilogy (desde la Trilogy UI), una sección adicional **Lock**

Model Output (bloquear la salida del modelo) le permite bloquear el archivo `.rds` de un ajuste determinado — junto con su base de predictores y la configuración compartida de CQA — como el modelo earth canónico que glmnetUI y mgcvUI importan.

9.6 Importación en glmnetUI

El mismo archivo `.rds` guardado para la rutina GAM también puede importarse en el **elastic net Post Processing Module** (módulo de posprocesamiento de red elástica) — incluido igualmente en el paquete earthUI y que se inicia con `earthUI::launch_glmnet()`. En la aplicación glmnet, diríjase a **Section 2: Import from earthUI** y utilice el botón **Browse** para seleccionar el archivo `.rds`.

glmnetUI utiliza el enfoque estándar de `earth::model.matrix()`: las funciones base del modelo earth (funciones bisagra, interacciones y términos lineales) se convierten en las columnas de la matriz de diseño de glmnet. glmnet realiza entonces una regresión regularizada sobre estas columnas base, seleccionándolas y contrayéndolas mediante lasso/red elástica. Esto combina la construcción adaptativa de bases de earth con la regularización de glmnet.

9.6.1. Beneficios clave

- **Interacciones interpretables:** las interacciones basadas en funciones bisagra de earth (incluidos los términos de 3 vías) tienen un significado geométrico claro, a diferencia de las interacciones nativas de producto cruzado de glmnet, que resultan difíciles de explicar en contextos judiciales o de auditoría.
- **No linealidad automática:** las funciones bisagra de earth capturan relaciones no lineales que un modelo glmnet estándar con predictores sin transformar pasaría por alto.
- **Selección de variables:** glmnet puede podar aún más la base de earth, descartando los términos bisagra que no contribuyen.

9.6.2. Importante: la columna de pesos debe coincidir

Advertencia

Si earthUI utiliza una columna de pesos (p. ej., para excluir la fila del inmueble a valuar con `weight=0`), la misma columna de pesos debe designarse en la configuración de variables de glmnetUI. Si las configuraciones de pesos difieren, los conteos de filas no coincidirán y la base de earth no podrá anexarse a la matriz del modelo de glmnet.

9.6.3. Flujo de trabajo

1. Ajuste un modelo earth en earthUI (de cualquier grado).
2. El archivo de resultados `.rds` se guarda automáticamente en la carpeta de salida del proyecto.
3. En glmnetUI, abra **Section 2: Import from earthUI** y navegue hasta el archivo `.rds`.
4. Verifique que se utilicen el **mismo archivo de datos** y la **misma columna de pesos** en ambas aplicaciones.
5. Configure los parámetros de glmnet en la Sección 5 (la Interaction Matrix se deshabilita automáticamente cuando hay una importación de earthUI activa).
6. Haga clic en **Fit Model**. Las columnas de la base de earth aparecen en las pestañas Equation, Coefficients y Contributions.

10 Cálculos RCA y descarga

El botón **Calculate RCA Adjustments & Download** (calcular ajustes RCA y descargar) aparece en la sección 7 de la barra lateral, visible únicamente en el modo de avalúo después de que se ha ajustado un modelo. RCA (Residual Constraint Approach, es.: enfoque por restricciones residuales) produce ajustes por comparable derivados del mercado, relativos al inmueble a valorar.

10.1 Apertura del diálogo RCA

Al hacer clic en el botón se abre un pequeño diálogo modal con:

- **Score type** (tipo de puntuación) — botones de opción para elegir **CQA** o **CQA per SF** (la opción por SF está disponible solo cuando se ha designado una columna `living_area`). Si elige “CQA per SF,” entonces, con base en la puntuación CQA que asigne al inmueble a valorar, su puntuación residual será su superficie habitable multiplicada por el `residual_sf` que corresponda a la puntuación `CQA_SF` dada.
- **CQA value** (valor CQA) — campo de entrada para la puntuación CQA del inmueble a valorar (0.00–10.00). Esta puntuación representa el lugar que, a su juicio, ocupa el inmueble a valorar en la distribución de calidad de los comparables.
- Botón **Generate** (generar) — calcula la salida RCA y la descarga como archivo de Excel. El botón permanece deshabilitado hasta que se introduce una puntuación válida.

10.2 Interpolación de la puntuación CQA

Cuando hace clic en Generate, earthUI interpola el residuo del inmueble a valorar a partir de los pares CQA/residuo de los comparables:

1. Las puntuaciones CQA y los residuos de los comparables se ordenan por CQA de forma ascendente
2. La interpolación lineal (`stats::approx()`) transforma el valor CQA introducido en un residuo
3. Si se utiliza CQA per SF, el residuo por SF se convierte de nuevo en un residuo total multiplicándolo por la superficie habitable del inmueble a valorar
4. El valor estimado del inmueble a valorar se calcula como: predicción del modelo + residuo interpolado

10.3 Columnas de salida

El archivo de Excel del RCA (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) incluye todas las columnas de salida intermedias más:

Columna	Descripción
<code>subject_value</code>	Predicción del modelo + residuo interpolado (solo fila 1)
<code>subject_cqa</code>	La puntuación CQA que usted introdujo (solo fila 1)
<code><variable>_adjustment</code>	Contribución del inmueble a valorar menos contribución del comparable (por función g)
<code>residual_adjustment</code>	Residuo del inmueble a valorar menos residuo del comparable
<code>net_adjustments</code>	Suma de todos los ajustes (contribución + residuo)
<code>gross_adjustments</code>	Suma de los valores absolutos de todos los ajustes
<code>residual_pct</code>	Ajuste residual como fracción del precio de venta del comparable

Columna	Descripción
net_adj_pct	Ajustes netos como fracción del precio de venta del comparable
gross_adj_pct	Ajustes brutos como fracción del precio de venta del comparable
adjusted_sale_price	Precio de venta del comparable + ajustes netos

Las columnas de ajustes indican, para cada comparable, cuánto de la diferencia atribuye el modelo a cada variable. **adjusted_sale_price** es el precio de venta del comparable después de aplicar todos los ajustes derivados del modelo — un conjunto de precios de venta ajustados que debería agruparse en torno al valor estimado del inmueble a valorar.

10.4 La pestaña RCA Adjustments

Una vez calculada la salida del RCA, la pestaña **RCA Adjustments** (ajustes RCA) del panel principal muestra histogramas de las distribuciones de Residual Adj %, Net Adj % y Gross Adj % entre los comparables (un conjunto por variable objetivo en los modelos multiobjetivo), lo que ofrece una comprobación visual rápida de cuán estrechamente se agrupan los ajustes.

Sugerencia: Una puntuación CQA de 5.00 sitúa el inmueble a valorar en la mediana de los comparables. Las puntuaciones superiores a 5 indican que el inmueble a valorar es de calidad superior al promedio; las inferiores a 5 indican calidad inferior al promedio.

11 Sales Comparison Grid

11.1 Descripción general y propósito

El **Sales Comparison Grid** (cuadro de comparación de ventas) es un libro de Excel generado en el modo de avalúo (sección 8 de la barra lateral) después de calcular los ajustes del RCA. Presenta el inmueble a valuar junto con los comparables seleccionados en un formato de cuadrícula estructurado, con fórmulas de Excel que calculan automáticamente los ajustes y un precio de venta ajustado para cada comparable.

La cuadrícula está diseñada para el expediente de trabajo del valuador — combina los ajustes derivados de la regresión del modelo Earth con celdas editables en las que el valuador puede asignar el residuo CQA a características específicas del inmueble. El nombre del archivo de salida es **SalesGrid_<YYYYMMDD_HHMMSS>.xlsx**.

11.2 Modal de selección de comparables

Al hacer clic en “Generate Sales Grid & Download” (generar y descargar el Sales Grid) se abre un cuadro de diálogo modal donde se seleccionan los comparables que se desea incluir:

- Los **comparables recomendados** aparecen premarcados. Se trata de comparables con un porcentaje de ajuste bruto inferior al 25 % del precio de venta, ordenados por antigüedad de la venta en orden ascendente (las ventas más recientes primero) y limitados a 30.
- Los **comparables adicionales** se listan a continuación, sin marcar de forma predeterminada. Estos presentan ajustes brutos del 25 % o más, pero siguen disponibles para su selección (se muestran hasta 50, ordenados por porcentaje de ajuste bruto ascendente).
- Puede seleccionarse un máximo de **30 comparables**, lo que produce hasta **10 hojas** (3 comparables por hoja).

Tras confirmar la selección, los comparables seleccionados se ordenan por porcentaje de ajuste bruto ascendente, de modo que los comparables más similares aparecen en la Hoja 1.

11.3 Estructura de la cuadrícula

Cada hoja tiene una estructura de 20 columnas que acomoda el inmueble a valuar más 3 comparables. Las columnas para cada entidad son:

Entidad	Columnas (5 por entidad)
Subject (inmueble a valuar)	Label, Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC)
Cada comparable	Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC), Adjustment

Las filas, de arriba hacia abajo, son:

1. **Title** — nombre de la hoja (p. ej., “Intermediate Sales Comparable Grid — Sheet 1 of 3”)
2. **Headers** — encabezados de columna “Subject” y de los comparables, con la dirección
3. **Address** — dirección completa del inmueble
4. **APN | MLS# | DOM | Subj.Prox** — número de parcela, identificador del listing, días en el mercado y distancia Haversine (en millas) respecto del inmueble a valuar
5. **Sales Price | Concess. | Net SP** — precio de venta, concesiones y fórmula del precio de venta neto
6. Fila de encabezado **Regression Features** (características de regresión)

7. **BASE VALUE** — el intercepto del modelo (valor base)
8. **Date of Sale | OffMkt | OnMkt** — fecha de contrato, antigüedad de la venta y DOM
9. **Filas agrupadas** (condicionales) — grupos de Location, Site Size y/o Age
10. **Filas de variables del modelo** — una fila por predictor (excluidas las variables agrupadas)
11. **Fila separadora en blanco**
12. Fila de encabezado **Residual Features** (características residuales)
13. **CQA|Residual** — puntuación CQA + fórmula del residuo restante
14. **Filas de características residuales** — filas con nombre + filas en blanco para captura por parte del valuador
15. **Total VC / Net Adjustment** (ajuste neto)
16. **Net Adjustment %**
17. **Gross Adjustment %** (porcentaje de ajuste bruto)
18. **Adjusted Sale Price** (precio de venta ajustado) — fila de fórmula
19. Pie de página de **Copyright**

11.4 Precio de venta, concesiones y Net SP

La fila 5 muestra tres valores para cada comparable:

- **Sale Price** — el precio de venta del comparable según los datos
- **Concessions** — el valor de la columna designada como **concessions** (0 si no se ha designado)
- **Net SP** — una fórmula de Excel: precio de venta – concesiones

La columna del inmueble a valorar muestra “N/A” como precio de venta (dado que se desconoce) y el valor de las concesiones, si está disponible.

11.5 Filas agrupadas (Location, Site, Age)

Cuando se designan ciertos tipos especiales y esas variables aparecen en el modelo, earthUI crea filas agrupadas que combinan variables relacionadas:

Loc: Long | Lat | Area — Aparece cuando cualquiera de **longitude**, **latitude** o **area** está en el modelo. Muestra los valores factuales de cada variable constituyente, una contribución al valor combinada (suma de las VC individuales) y, para los comparables, un ajuste (VC combinada del inmueble a valorar – VC combinada del comparable). Las celdas de VC se presentan con fondo azul claro.

Site Size | Dimensions — Aparece cuando **lot_size** o **site_dimensions** están en el modelo. Misma estructura que el grupo de Location.

Actual Age | Effective Age — Aparece cuando **actual_age** o **effective_age** están en el modelo. Misma estructura.

Las variables consumidas por las filas agrupadas quedan **excluidas de las filas individuales de variables del modelo** que aparecen debajo, lo que evita la doble contabilización en los totales de ajustes.

11.6 Filas de variables del modelo

Debajo de las filas agrupadas (o directamente debajo de BASE VALUE si no hay filas agrupadas), una fila por cada predictor restante del modelo muestra:

- **Valores factuales** para el inmueble a valorar y cada comparable (los valores reales de los datos)
- **Value Contribution (VC)** — la contribución de la función g al valor predicho para esa fila
- **Adjustment** (solo comparables) — VC del inmueble a valorar menos VC del comparable

11.7 CQA|Residual y filas de características residuales

La fila **CQA|Residual** muestra la puntuación CQA de cada inmueble y contiene una fórmula para el **residuo restante** (Remaining Residual) — la porción del residuo aún no asignada a características específicas. La fórmula es:

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

Debajo se encuentran las **filas de características residuales** — 5 filas con nombre (Location, View/Appeal, Condition, Quality, Other) más 6 filas en blanco. Son celdas de entrada donde el valuador captura contribuciones al valor correspondientes a características no capturadas por el modelo. A medida que el valuador introduce valores, la fórmula del residuo restante disminuye automáticamente.

Para cada comparable, la columna de ajuste contiene una fórmula: VC de la característica del inmueble a valorar — VC de la característica del comparable.

11.8 Fórmula del Adjusted Sale Price

La fila **Adjusted Sale Price** contiene fórmulas de Excel:

- **Inmueble a valorar:** suma de todas las celdas de contribución al valor desde BASE VALUE hasta la última fila de características residuales. Representa la predicción total del modelo para el inmueble a valorar más las características residuales asignadas por el valuador.
- **Comparables:** Net SP + suma de todas las celdas de ajuste desde la primera fila agrupada o de variable hasta la última fila de características residuales. Esto proporciona el precio de venta del comparable después de todos los ajustes derivados de la regresión y los introducidos por el valuador.

11.9 Protección de hojas y celdas editables

Cada hoja está **protegida** para evitar la modificación accidental de fórmulas y datos. La protección bloquea:

- Todas las celdas de fórmula (Net SP, residuo restante, ajustes, precio de venta ajustado, etc.)
- Todas las celdas de datos (direcciones, precios de venta, valores factuales, contribuciones al valor provenientes del modelo)
- Etiquetas y encabezados

Las únicas celdas **desbloqueadas** (editables) son las entradas de contribución al valor de las características residuales — las celdas bajo la fila **CQA|Residual** donde el valuador captura los desgloses. Estas celdas se presentan con fondo amarillo claro para indicar que son editables.

11.10 Trabajo con la cuadrícula en Excel

1. **Abra el archivo** en Excel o en una aplicación de hoja de cálculo compatible.
2. **Revise los ajustes de la regresión** — los ajustes derivados del modelo vienen prellenados y bloqueados.
3. **Asigne el residuo** — en las filas de características residuales (celdas amarillas), capture su estimación de qué parte del residuo restante es atribuible a cada característica (Location, View, Condition, Quality, etc.). La fórmula del residuo restante disminuirá a medida que asigne valores.
4. **Verifique el Adjusted Sale Price** — esta fórmula se actualiza automáticamente conforme se capturan los valores de las características residuales.

5. **Hojas múltiples** — si seleccionó más de 3 comparables, navegue entre las hojas. Cada hoja contiene el mismo inmueble a valuar en la columna izquierda con 3 comparables distintos.

Sugerencia: el objetivo es asignar el residuo CQA entre las filas de características hasta que el residuo restante se aproxime a cero. Esto demuestra que es posible explicar la diferencia entre la estimación de regresión de cada comparable y su precio de venta real.

12 Descarga de informes

Después de ajustar un modelo, la generación de informes es un flujo de trabajo de Quarto en dos pasos: **Generate Quarto Report** (generar informe Quarto; sección 9 de la barra lateral en el modo de avalúo, 7 en los demás casos) escribe un paquete de informe autocontenido, y **Convert Quarto Report** (convertir informe Quarto; sección 10) lo renderiza — al igual que cualquier otro archivo Quarto — a los formatos finales.

12.1 Generate Quarto Report

Al hacer clic en **Generate Quarto Report** se escribe un paquete Quarto autocontenido bajo la carpeta de salida del proyecto, en `<base>_qmd/`, que contiene el archivo fuente `.qmd` ya poblado, todos los recursos gráficos pregenerados (PNG y PDF), los datos del informe y una plantilla de referencia de Word. En este paso no se realiza ninguna renderización — el paquete puede editarse a mano, combinarse con otras fuentes Quarto mediante *includes* de Quarto, o previsualizarse de manera independiente con `quarto preview`.

12.2 Convert Quarto Report

La sección 10 ofrece un selector de archivos (que apunta por defecto al paquete generado más recientemente en el proyecto actual — aunque puede navegarse hasta cualquier archivo `.qmd`) y casillas de verificación de formato:

- **HTML** — un archivo `.html` con renderización matemática mediante KaTeX, imágenes incrustadas y una tabla de contenido. Utiliza el tema Flatly de Bootstrap y la fuente Roboto Condensed.
- **Word** — un archivo `.docx` apto para edición y distribución. Incluye una tabla de contenido y números de página, y utiliza una plantilla de referencia personalizada para un estilo uniforme.
- **PDF** — compuesto tipográficamente con LuaLaTeX para una salida de calidad profesional. El tamaño de papel sigue la configuración regional (Letter o A4). Incluye páginas apaisadas para matrices de interacción grandes y utiliza Roboto Condensed con Latin Modern Math para las ecuaciones. La opción PDF se ofrece únicamente cuando se detecta una distribución de LaTeX.

Haga clic en **Convert** para renderizar los formatos seleccionados. La renderización se ejecuta en segundo plano — un cuadro de diálogo modal muestra el tiempo transcurrido y el progreso de Quarto mientras la aplicación permanece receptiva. Una notificación confirma la ubicación del archivo al finalizar.

Todas las operaciones (ajuste de modelos, descargas de datos, cálculos RCA, generación de Sales Grids y renderización de informes) se registran con marcas de tiempo de inicio/fin y tiempos transcurridos en `<datafile>_earthui_log.txt` dentro de la carpeta de salida, para diagnóstico de problemas y supervisión del rendimiento.

12.3 Contenido del informe

Todo informe generado incluye las secciones siguientes:

1. **Descripción del conjunto de datos** — número de observaciones, variable(s) objetivo, predictores utilizados, predictores categóricos
2. **Especificación del modelo** — grado, estado de la validación cruzada, número de términos, predictores incluidos en el modelo
3. **Interacciones permitidas** (solo con grado ≥ 2) — la matriz de interacciones con marcas de verificación, formateada según el tipo de salida

4. **Resumen de resultados** — R^2 , GRSq, GCV, RSS y CV R^2 (si la validación cruzada estaba habilitada)
5. **Ecuación del modelo** — la ecuación completa del modelo Earth en notación LaTeX
6. **Coeficientes y funciones base** — tabla de todos los términos, sus coeficientes y las funciones base que los definen
7. **Importancia de las variables** — gráfico de barras y tabla ordenada de las puntuaciones de importancia de los predictores
8. **Contribuciones de las funciones g** — gráficos para cada función g: gráficos de líneas para los términos univariados, gráficos de perspectiva 3D y de contorno para las interacciones bivariadas
9. **Matriz de correlación** — mapa de calor de las correlaciones entre predictores
10. **Diagnósticos** — gráficos de Residuals vs Fitted (residuos frente a valores ajustados), Normal Q-Q y Actual vs Predicted (valores reales frente a predichos)
11. **Descomposición ANOVA** — tabla que muestra la contribución de cada función base al RSS
12. **Salida de Earth** — salida de texto sin procesar de `earth::summary()`, incluidos los detalles de la pasada de poda

En los modelos de respuesta múltiple, las secciones 4–8 y 10 se repiten para cada variable objetivo.

Sugerencia: Los informes requieren el paquete de R *quarto* y una instalación de Quarto. Para la salida en PDF debe instalarse además una distribución de LaTeX (p. ej., mediante `tinytex::install_tinytex()`) — de lo contrario, la casilla PDF permanece oculta.

13 Los módulos de posprocesamiento: Elastic Net y GAM

El paquete earthUI incluye dos módulos de posprocesamiento — una rutina de elastic net (red elástica) construida sobre `glmnet` y una rutina GAM construida sobre `mgcv`. Cada una se trata en profundidad en su propio artículo de este número; esta sección explica *para qué* sirven dentro del flujo de trabajo de earthUI y cómo cooperan las tres rutinas.

13.1 ¿Por qué ajustes de apoyo?

El modelo earth produce la conclusión de valor. Los módulos de posprocesamiento existen para **corroborarla**: dos estimaciones adicionales del mismo inmueble, producidas por motores metodológicamente distintos y ajustadas a los mismos datos del proyecto. Cuando una reestimación mediante elastic net de la base de earth y un suavizado GAM independiente llegan a un valor sustancialmente igual, la conclusión del valuador ya no descansa sobre los supuestos de un solo modelo — una ventaja significativa en contextos de revisión y litigio. Cuando *discrepan*, esa discrepancia es en sí misma diagnóstica: por lo general apunta a datos escasos, a un término excesivamente flexible o a un comparable atípico que conviene reexaminar. Ninguno de los módulos prevalece sobre el modelo earth; constituyen evidencia de apoyo, presentada deliberadamente sin ponderación.

13.2 Ejecución y estado compartido

Cada rutina es una aplicación Shiny local independiente:

- `earthUI::launch_glmnet()` — la aplicación de elastic net, en el puerto 7879.
- `earthUI::launch_mgcv()` — la aplicación GAM, en el puerto 7880.

Las tres aplicaciones leen el mismo árbol de proyecto y la misma base de datos de configuración, de modo que el proyecto abierto en earthUI queda inmediatamente disponible en los módulos — junto con los campos de arrastre que earthUI guarda para él: la Response (variable objetivo), la Effective Date (fecha efectiva), el tipo y valor de CQA del inmueble a valuar para el RCA, y la designación de superficie habitable. No es necesario reingresarlos; los módulos los leen para que los mismos valores fluyan a través de los tres ajustes.

13.3 La rutina de Elastic Net

La rutina de `glmnet` importa las funciones base del modelo earth — sus funciones bisagra (hinges), interacciones y términos lineales — como columnas de una matriz de diseño y las reestima bajo regularización lasso/elastic-net (véase “`glmnetUI Import`” más arriba para los detalles del mecanismo y la salvedad sobre la columna de pesos). El resultado conserva los términos interpretables y geométricamente significativos de earth, a la vez que permite que el regularizador contraiga o elimine aquellos que no justifican su permanencia. Véase el artículo sobre `glmnetUI` en este número para el tratamiento completo.

13.4 La rutina GAM

La rutina de `mgcv` va en la dirección opuesta: en lugar de reutilizar la base de earth, ajusta términos suaves penalizados nuevos, usando las ubicaciones de los nodos del modelo earth como puntos de partida (véase “`mgcvUI Auto-Export`” más arriba). Dado que el GAM vuelve a derivar las formas funcionales bajo un paradigma de suavizado distinto, su concordancia con el modelo earth constituye una corroboración genuinamente independiente y no un mero eco. Se admiten modelos earth con $\text{degree} \leq 2$. Véase el artículo sobre `mgcvUI` en este número para el tratamiento completo.

13.5 Un flujo de trabajo práctico

1. Ajuste y refine **primero el modelo earth**, en earthUI, hasta quedar satisfecho con él. Cada ajuste exitoso ($\text{grado} \leq 2$) exporta automáticamente el archivo **.rds** que consumen los módulos.
2. Abra las rutinas de **glmnet** y **mgcv** sobre el mismo proyecto, importe el resultado de earth y ajuste. Trabaje en una rutina a la vez — los ajustes son de larga duración y las rutinas se coordinan a través del proyecto compartido, por lo que el ritmo sensato es secuencial, no paralelo.
3. Itere con libertad. Cada aplicación conserva sus resultados mientras usted alterna entre ellas, y nada obliga a generar informes por el camino.
4. Solo cuando los tres resultados le satisfagan, regrese a cada rutina y genere su informe Cuarto. Cada método registra su ajuste y su conclusión de valor en el registro de trilogía del proyecto, manteniendo los tres informes trazables a una misma corrida comparativa.

***Sugerencia:** Los módulos son opcionales en todos los niveles: el flujo de trabajo propio de earthUI — ajuste, RCA, Sales Grids, informes — está completo sin ellos. Recurra a ellos cuando una conclusión amerite corroboración: un inmueble a valuar difícil, un segmento de mercado con pocos datos o un encargo con probabilidades de ser objeto de revisión.*

14 Conjunto de datos de demostración: Appraisal_1.csv

earthUI incluye un conjunto de datos MLS de demostración para explorar el flujo de trabajo de valuación. Carguelo programáticamente con:

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

O bien, en la aplicación Shiny, cópielo en la carpeta `in/` de un proyecto (o adjúntelo como archivo de datos inicial al crear el proyecto) y selecciónelo en la Sección 2 (Import Data, importación de datos).

14.1 Descripción

El archivo contiene 1,501 ventas residenciales más 1 inmueble a valuar en la fila 1 (1,502 filas en total), provenientes de una exportación MLS simulada. Los datos representan ventas de viviendas unifamiliares en un mercado con múltiples zonas, con una variedad de tamaños, antigüedades y ubicaciones de los inmuebles.

No se trata de datos reales, sino que se basan en un vecindario realista del norte de California. Toda la información de identificación ha sido alterada o eliminada.

14.2 Columnas

Columna	Tipo	Tipo especial	Descripción
weight	numeric	—	Peso de la observación (0 = excluir del ajuste)
id	numeric	display_only	ID interno del registro
property_id	numeric	display_only	Identificador MLS del inmueble
listing_id	character	display_only	Número de listing MLS
parcel_number	character	display_only	Número de parcela (APN) del Assessor del condado (autoridad de valuación fiscal)
street_address	character	display_only	Dirección del inmueble
city_name	character	display_only	Ciudad
postal_code	character	display_only	Código postal (ZIP)
county_name	character	display_only	Condado
contract_date	Date	contract_date	Fecha del contrato de venta (calcula sale_age)
sale_age	numeric	—	Días desde la fecha del contrato hasta la fecha efectiva (precalculado)
coe_date	Date	display_only	Fecha de cierre de escrow
listing_status	character	display_only	Estado del listing (p. ej., "Sold")
sale_price	numeric	(target)	Precio de venta — variable de respuesta
rent	numeric	—	Renta mensual (para modelos con múltiples variables objetivo)
list_price	numeric	display_only	Precio del listing
original_list_price	numeric	display_only	Precio original del listing
living_sqft	numeric	living_area	Superficie habitable bruta en pies cuadrados
beds_total	integer	—	Número de dormitorios
baths_total	numeric	—	Número total de baños (p. ej., 2.5 = 2 completos + 1 medio baño)
lot_size	numeric	lot_size	Tamaño del lote en pies cuadrados
area_id	integer	area	Identificador de zona MLS
area_text	character	display_only	Nombre de la zona
age	numeric	actual_age	Antigüedad del inmueble en años
year_built	integer	display_only	Año de construcción
latitude	numeric	latitude	Latitud (redondeada a 3 decimales para el modelo)
longitude	numeric	longitude	Longitud (redondeada a 3 decimales para el modelo)
latitude6	numeric	display_only	Latitud con precisión completa
longitude6	numeric	display_only	Longitud con precisión completa
garage_spaces	integer	—	Número de plazas de garaje
fp_count	integer	—	Número de chimeneas
no_of_stories	numeric	—	Número de pisos
style	character	—	Estilo arquitectónico
view	character	—	Tipo de vista (p. ej., "Neighborhood", "Hills")
days_on_market	integer	dom	Días en el mercado
listing_date	Date	listing_date	Fecha del listing
sale_concessions	numeric	concessions	Concesiones del vendedor

14.3 Inicio rápido sugerido

1. Inicie earthUI: `earthUI::launch()`
2. Cree un proyecto con el Purpose (propósito) **For Appraisal** (para avalúo) en el asistente New Project (nuevo proyecto), adjuntando **Appraisal_1.csv** como archivo de datos inicial
3. Seleccione **Appraisal_1.csv** en la Sección 2 (Import Data)
4. Seleccione **sale_price** como variable objetivo
5. Asigne los tipos especiales como se muestra en la tabla anterior
6. Incluya los predictores: **sale_age**, **living_sqft**, **baths_total**, **lot_size**, **area_id** (como factor), **age**, **latitude**, **longitude**, **garage_spaces**
7. Establezca el grado en 1 y haga clic en **Fit Earth Model** (ajustar el modelo earth)
8. Descargue la salida intermedia (Paso 6) y revise la clasificación CQA
9. Calcule los ajustes RCA (Paso 7) con una puntuación CQA de ~5.00

10. Genere el Sales Comparison Grid (Paso 8)

Sugerencia: El conjunto de datos de demostración es una buena manera de explorar todas las funciones de earthUI antes de importar sus propios datos MLS. Los predictores sugeridos arriba producen típicamente un modelo con R^2 superior a 0.85.

15 Experimental: extracción de texto con Prolog & reglas de derivación

La versión 0.10.0 introduce una facilidad **opcional y genuinamente experimental** para extraer características listas para el modelo a partir de las columnas de texto descriptivo de una exportación MLS — ante todo la columna “Public Remarks”, pero también los comentarios del agente y las listas de características delimitadas por comas, como las características de las habitaciones o las estructuras existentes. El motor de extracción es Prolog y, en particular, las **gramáticas de cláusulas definidas (Definite Clause Grammars, DCGs)**: las reglas que reconocen frases como “move in ready,” “3 car garage,” u “owned solar” en el texto de los comentarios se escriben como reglas gramaticales de Prolog.

Esta funcionalidad aún se encuentra en desarrollo. Hoy puede utilizarse *parcialmente* para la extracción — la gramática incluida ya reconoce un conjunto útil de características —, pero adaptarla al vocabulario de su propio MLS exige leer y escribir Prolog y, en particular, DCGs. Si ese no es su caso, omita este capítulo por completo: la funcionalidad está desactivada por defecto y el resto de earthUI funciona exactamente igual sin ella.

Advertencia

Se trata de una funcionalidad experimental en desarrollo activo. Su gramática de extracción no está completa ni validada para trabajo de valuación en producción — las columnas extraídas deben cotejarse con el texto de origen antes de utilizarse con confianza en un modelo. Usar la funcionalidad con eficacia requiere conocimientos prácticos de Prolog y de DCGs. Todo lo que hace es aditivo y reversible; si algo funciona mal, desactive el interruptor de Settings (configuración) y continúe con normalidad.

15.1 Qué hace

Los listings de la MLS contienen una gran cantidad de información relevante para el valor en texto libre y en campos de listas delimitadas. Este paso opcional convierte ese texto en columnas ordinarias que su regresión puede utilizar, mediante dos roles especiales de columna (Capítulo 6):

- Las **columnas remarks** son analizadas por la gramática DCG de Prolog para producir una columna de característica por cada atributo extraído. El prefijo de columna se construye con las primeras letras del nombre de la columna de origen: **public_remarks** produce columnas **pr_***, **agent_remark** produce **ar_***, y así sucesivamente.
- Las **columnas list** (p. ej., **existing_structures = "Shed, Workshop"**) se dividen en columnas one-hot **<col>_<value>** de tipo TRUE/FALSE. Esta parte es puro R y no necesita Prolog en absoluto.

La gramática incluida (**inst/prolog/mls_remarks.pl**) extrae actualmente, entre otras cosas: un nivel graduado de estado con una puntuación **condition_score** de 1–5; indicadores de vista y tipos de vista; indicadores de remodelación de cocina y baño; pisos de madera y de otros materiales; una lista y un conteo de mejoras; plazas de garaje y superficie de garaje; piscina, spa y otras características exteriores; número de chimeneas; construcciones auxiliares (cobertizo, granero, taller, casita, casa de huéspedes, perrera, etc.) con tamaños obtenidos de dimensiones como “40x50”; estacionamiento y almacenamiento para vehículos recreativos (RV); propiedad versus arrendamiento de paneles solares, almacenamiento en baterías; tamaño del lote (incluida la conversión de acres a pies cuadrados); montos de precio ligados al contexto de la venta y asociados a un sujeto (reducción de precio, cuotas de HOA, arrendamiento o liquidación del sistema solar, Mello-Roos); indicios de ubicación; y una puntuación de sentimiento de aspectos destacados/preocupaciones en $[-1, 1]$.

15.2 Requisitos: vProlog (Trealla), no SWI-Prolog

El análisis de los comentarios (remarks) se ejecuta sobre el paquete de R **vProlog** — un motor Trealla Prolog autocontenido. No está en CRAN; instálelo desde el repositorio r-universe del autor, como se muestra en “Instalación de earthUI y vProlog” al comienzo de este artículo (binarios precompilados para Windows y macOS; código fuente en github.com/wcraytor/vProlog). No se requiere ni se utiliza ninguna instalación de Prolog en el sistema (SWI-Prolog u otra). Si vProlog no está presente, el análisis de comentarios se omite con una nota, mientras que la división de columnas `list` se ejecuta de todos modos. La función auxiliar `vProlog_available()` informa si tanto el motor como la gramática incluida están presentes.

15.3 Cómo habilitarla y ejecutarla

1. Abra **Settings** (engranaje) y marque **Enable Prolog / list column processing** (habilitar el procesamiento de columnas Prolog/list). El interruptor está desactivado por defecto, lleva la nota “advanced feature — for users with Prolog expertise” (funcionalidad avanzada, para usuarios con experiencia en Prolog), se recuerda por proyecto y se muestra en el diálogo Project Information (información del proyecto).
2. En la tabla de predictores, establezca el rol **Special** de las columnas pertinentes en `remarks` o `list` (varias columnas pueden llevar cada rol).
3. Utilice la sección **Execute Prolog Processing** (ejecutar el procesamiento Prolog), situada entre las Secciones 4 y 5 de la barra lateral. Ofrece dos pasos:

1. Expand lists & remarks (expandir listas y comentarios) — ejecuta la gramática DCG sobre cada columna `remarks` y divide en columnas one-hot cada columna `list`. La expansión es *incremental*: solo se procesan las columnas designadas desde la última ejecución, y el botón se desactiva una vez que todo lo designado ya ha sido expandido. Una barra de progreso da seguimiento al análisis.

2. Execute derivation rules (ejecutar reglas de derivación) — aplica sus propias reglas `derive/2` de Prolog sobre las columnas expandidas (véase más abajo). Las reglas se someten a una verificación de sintaxis y a una validación de ejecución en seco antes de ejecutarse.

15.4 Reglas de derivación

Para cada fila de datos, earthUI aserta las columnas de la fila como hechos `cell(Name, Value)` y recoge todas las soluciones de sus reglas `derive(Col, Val)` — cada `Col` distinto se convierte en una nueva columna, y gana el *primer* valor encontrado por columna, de modo que puede añadirse a continuación una cláusula de respaldo. Los predicados auxiliares (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) dan soporte a la agregación y al manejo de NA, y los hechos `drop/1` descartan columnas después de ejecutarse las reglas. Por ejemplo:

```
% 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).
```

Las reglas residen en un archivo `<project>_rules.pl` por proyecto, generado a partir de una plantilla completamente comentada, y se editan mediante **Edit derivation rules...** (editar reglas de derivación) en un editor de código flotante y arrastrable que también puede acoplarse a la barra lateral. Un botón **Check** (verificar) valida las reglas (exploración de sintaxis, verificación de carga y una ejecución en seco contra la fila 1 de sus datos) sin ejecutarlas.

15.5 Archivo de datos activo versionado

Después de cualquiera de los dos pasos, el data frame aumentado se escribe en la carpeta de entrada del proyecto como `<original_name>_<YYMMDD_HHMMSS>.xlsx` y se convierte en el archivo de datos **activo** del proyecto — al reabrir el proyecto se cargan los datos aumentados, de modo que las expansiones nunca se recalculan, y ese mismo archivo es el que consumen glmnetUI y mgcvUI. Las columnas de origen con los comentarios sin procesar se excluyen automáticamente del modelo; las columnas generadas *no* se incluyen automáticamente como predictores. El análisis de texto puede producir muchas columnas, así que marque como **Latent** (latente) las que (todavía) no desee en la regresión, o desactívelas con $\times$ (Capítulo 6).

15.6 Interfaz programática & archivos incluidos

Toda la cadena de procesamiento está exportada para uso por lotes: `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` orquesta las pasadas, y está construida sobre `remarks_to_columns()`, `split_list_column()`, `apply_user_rules()` y `validate_rules()`. Las fuentes Prolog incluidas bajo `inst/prolog/` son `mls_remarks.pl` (la gramática DCG y la agregación de características), `mls_analyze.pl` (un controlador de entrada/salida de archivos para el análisis por lotes), `mls_rules.pl` (los predicados auxiliares disponibles para las reglas de derivación) y `climate_regions_ca.pl` (más abajo).

15.7 Experimento relacionado: priores por región climática

Una facilidad experimental complementaria aborda las características *de datos escasos* — aquellas que la regresión no puede estimar porque aparecen solo en un puñado de comparables (un enfriador evaporativo, digamos). Tres funciones auxiliares de R — `climate_region_for()`, `market_area_profile()` y `climate_feature_priors()` — clasifican la ubicación de un proyecto (condado + ciudad) en una de siete regiones climáticas de California reconocidas por los compradores (una reducción práctica de las 16 zonas climáticas de edificación de la California Energy Commission) y devuelven priores ordinales de contribución al valor para características de climatización (HVAC) y enfriamiento (calefacción, aire acondicionado, enfriador evaporativo, ventilador de casa completa, filtración) por región. El módulo Prolog `climate_regions_ca.pl` refleja la misma clasificación para su uso dentro de las reglas de derivación; ambos lados se generan a partir de una única fuente canónica en R, de modo que no pueden divergir. Los priores están concebidos para *informar* la valuación de características y servir como verificación de plausibilidad de los coeficientes estimados — las asignaciones de región (en particular para condados que abarcan varias regiones) deben ser revisadas por el valuator.

15.7.1. Más allá de California: paquetes jurisdiccionales y módulos de zona de mercado

California es la primera jurisdicción, no la frontera. earthUI se utiliza internacionalmente, y el conocimiento de los priores climáticos se está reestructurando en forma de **paquetes**

jurisdiccionales (jurisdiction packs): tablas de hechos autocontenidas, con forma de datos (definiciones de zonas, asignaciones de unidades administrativas, priores de características), indexadas por los códigos de país y de nivel administrativo del sistema del proyecto, de modo que un paquete puede describir un estado de los EE. UU., un Land alemán o directamente un país pequeño. Una capa complementaria de **módulos de zona de mercado** captura el vocabulario local y priores refinados para una zona de mercado con nombre propio.

Ambos están diseñados para ser creados *en el terreno* por Valuation Engineers y valuadores, en su propio idioma — y, siendo realistas, con asistencia de LLM: el mundo no cuenta con muchos programadores de Prolog, por lo que los artefactos de campo son deliberadamente tablas y no programas. El formato de paquete es un esquema documentado con un ejemplo completo desarrollado (la propia California), expresamente para que un asistente como Claude pueda redactar un paquete válido a partir de la descripción de un mercado en lenguaje llano; earthUI genera el Prolog subyacente a partir del paquete, y un validador bloqueante (verificaciones de esquema, de cobertura, de integridad referencial y de ejecución en seco) debe aprobarse antes de que un paquete pueda habilitarse. Todo paquete creado en el terreno lleva una procedencia obligatoria — el autor responsable, la fecha y la autoridad citada —, de modo que quien respalda el conocimiento es el profesional responsable, no la herramienta de redacción, y los informes pueden citar el paquete y la versión exactos utilizados.

Sugerencia: Un patrón práctico: expanda los comentarios, consolide los indicadores granulares *pr_** en un puñado de conteos o booleanos derivados con reglas *derive/2*, elimine con *drop/1* los indicadores sin procesar que consolidó y conserve solo las columnas derivadas como predictores candidatos. Esto mantiene corta la lista de predictores y manejable la colinealidad.

Referencias

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

glmnetUI: una interfaz interactiva para el paquete glmnet (Lasso y Elastic Net)

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

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

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

Este artículo de acceso abierto se distribuye bajo los términos de la [licencia Creative Commons Atribución 4.0](#).

Resumen. glmnetUI es una interfaz gráfica de usuario para el paquete de R glmnet, que ajusta modelos lineales generalizados regularizados mediante las penalizaciones Lasso, ridge y elastic-net. Ofrece tres modos de uso — modelado predictivo general, tasación inmobiliaria y análisis de zona de mercado — y guía al usuario a través de la importación de datos, la configuración del modelo, el ajuste con validación cruzada, las trayectorias de coeficientes y los informes descargables. Este artículo documenta los requisitos de formato de datos de glmnetUI, el flujo de trabajo de modelado, las pantallas de salida y la referencia completa de funciones.

Palabras clave: Lasso, regresión ridge, elastic net, regresión regularizada, glmnet, R, interfaz gráfica de usuario, valuación inmobiliaria

Nota del editor (julio de 2026): Poco antes de la publicación, glmnetUI se consolidó dentro de earthUI (versión 0.11.0 y posteriores), que ahora ofrece el flujo de trabajo de elastic net aquí descrito junto con los flujos de trabajo de MARS y GAM. glmnetUI sigue disponible en su forma final publicada (0.5.0) en <https://github.com/wcraytor/glmnetUI>; el desarrollo futuro continúa en earthUI. Este artículo se refiere a glmnetUI tal como fue publicado.

Índice

1. Introducción	3
1.1. Historia y antecedentes	3
1.1.1. El Lasso	3
1.1.2. Regresión ridge y elastic net	4
1.1.3. El paquete glmnet	4
1.1.4. Comparación con MARS (earth) y GAM (mgcv)	4
1.1.5. Flujo de trabajo recomendado	5
1.2. ¿Qué es glmnetUI?	6
1.3. Tres modos de uso	6
1.4. Funciones específicas del sector inmobiliario	7
1.5. Primeros pasos	7
2. Requisitos de los datos de entrada del MLS	9
2.1. Formato y estructura del archivo	9
2.2. Columnas requeridas en el modo de tasación	9
2.3. Convenciones de nomenclatura de columnas especiales	10
2.4. Calidad e integridad de los datos	10
2.5. Colocación de la fila del sujeto	11

3. Modo de uso general	12
3.1. Panorama general	12
3.1.1. Skip First Row (omitir la primera fila)	12
3.1.2. Ajustes por propósito	12
3.2. El flujo de trabajo de la barra lateral	12
3.3. Proyectos	13
3.4. Importación desde earthUI (opcional)	13
3.5. Pestañas del panel principal	14
3.6. Persistencia de los ajustes	14
3.7. Modo oscuro	15
3.8. Configuración regional y de idioma	15
3.8.1. Guardar los valores predeterminados	15
4. Modo de tasación	16
4.1. Tratamiento de la fila del sujeto	16
4.2. Fecha efectiva y edad de la venta	16
4.3. Designaciones especiales de columnas	16
4.4. Panorama de los ajustes RCA	16
5. Modo de análisis de zona de mercado	17
5.1. Diferencias respecto al modo de tasación	17
5.2. Cuándo usar el modo de mercado	17
6. Selección de variables	18
6.1. Variable objetivo	18
6.2. La tabla de predictores	18
6.3. Detección de tipos de datos y anulaciones	18
6.4. Signos esperados y su aplicación	18
6.5. Referencia de los tipos de columna especiales	19
7. Selección de parámetros	20
7.1. Alpha (parámetro de mezcla)	20
7.2. Selección de lambda	20
7.3. Family (familia)	20
7.4. Standardize (estandarizar)	21
7.5. Restricciones de signo	21
7.6. Relaxed lasso	21
7.7. Semilla aleatoria	21
7.8. Matriz de interacciones	22
7.9. Bloquear el efecto principal (solo interacción)	22
7.10. Parámetros avanzados	22
7.11. Valores predeterminados de los ajustes	23
8. Ajuste del modelo	24
8.1. El botón de ajuste	24
8.2. Qué ocurre durante el ajuste	24
9. Pestañas de resultados	25
9.1. Data Preview	25
9.2. Equation	25
9.3. Correlation	25
9.4. Summary	25

9.5. Coefficients	26
9.6. Variable Importance	26
9.7. Contributions	26
9.8. ANOVA	26
9.9. Diagnostics	26
9.10. Glmnet Output	27
9.11. Report	27
10.Descarga de datos	28
10.1. Columnas de salida	28
10.2. Puntuaciones CQA	28
11.Cálculos RCA y descarga	29
11.1. Apertura del diálogo RCA	29
11.2. Interpolación de la puntuación CQA	29
11.3. Columnas de salida	29
12.Sales Comparison Grid	30
12.1. Diálogo de selección de comparables	30
12.2. Estructura del libro	30
12.3. Protección de la hoja y celdas residuales	30
13.Descarga de informes	31
13.1. El flujo de trabajo Quarto de dos pasos	31
13.2. Formatos de informe	31
13.3. Contenido del informe	31
14.Comparación con earthUI	33
14.1. Diferencias principales	33
14.2. Funciones compartidas	33
14.3. Cuál usar en cada caso	33
15.Conjunto de datos de demostración: Appraisal_1.csv	35
15.1. Descripción	35
15.2. Columnas	35
15.3. Inicio rápido sugerido	36
16.Requisitos del sistema y resolución de problemas	37
16.1. Plataformas admitidas	37
16.2. Dependencias opcionales	37
16.3. Notas por plataforma	37
16.4. Degradación elegante	37
16.5. Resolución de problemas	37

1 Introducción

1.1 Historia y antecedentes

1.1.1. El Lasso

El **Lasso** (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator) fue introducido por Robert Tibshirani en 1996. El nombre es un juego de palabras deliberado con el lazo del vaquero — una

cuerda que restringe y selecciona. Tibshirani lo eligió porque el método literalmente “enlaza” los coeficientes, restringiendo sus valores absolutos y llevando algunos exactamente a cero, con lo cual selecciona qué variables permanecen en el modelo.

La intuición de Tibshirani consistió en añadir una penalización L_1 (suma de los valores absolutos de los coeficientes) al objetivo de mínimos cuadrados. A diferencia de la penalización L_2 de la regresión ridge (Hoerl y Kennard, 1970), que contrae los coeficientes hacia cero pero nunca lo alcanza, la penalización L_1 produce soluciones *dispersas* en las que muchos coeficientes son exactamente cero. Esto convierte al Lasso a la vez en un método de regularización y en un método de selección de variables.

1.1.2. Regresión ridge y elastic net

La **regresión ridge** fue propuesta por Arthur Hoerl y Robert Kennard en 1970. Añade una penalización L_2 (suma de los cuadrados de los coeficientes) que contrae todos los coeficientes proporcionalmente pero mantiene todas las variables en el modelo. Ridge es eficaz cuando los predictores están correlacionados (multicolinealidad), pero no simplifica el modelo.

El **elastic net**, introducido por Hui Zou y Trevor Hastie en 2005, combina ambas penalizaciones. El parámetro de mezcla α controla la combinación: $\alpha = 1$ es Lasso puro, $\alpha = 0$ es ridge puro, y los valores intermedios dan un compromiso. El elastic net supera una limitación del Lasso: cuando los predictores están muy correlacionados, el Lasso tiende a seleccionar uno e ignorar el resto, mientras que el elastic net agrupa los predictores correlacionados.

1.1.3. El paquete glmnet

El paquete de R **glmnet** fue desarrollado por Jerome Friedman, Trevor Hastie y Robert Tibshirani en la Universidad de Stanford. Implementa la regularización elastic net para modelos lineales generalizados mediante un algoritmo de descenso por coordenadas extremadamente eficiente (Friedman, Hastie y Tibshirani 2010), extendido después a los modelos de riesgos proporcionales de Cox (Simon et al. 2011) y a todas las familias de modelos lineales generalizados (Tay, Narasimhan y Hastie 2023). El paquete admite las familias gaussian (lineal), binomial (logística), poisson y otras familias de GLM.

Entre las características principales de **glmnet** se incluyen:

- **Trayectoria de regularización:** ajusta el modelo sobre una malla de valores de λ en una sola pasada, mucho más rápido que ajustarlo por separado en cada λ .
- **Validación cruzada:** `cv.glmnet()` selecciona automáticamente la intensidad óptima de regularización mediante validación cruzada de k pliegues.
- **Cotas de los coeficientes:** admite límites superiores e inferiores para coeficientes individuales (restricciones de signo).
- **Relaxed lasso:** reajusta el modelo sobre las variables seleccionadas con penalización reducida o nula, lo que disminuye el sesgo.

1.1.4. Comparación con MARS (earth) y GAM (mgcv)

La aplicación **earth** y sus dos Post Processing Modules (módulos de posprocesamiento; **glmnet** y **mgcv**) utilizan motores de modelado distintos. La tabla siguiente resume las diferencias principales:

Característica	glmnet (Elastic Net)	earth (MARS)	mgcv (GAM)
Tipo de modelo	Lineal (con regularización)	Lineal por tramos (splines adaptativos)	No lineal suave (splines penalizados)
No linealidad	Solo mediante interacciones o expansión de base	Automática mediante funciones bisagra con nodos guiados por los datos	Automática mediante funciones suaves (thin plate, cúbicas, etc.)
Selección de variables	Incorporada mediante la penalización L_1 (Lasso)	Incorporada mediante pasos hacia adelante/atrás con poda por GCV	Opcional mediante contracción de doble penalización (<code>select=TRUE</code>)
Interacciones	Productos cruzados por pares (difíciles de interpretar)	Productos basados en bisagras (interpretables, hasta grado 3)	Suavizados de producto tensorial (<code>te()</code> , <code>ti()</code>)
Interpretabilidad	Los coeficientes son pesos lineales; fáciles de explicar individualmente	Las funciones g muestran efectos parciales lineales por tramos para cada variable	Curvas suaves de efecto parcial; muy intuitivas
Más indicado para	Datos de alta dimensión, selección de variables, muestras pequeñas	Detección automática de no linealidad e interacciones	Relaciones no lineales suaves, modelado flexible
Ajustes RCA	Ajustes lineales por variable	Ajustes lineales por tramos mediante funciones g	Ajustes suaves mediante efectos parciales
Control del sobreajuste	Lambda validado por validación cruzada	Poda basada en GCV	Estimación del parámetro de suavizado por REML/GCV

MARS (Multivariate Adaptive Regression Splines) fue introducido por Jerome Friedman en 1991. Construye modelos lineales por tramos seleccionando de manera adaptativa funciones bisagra $\max(0, x-k)$ y las posiciones de sus nodos a partir de los datos. La implementación en R es el paquete **earth** de Stephen Milborrow. MARS destaca por descubrir automáticamente relaciones no lineales e interacciones sin que el usuario tenga que especificarlas. Sus funciones g (términos de base agrupados por variable) proporcionan curvas de efecto parcial muy interpretables.

Los **GAM** (Generalized Additive Models, modelos aditivos generalizados) fueron formalizados por Hastie y Tibshirani (1990). La implementación en R **mgcv**, de Simon Wood, ofrece splines de regresión penalizados con selección automática del suavizado mediante REML o GCV. Los GAM producen curvas suaves de efecto parcial que resultan intuitivas de visualizar y de explicar. Cuando **earthUI** exporta las posiciones de los nodos a **mgcvUI**, los nodos derivados de **earth** sirven de punto de partida para los términos suaves del GAM, combinando la colocación adaptativa de nodos de MARS con la estimación suave del GAM.

1.1.5. Flujo de trabajo recomendado

Para la tasación inmobiliaria y aplicaciones similares que requieren modelos interpretables y defendibles:

1. **Comience con earthUI** para descubrir las variables importantes, las relaciones no lineales y las interacciones presentes en sus datos. La selección automática de base y las funciones `g` de `earth` proporcionan excelentes intuiciones iniciales.
2. **Refine con mgcvUI** si desea efectos parciales más suaves. Importe los nodos de `earthUI` en `mgcvUI` para una transición fluida de modelos lineales por tramos a modelos suaves.
3. **Utilice glmnetUI** cuando necesite una selección de variables agresiva (Lasso), cuando el número de predictores se acerque o supere el número de observaciones, o cuando desee un modelo puramente lineal con regularización por motivos de defendibilidad.

1.2 ¿Qué es glmnetUI?

`glmnetUI` es una interfaz gráfica de usuario para el paquete `glmnet` de R. Se ejecuta como una aplicación Shiny local — no hay inicio de sesión, ni servidor, ni cuentas. Se inicia desde R, se importa un conjunto de datos (CSV o Excel), se configura el modelo y se ajusta de manera interactiva.

A partir de la versión actual, `glmnetUI` ya no es un paquete que se instale por separado: se distribuye *dentro* del paquete `earthUI` como su Post Processing Module de elastic net, y se inicia con `earthUI::launch_glmnet()`. La aplicación en sí — y todo lo descrito en este artículo — no ha cambiado; solo difieren la instalación y el comando de inicio. Las tres rutinas (`earth`, `glmnet`, `mgcv`) comparten un mismo árbol de proyectos, una misma base de datos de configuración y una misma copia de la infraestructura de soporte.

La aplicación proporciona un flujo de trabajo completo: importación de datos, configuración de variables, ajuste del modelo, gráficos de diagnóstico, importancia de las variables, ecuaciones del modelo e informes descargables en formato Word, PDF o HTML.

La **regresión elastic net** combina dos técnicas de regularización:

- **Regresión ridge** ($\alpha = 0$): contrae proporcionalmente todos los coeficientes hacia cero. Adecuada cuando se considera que todas las variables son relevantes.
- **Regresión Lasso** ($\alpha = 1$): puede contraer coeficientes exactamente a cero, realizando una selección automática de variables. Produce modelos más sencillos e interpretables.
- **Elastic net** ($0 < \alpha < 1$): una combinación de ambos enfoques. Resulta útil cuando los predictores están correlacionados.

1.3 Tres modos de uso

Todo proyecto de `glmnetUI` tiene un **purpose** (propósito o modo de uso), que se elige al crearlo (Sección 1, *Project*). El propósito del proyecto activo determina qué herramientas y elementos de la interfaz aparecen; para trabajar en otro modo, cree o abra un proyecto con ese propósito. Los tres modos son:

- **General** — regresión elastic net para cualquier tipo de población o conjunto de datos. Este es el modo predeterminado. Ofrece el flujo de trabajo completo de regresión regularizada sin añadidos específicos de ningún dominio.
- **For Appraisal** (para tasación) — regresión elastic net adaptada a la tasación inmobiliaria. Añade funciones específicas para la valuación de una sola propiedad, incluidos el tratamiento de la propiedad sujeto, las designaciones especiales de columnas y el Residual Constraint Approach (RCA).
- **Market Area Analysis** (análisis de zona de mercado) — regresión elastic net adaptada a los estudios de zona de mercado. Añade funciones para analizar grupos de propiedades dentro de un mercado definido, con exclusión opcional de la fila del sujeto.

En los tres modos el motor de modelado subyacente es idéntico — siempre se ajusta un modelo elastic net (`glmnet`). El ajuste del propósito controla qué herramientas y elementos de interfaz adicionales están disponibles.

1.4 Funciones específicas del sector inmobiliario

Cuando se selecciona **For Appraisal** o **Market Area Analysis**, glmnetUI activa varias funciones diseñadas para el análisis inmobiliario:

- **Designaciones especiales de columnas** — cada predictor puede etiquetarse con un papel especial como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` o `display_only`. Estas designaciones controlan cómo se trata la columna durante el ajuste y en las salidas.
- **Redondeo de latitud y longitud** — las columnas designadas como latitud o longitud se redondean automáticamente a 3 decimales para evitar el sobreajuste.
- **Columna Sale Age** — cuando una columna se designa como `contract_date` y se proporciona una fecha efectiva, glmnetUI calcula una columna `sale_age` (días entre la fecha de venta y la fecha efectiva) y la sustituye como predictor.
- **Cálculos RCA (solo en tasación)** — en el modo de tasación, después de ajustar el modelo glmnetUI puede calcular la salida del Residual Constraint Approach (RCA), que produce ajustes por comparable, resúmenes de ajuste neto/bruto y un precio de venta ajustado para la propiedad sujeto.

***Consejo:** El modo General funciona con cualquier conjunto de datos — financiero, científico, de ingeniería o inmobiliario. Los modos de tasación y de mercado simplemente añaden funciones de conveniencia para los profesionales del sector inmobiliario.*

1.5 Primeros pasos

Para usar glmnetUI:

1. **Instale earthUI** en R: `install.packages("earthUI")` — la rutina de glmnet está incluida. (Véase “Instalación de earthUI y vProlog” en el artículo sobre earthUI de este número para la versión de desarrollo y los componentes opcionales).
2. **Inicie la aplicación:** ejecute `earthUI::launch_glmnet()` en R. La aplicación se abre en su navegador web en el puerto 7879. También puede acceder directamente a la aplicación navegando a `http://localhost:7879` en su navegador. La aplicación recuerda el último modo de uso empleado y lo restaura automáticamente.
3. **Cree o abra un proyecto** en la Sección 1 de la barra lateral. Un proyecto fija el propósito y la ubicación de trabajo (país, estado, condado, ciudad) y sus carpetas de entrada y salida. Los proyectos se almacenan bajo una raíz `regProj` compartida y se comparten con las aplicaciones hermanas earthUI y mgcvUI. Abrir un proyecto con un propósito distinto es la forma de cambiar de modo.
4. **Importe sus datos** en la Sección 2 seleccionando un archivo CSV o Excel de la carpeta de entrada del proyecto activo. Añada archivos a esa carpeta y haga clic en **Refresh** (actualizar) para listarlos.
5. **Importe desde earthUI (opcional)** — la Sección 3 permite importar un archivo `.rds` de resultados de earthUI para utilizar las funciones de base bisagra de earth con la regularización de glmnet. Omita este paso si no dispone de un resultado de earthUI.
6. **Configure las variables** — elija su variable objetivo y sus predictores, fije los tipos de datos, los signos esperados y asigne los papeles especiales de columna.
7. **Fije los parámetros de glmnet** — `alpha`, `lambda`, `family`, restricciones de signo, interacciones y otras opciones.
8. **Ajuste el modelo** — haga clic en “Fit Glmnet Model” (ajustar el modelo glmnet) y revise los resultados en el panel principal.
9. **Exporte** — descargue las predicciones en Excel, genere y convierta un informe Quarto o (en modo de tasación) calcule los ajustes RCA y una Sales Comparison Grid (cuadrícula

de comparación de ventas).

Los ajustes se guardan automáticamente en el almacenamiento local de su navegador y se restauran cuando vuelve a cargar el mismo archivo de entrada.

2 Requisitos de los datos de entrada del MLS

Para los flujos de trabajo de tasación inmobiliaria y análisis de mercado, sus datos de entrada provienen normalmente de una exportación de un Multiple Listing Service (MLS, servicio de listado múltiple). Este capítulo describe la estructura de archivo esperada y las columnas que glmnetUI puede utilizar.

2.1 Formato y estructura del archivo

glmnetUI acepta archivos **CSV** y **Excel** (.xlsx, .xls). Al importar, los nombres de las columnas se convierten automáticamente a **snake_case** — por ejemplo, “Living SqFt” pasa a ser **living_sqft**, “Contract Date” pasa a ser **contract_date** y “Sale Price” pasa a ser **sale_price**. Esta normalización garantiza referencias de columna coherentes en todo el flujo de trabajo. El separador de CSV y el signo decimal utilizados durante la importación vienen determinados por la configuración regional (véase el capítulo 3, “Locale & Regional Settings”).

Su archivo de datos debe ser una tabla plana con una fila por propiedad y una columna por atributo. La primera fila del archivo debe contener los encabezados de columna.

2.2 Columnas requeridas en el modo de tasación

Aunque glmnetUI funciona con cualquier conjunto de columnas, el flujo de trabajo completo de tasación se beneficia de disponer de las columnas siguientes:

Columna	Tipo especial	Finalidad
Sale Price	(objetivo)	Variable respuesta del modelo
Contract Date	contract_date	Se usa para calcular sale_age (días desde la fecha efectiva)
Listing Date	listing_date	Se usa con la fecha de contrato para calcular el DOM si no existe columna DOM
Days on Market	dom	Días en el mercado; se muestra en las exportaciones
Concessions	concessions	Concesiones de venta; Net SP = Sale Price – Concessions
Living Area (SF)	living_area	Permite residuos por pie cuadrado (residual_sf , cqa_sf)
Lot Size	lot_size	Columna de tamaño del terreno
Site Dimensions	site_dimensions	Agrupada con el tamaño del terreno
Latitude	latitude	Redondeada a 3 decimales
Longitude	longitude	Redondeada a 3 decimales
Area ID	area	Identificador de zona de mercado / vecindario
Actual Age	actual_age	Edad de la propiedad
Effective Age	effective_age	Edad efectiva de la propiedad
Address	display_only	Se muestra en las exportaciones; se excluye del modelo

Los nombres de columna de la hoja de cálculo pueden estar en otro idioma — los nombres “especiales” están en inglés para que el programa de R pueda darles un tratamiento especial. Por lo demás, los nombres de columna suministrados aparecen en los modelos de regresión, en los gráficos y, si se hacen tasaciones, en los informes de salida.

No todas las columnas son obligatorias. glmnetUI se adapta — si falta una columna, la función correspondiente simplemente se omite. No obstante, para los modelos de precios inmobiliarios se recomiendan encarecidamente ciertas columnas para lograr un ajuste aceptable:

1. **Sale Age** — el número de días entre la fecha de venta según contrato y la fecha efectiva de la tasación o del análisis. Si se emplea un historial de ventas de varios años, especialmente para períodos superiores a 5 años, **sale_age** suele desempeñar un papel central en la estimación del precio de venta.
2. **Living Area** — también aparece con nombres como “Living Sqft”, “GLA” (gross living area, superficie habitable bruta), etc. Es otro determinante principal del precio de venta.
3. **Total Bath Count** — el número total de baños completos, de un cuarto, de medio baño y de tres cuartos. Por ejemplo, dos baños completos y un medio baño darían un valor de 2,5.
4. **Garage Bays** o **Garage Area** — el número de plazas de garaje o la superficie del garaje.
5. **Lot Size** — la superficie de terreno de la propiedad, normalmente en pies cuadrados o acres.
6. **Longitude**, **Latitude** y, si está disponible, **Area ID**. Estas variables de localización ayudan al modelo a tener en cuenta la variación geográfica de los precios.

2.3 Convenciones de nomenclatura de columnas especiales

glmnetUI identifica las columnas por su **designación de tipo especial**, no por su nombre de columna. Puede dar a sus columnas el nombre que desee en la exportación del MLS — lo que importa es que asigne el tipo especial correcto en la tabla de configuración de variables (capítulo 6).

Por ejemplo, su MLS podría exportar la superficie habitable como “GLA”, “Living SqFt”, “liv_area” o “gross_living_area”. Tras la importación (donde pasa a snake_case), basta con designarla como **living_area** en el desplegable Special. glmnetUI la utilizará entonces para los cálculos de residuos por pie cuadrado con independencia de su nombre original.

2.4 Calidad e integridad de los datos

- **Valores faltantes (NA)**: las filas con valores NA en cualquier columna predictora u objetivo se eliminan automáticamente antes del ajuste.
- **Columnas de fecha**: las columnas de texto que coinciden con formatos de fecha habituales se analizan automáticamente al importar. Los formatos con año de 2 dígitos tienen prioridad cuando no se detecta un año de 4 dígitos.
- **Columnas numéricas**: el precio de venta, la superficie habitable, el tamaño del terreno, las concesiones y campos similares deben ser numéricos. Si su MLS exporta los precios con símbolos de moneda o separadores de miles (por ejemplo, “\$350,000” o “350.000,00”), es posible que deba limpiarlos antes de la importación.
- **Columnas de factor**: las columnas de texto, TRUE/FALSE y de tipo factor de R se tratan automáticamente como categóricas — una regresión solo puede manejarlas como factores. Una columna **numérica** (por ejemplo, un código numérico de zona o de estilo) se convierte en factor solo cuando se marca su casilla **Factor**; glmnetUI nunca infiere el carácter categórico a partir del rango de valores de una columna numérica, de modo que predictores numéricos discretos como **bath_count** permanecen continuos. El papel Special es independiente de la decisión sobre el factor.

***Consejo:** Revise la columna NA en la tabla de predictores después de la importación. Las columnas con muchos valores faltantes pueden provocar la eliminación de filas. Considere excluir las columnas con muchos NA o limpiar los datos antes de importarlos.*

2.5 Colocación de la fila del sujeto

En el modo **Appraisal**, la fila 1 debe ser la propiedad sujeto. Todas las filas restantes son ventas comparables. La fila del sujeto se excluye del ajuste del modelo. Tras el ajuste, el modelo genera igualmente predicciones para la fila del sujeto.

En el modo **Market Area Analysis**, colocar el sujeto en la fila 1 es opcional.

En el modo **General** no hay tratamiento especial de filas — todas las filas se tratan por igual.

3 Modo de uso general

3.1 Panorama general

El modo General Purpose (uso general) es el predeterminado al iniciar glmnetUI. Ofrece el flujo de trabajo completo de regresión elastic net para cualquier conjunto de datos — no solo inmobiliarios. Puede usar glmnetUI para datos científicos, análisis financiero, estudios de ingeniería o cualquier problema de regresión.

En el modo General, la interfaz omite las funciones específicas del sector inmobiliario (columnas especiales, edad de la venta, redondeo de coordenadas, RCA). La barra lateral se simplifica para centrarse en la selección de variables, la configuración de parámetros, el ajuste del modelo y la exportación.

3.1.1. Skip First Row (omitir la primera fila)

Una casilla **Skip first row** (omitir la primera fila) aparece en la parte superior de la barra lateral una vez abierto un proyecto General o Market. Cuando está marcada, la fila 1 se excluye del ajuste del modelo. Esto resulta útil cuando la fila 1 contiene una observación objetivo o de referencia que no debe influir en el modelo. En el modo Appraisal, la fila 1 (la propiedad sujeto) siempre se excluye automáticamente.

3.1.2. Ajustes por propósito

Los ajustes (selección de predictores, parámetros del modelo, interacciones) se guardan por separado para cada combinación de archivo de entrada y modo de uso. Cambiar entre los modos General, Appraisal y Market conserva los ajustes de cada modo de forma independiente.

3.2 El flujo de trabajo de la barra lateral

La barra lateral se organiza en secciones numeradas y plegables que le guían desde la selección del proyecto hasta la exportación del informe. La mayoría de las secciones aparecen solo una vez abierto un proyecto activo, y varias son específicas del propósito del proyecto. La carpeta de salida se deriva del proyecto activo — no hay un campo aparte para la carpeta de salida.

1. Project — crear o abrir un proyecto. Un proyecto fija el propósito y la ubicación de trabajo (país, estado, condado, ciudad) y sus carpetas de entrada (`<os>_in/`) y de salida. Los proyectos se almacenan bajo una raíz `regProj` compartida (por omisión `~/regProj`, modificable con la variable de entorno `REGPROJ_ROOT`) y se comparten con las aplicaciones hermanas `earthUI` y `mgcvUI`, de modo que los modelos de las tres conviven lado a lado. Haga clic en **Close Project** (cerrar proyecto) para cambiar.

2. Import Data — seleccionar un archivo CSV o Excel de la carpeta de entrada del proyecto activo. Añada archivos a la carpeta mediante el Finder o el Explorador, y haga clic en **Refresh** para listarlos. Para los archivos de Excel con varias hojas aparece un selector de hoja. Los nombres de columna se convierten automáticamente a `snake_case`.

3. Import from earthUI (opcional) — importar un archivo `.rds` de resultados de `earthUI` para utilizar las funciones de base bisagra de `earth` con la regularización elastic net de `glmnet`. El botón de exploración tiene el mismo estilo que el selector de archivos de la Sección 2. Este paso es opcional — omítalo si no dispone de un resultado de `earthUI` que importar.

4. Variable Configuration — selector de la variable objetivo, tabla de predictores con casillas Include, Factor y Force, además de los desplegables Sign. La columna Special (para designar fechas de contrato, coordenadas, etc.) aparece solo en los modos Appraisal y Market. Véase el capítulo 6 para todos los detalles.

5. glmnet Call Parameters — toda la configuración del modelo: alpha, lambda, family, standardize, restricciones de signo, relaxed lasso, semilla aleatoria, matriz de interacciones y parámetros avanzados (lambda.min.ratio, nlambda, métrica de pérdida de CV, umbral de convergencia, iteraciones máximas, intercepto). Véase el capítulo 7 para la referencia completa de parámetros.

6. Fit Glmnet Model — el botón que ejecuta el modelo. Los resultados aparecen en las pestañas de la derecha.

7. Download Output — exporta las predicciones, los residuos, las puntuaciones CQA y las contribuciones por variable como archivo de Excel a la carpeta de salida del proyecto. Disponible en todos los modos de uso una vez ajustado el modelo.

8. Calculate RCA Adjustments & Download — calcula los ajustes RCA por variable para cada comparable e interpola el residuo del sujeto mediante CQA. Aparece solo en los modos Appraisal y Market, después del ajuste. Véase el capítulo 11.

9. Generate Sales Grid & Download — genera la Sales Comparison Grid intermedia, ordenando los comparables por porcentaje de ajuste bruto. Aparece solo en el modo Appraisal, después de calcular los ajustes RCA. Véase el capítulo 12.

10. Generate Quarto Report — escribe un paquete de informe Quarto (.qmd) autocontenido — con todos los gráficos y tablas — en la carpeta de salida del proyecto. Disponible una vez ajustado el modelo.

11. Convert Quarto Report — renderiza el paquete .qmd generado a HTML, Word o PDF. El PDF requiere una instalación de LaTeX; si no se detecta ninguna, la opción de PDF queda oculta. Véase el capítulo 13.

En los modos Market y General se omiten los pasos exclusivos de tasación y las dos secciones de informe se renumeran en consecuencia.

3.3 Proyectos

glmnetUI organiza el trabajo en **proyectos** bajo una carpeta raíz **regProj** compartida (el mismo árbol que usan earthUI y mgcvUI). Cada proyecto posee una carpeta de entrada (<os>_in) y carpetas de salida por motor (por ejemplo, <os>_out_glmnet).

Para crear un proyecto, haga clic en **+ New...** en la Sección 1 y proporcione:

- **Purpose** — General, Appraisal o Market Area.
- **Location** — la jerarquía política/administrativa del proyecto. El conjunto de niveles que se rellenan depende del país que se elija, de modo que las convenciones de denominación extranjeras se admiten automáticamente — un proyecto estadounidense usa estado/condado/ciudad, uno francés usa región/département/commune, uno japonés usa prefectura/ciudad-o-barrio/distrito, y así sucesivamente.
- **Project Name** — una etiqueta corta, limitada a 8 caracteres (letras, dígitos, _ y -).

La aplicación no usa literalmente el nombre que usted teclea: construye un nombre de carpeta único con marca de tiempo anteponiendo los códigos de ubicación y la fecha y hora de creación, con la forma <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name> — por ejemplo, un proyecto estadounidense llamado **bur1** pasa a ser **us_ca_081_bur1in_20260604-131500_bur1**. El límite de 8 caracteres se aplica solo a la parte que usted teclea. Solo los proyectos recién creados reciben el prefijo de marca de tiempo; los proyectos ya existentes en disco conservan sus nombres originales y siguen siendo plenamente accesibles.

3.4 Importación desde earthUI (opcional)

La Sección 3 de la barra lateral — **Import from earthUI** — permite importar un archivo .rds de resultados de earthUI. Esto permite a glmnetUI usar las funciones de base bisagra de earth (términos lineales por tramos) con la regularización elastic net de glmnet, combinando la

detección automática de no linealidad de earth con la selección de variables y las restricciones de coeficientes de glmnet.

Cuando se importa un resultado de earthUI, son los predictores del modelo earth — y no las casillas Include — los que determinan el ajuste de glmnet, porque glmnet se ajusta sobre la expansión de base de earth. La tabla de configuración de variables lo deja explícito: **Include** y **Type** quedan bloqueados, las filas de los predictores que el modelo earth no utilizó se deshabilitan y una nota explica que el conjunto de predictores procede del modelo earth. **Force** y **Sign** siguen siendo editables para los predictores de earth, ya que estos aún se aplican a la base de earth en el momento del ajuste.

El botón de exploración tiene el mismo estilo que el selector de archivos de la Sección 2. Este paso es totalmente opcional — si no dispone de un archivo de resultados de earthUI, simplemente omita la Sección 3 y pase a la Sección 4.

3.5 Pestañas del panel principal

Tras el ajuste, el panel principal ofrece las pestañas siguientes:

Pestaña	Contenido
Data Preview	Vista previa de los datos importados. En modo de tasación, se divide en las tablas Subject Property y Comparable Sales.
Equation	La ecuación del modelo ajustado, mostrada en LaTeX/MathJax, con cada coeficiente no nulo y los términos de interacción.
Correlation	Mapa de calor de las correlaciones entre predictores numéricos (disponible antes del ajuste).
Summary	Métricas clave (R^2 , R^2 ajustado, GR^2 , R^2 de CV, RMSE, MAE) y métricas de tasación (COD, PRD, razón mediana).
Coefficients	Tabla de coeficientes no nulos con avisos de violación de signo.
Variable Importance	Diagrama de barras y tabla ordenada de la importancia de los predictores ($ \beta \times sd(x)$). Interacción con plotly al pasar el cursor, cuando está disponible.
Contributions	Gráficos de efecto parcial por variable con etiquetas de pendiente. Dispersión + línea de ajuste para predictores numéricos, diagramas de caja para factores, superficie/dispersión/mapa de calor 3D para interacciones.
ANOVA	Tabla de descomposición de la varianza: SS por variable, % de la SS del modelo, coeficiente.
Diagnostics	Cuatro gráficos: trayectoria de coeficientes, error de CV, valores reales frente a predichos y residuos frente a valores ajustados.
Glmnet Output	Salida bruta del modelo: semilla aleatoria empleada, impresión del modelo, lambda seleccionado, lambda.min/1se, gamma (si es relaxed) y vector completo de coeficientes.
Report	Campos de configuración y exportación del informe.
RCA Adjustments	Resultados del análisis RCA (modo de tasación, tras el cálculo).

3.6 Persistencia de los ajustes

glmnetUI guarda automáticamente su configuración en el almacenamiento local del navegador, indexada por el nombre del archivo de entrada. Cuando vuelve a cargar el mismo archivo, se restauran todos los ajustes: selección del objetivo, casillas de los predictores, tipos de datos, signos esperados, parámetros de glmnet y matriz de interacciones. El último modo de uso empleado

también se guarda globalmente y se restaura al volver a iniciar la aplicación. Hay tres opciones disponibles:

- **Use last settings for input file** — restaurar los ajustes guardados para este archivo concreto
- **Use default settings** — aplicar los valores predeterminados globales guardados
- **glmnet defaults** — restablecer los valores de fábrica ($\alpha = 1$, CV lambda.1se, 10 pliegues, gaussian, standardize activado)

Un botón **Save current as default** guarda todos los ajustes actuales como valor predeterminado global.

3.7 Modo oscuro

Haga clic en el icono de luna/sol de la esquina superior derecha para alternar entre los temas claro y oscuro. La preferencia de tema se guarda en el almacenamiento local y persiste entre sesiones. Todos los elementos de la interfaz (tablas, gráficos, tarjetas, botones) se adaptan al tema seleccionado mediante variables CSS.

3.8 Configuración regional y de idioma

glmnetUI admite convenciones internacionales de formato numérico y de CSV mediante un sistema de configuración regional basado en el país. El desplegable **Settings** de la barra de título ofrece selectores **Country** y **Paper** para 31 países admitidos. Cada preajuste configura:

- **Separador de CSV** — coma (,) para EE. UU., Reino Unido y Japón, o punto y coma (;) para gran parte de Europa, donde la coma se usa como signo decimal.
- **Signo decimal** — punto (.) o coma (,).
- **Separador de millares** — coma (EE. UU./Reino Unido/Japón), punto (Alemania/Italia/España), espacio (Finlandia/Francia/Polonia/Bálticos/Ucrania/Rusia) o apóstrofo (Suiza).
- **Tamaño de papel** — Letter (EE. UU./Canadá/México) o A4 (en el resto de los casos).

3.8.1. Guardar los valores predeterminados

Haga clic en **Save as my default** para almacenar sus preferencias regionales globalmente. Estos valores predeterminados se aplican a todas las sesiones futuras con independencia del archivo de datos que cargue.

***Consejo:** El formato numérico de los ejes de los gráficos y de las etiquetas de pendiente se adapta automáticamente a su configuración regional. La configuración alemana usa puntos para los millares (200.000), la finlandesa usa espacios (200 000) y la suiza usa apóstrofes (200'000). No se muestran símbolos de moneda — glmnetUI es agnóstico respecto de la moneda.*

4 Modo de tasación

Cuando selecciona **For Appraisal** como propósito, glmnetUI se configura para la valuación de una sola propiedad. Todas las funciones descritas en el capítulo 3 siguen disponibles; este capítulo cubre únicamente las adiciones específicas de la tasación.

4.1 Tratamiento de la fila del sujeto

En el modo de tasación, **la fila 1 de su conjunto de datos es la propiedad sujeto** y todas las filas restantes son ventas comparables. Su archivo de entrada debe organizarse en consecuencia (véase el capítulo 2). El precio de venta del sujeto puede dejarse en blanco o fijarse en cualquier valor — glmnetUI lo trata automáticamente como NA durante el ajuste.

Tras la importación, la pestaña Data Preview se divide en dos secciones: **Subject Property** (fila 1) y **Comparable Sales** (filas 2 y siguientes). La fila 1 siempre se excluye del ajuste del modelo. Tras el ajuste, el modelo genera igualmente predicciones para la fila del sujeto.

4.2 Fecha efectiva y edad de la venta

En los modos de tasación y de mercado aparece un campo **Effective Date** (fecha efectiva) en la sección de configuración de variables (con la fecha de hoy por omisión). Si designa una columna como **contract_date** en el desplegable de la columna Special, glmnetUI calcula una columna **sale_age** — el número entero de días entre la fecha de contrato de cada venta y la fecha efectiva. Esta columna se añade como predictor.

Cuando cambia la fecha efectiva, **sale_age** se recalcula automáticamente.

4.3 Designaciones especiales de columnas

En los modos de tasación y de mercado aparece un desplegable **Special** para cada predictor en la tabla de configuración de variables. Véase el capítulo 6 para la referencia completa de los tipos especiales y sus efectos.

4.4 Panorama de los ajustes RCA

El botón **Calculate RCA Adjustments & Download** (sección 8 de la barra lateral, visible en los modos Appraisal y Market tras el ajuste) calcula ajustes derivados del mercado para cada comparable en relación con el sujeto. El flujo de trabajo RCA completo se describe en el capítulo 11.

***Consejo:** La puntuación CQA que asigne al sujeto controla qué parte de la distribución de residuos se atribuye al sujeto. Una puntuación de 5,00 sitúa al sujeto en la mediana de los comparables.*

5 Modo de análisis de zona de mercado

Cuando selecciona **Market Area Analysis** como propósito, glmnetUI ofrece las mismas funciones específicas del sector inmobiliario que el modo de tasación (columnas especiales, edad de la venta, redondeo de coordenadas), pero está orientado a analizar un grupo de propiedades en lugar de valorar un único sujeto.

5.1 Diferencias respecto al modo de tasación

- **El tratamiento de la fila del sujeto es flexible** — en el modo de mercado, la fila 1 no se trata automáticamente como propiedad sujeto.
- **No hay Sales Comparison Grid** — el paso “Calculate RCA Adjustments & Download” (Sección 8) está disponible en el modo de mercado, pero el paso de Sales Grid (Sección 9) es exclusivo de la tasación.

5.2 Cuándo usar el modo de mercado

El modo de análisis de zona de mercado es apropiado cuando usted está:

- Construyendo un modelo de regresión para un vecindario o zona de mercado con el fin de entender los factores de valor
- Analizando cómo variables como la superficie, la antigüedad, el tamaño del terreno y la ubicación afectan a los precios de venta en un grupo de propiedades
- Preparando el respaldo de un análisis de condiciones de mercado o de una delimitación de vecindario
- Trabajando con un conjunto de datos en el que desea las funciones de columna especial (edad de la venta, redondeo de coordenadas) pero no necesita el flujo de trabajo de ajustes RCA

***Consejo:** El modo de mercado también es útil para la regresión inmobiliaria general cuando se desean las funciones de columna especial pero no se necesita el flujo de trabajo de ajustes RCA.*

6 Selección de variables

La Sección 4 de la barra lateral — **Variable Configuration** — es donde se elige qué columnas participan en el modelo y cómo se tratan.

6.1 Variable objetivo

El desplegable **Target (response) variable** situado en la parte superior de la Sección 4 enumera todas las columnas de su conjunto de datos. Seleccione una columna como variable respuesta (por ejemplo, el precio de venta). La columna objetivo se excluye automáticamente de la lista de predictores.

6.2 La tabla de predictores

Debajo del selector de objetivo, una tabla enumera todas las columnas restantes con los campos siguientes:

Columna	Descripción
Variable	Nombre de la columna
Type	Desplegable de tipo de dato: numeric , integer , character , Date , POSIXct
Include	Casilla — incluir esta columna como predictor en el modelo
Factor	Casilla — tratar la variable como factor categórico (crea columnas indicadoras en la matriz del modelo)
Force	Casilla — forzar su entrada en el modelo (factor de penalización = 0, nunca descartada por el lasso)
Special	Desplegable (solo en tasación/mercado) — véase la referencia de tipos de columna especiales más abajo
Sign	Signo esperado del coeficiente: positivo, negativo o cualquiera
NAs	Recuento de valores faltantes

6.3 Detección de tipos de datos y anulaciones

glmnetUI detecta automáticamente los tipos de datos al importar. Se reconocen las columnas numéricas, enteras y de fecha. Las columnas de texto que parecen fechas (patrones de formato de fecha habituales) se clasifican como **Date**.

Puede anular cualquier detección cambiando el desplegable **Type**. Cambiar los tipos afecta a cómo se codifica la columna en la matriz del modelo.

Las columnas de texto, TRUE/FALSE y de tipo factor de R se tratan automáticamente como categóricas. Una columna numérica se convierte en factor solo cuando se marca su casilla **Factor** — glmnetUI nunca infiere el carácter categórico a partir del rango de valores de una columna numérica, de modo que predictores numéricos discretos como un recuento de baños permanecen continuos salvo indicación en contrario. El ajuste Factor persiste mediante localStorage.

6.4 Signos esperados y su aplicación

Para cada predictor, el desplegable **Sign** especifica si se espera que su coeficiente sea positivo, negativo o de cualquier signo. Tras el ajuste:

- Los coeficientes con signos inesperados se **marcan en la tabla Coefficients**

- Los avisos de signo aparecen en la pestaña Summary

Cuando la casilla **Enforce Sign Constraints** está activada en los parámetros de glmnet (capítulo 7), se obliga al modelo a producir coeficientes que coincidan con los signos esperados. Esto se implementa mediante `upper.limits` y `lower.limits` en glmnet.

6.5 Referencia de los tipos de columna especiales

En los modos de tasación y de mercado, el desplegable **Special** ofrece las opciones siguientes:

Ponderación:

- **weight** — columna de peso de la observación (disponible en todos los modos; solo se permite una; las filas con peso = 0 se excluyen del ajuste)

Tipos de fecha y hora:

- **contract_date** — activa el cálculo automático de **sale_age** a partir de la fecha efectiva
- **listing_date** — se usa como alternativa para calcular los días en el mercado
- **dom** — identifica la columna de días en el mercado
- **sale_age** — designa una columna de edad de venta ya existente, de modo que cualquier columna pueda servir como edad de la venta en lugar de que glmnetUI calcule una a partir de **contract_date**

Tipos monetarios:

- **concessions** — identifica las concesiones de venta (créditos del vendedor, incentivos al comprador, etc.)

Tipos de transacción:

- **sale_type** — identifica la columna de tipo de venta o transacción

Tipos de tamaño y localización:

- **latitude** — los valores se redondean automáticamente a 3 decimales
- **longitude** — mismo tratamiento de redondeo que la latitud
- **area** — identificador de zona de mercado o de vecindario (por ejemplo, **area_id**, **neighborhood**, **market_area**)
- **living_area** — permite los cálculos de residuos por pie cuadrado (**residual_sf** y **cqa_sf**)
- **lot_size** — columna de tamaño del terreno
- **site_dimensions** — agrupada con el tamaño del terreno (por ejemplo, “75x120”)

Tipos de antigüedad:

- **actual_age** — columna de edad de la propiedad
- **effective_age** — edad efectiva de la propiedad

Tipos de visualización:

- **display_only** — la columna se incluye en las exportaciones a Excel pero queda totalmente excluida del ajuste del modelo. Úsela para campos de dirección, números de MLS, identificadores de parcela u otros datos de referencia. Varias columnas pueden llevar esta designación.

***Consejo:** Las columnas designadas como **display_only** permanecen en su conjunto de datos y aparecen en la exportación a Excel, pero quedan totalmente excluidas de la tabla de predictores. Úselas para columnas de identificador, direcciones u otros campos de referencia.*

7 Selección de parámetros

La Sección 5 de la barra lateral — **glmnet Call Parameters** — da acceso a todas las opciones de configuración del modelo elastic net. Cada parámetro tiene un icono azul de ayuda (?) con una explicación emergente.

7.1 Alpha (parámetro de mezcla)

El parámetro alpha controla el tipo de regularización:

Valor de alpha	Tipo	Comportamiento
0	Ridge	Contrae todos los coeficientes proporcionalmente; nunca fija ninguno en cero
1	Lasso	Puede fijar coeficientes exactamente en cero (selección de variables)
0 a 1	Elastic Net	Combinación de ridge y lasso

Hay dos métodos de selección de alpha disponibles:

- **Fixed** — fijar un único valor de alpha con el deslizador (por omisión: 1,0)
- **Grid Search** — probar varios valores de alpha y seleccionar el que tenga el menor error de CV. Se configuran el rango y el número de valores.

7.2 Selección de lambda

Lambda controla la intensidad de la regularización:

Método	Descripción
Cross-validation	Recomendado. Prueba muchos valores de lambda mediante validación cruzada de k pliegues.
Manual	Introducir un valor concreto de lambda si tiene motivos para ello.

Al usar la validación cruzada hay dos opciones de lambda disponibles:

- **lambda.1se** (por omisión) — el mayor lambda dentro de 1 error estándar del mínimo. Produce modelos más sencillos y más regularizados. Recomendado para el trabajo de tasación.
- **lambda.min** — el lambda que minimiza el error de validación cruzada. Produce el modelo con mejor ajuste, pero potencialmente más complejo.

El parámetro **Number of CV Folds** controla cómo se divide el conjunto de datos para la validación cruzada (por omisión: 10 pliegues).

7.3 Family (familia)

Elija la familia de distribución de su variable respuesta:

Familia	Caso de uso
gaussian	Respuestas continuas (p. ej., el precio de venta). La más habitual.
binomial	Resultados binarios (p. ej., vendido/no vendido).
poisson	Datos de recuento (p. ej., número de ventas).

7.4 Standardize (estandarizar)

Cuando está marcada (por omisión), todos los predictores se escalan para tener media 0 y desviación típica 1 antes del ajuste. Esto garantiza que la penalización trate a todos los predictores por igual, con independencia de su escala original. Los coeficientes se devuelven en la escala original. Normalmente conviene dejarla activada.

7.5 Restricciones de signo

Cuando **Enforce Sign Constraints** está activado, los coeficientes se restringen para que coincidan con los signos esperados fijados en la tabla de variables (capítulo 6):

- Un signo “positive” obliga a que el coeficiente sea ≥ 0
- Un signo “negative” obliga a que el coeficiente sea ≤ 0
- Las variables fijadas en “either” quedan sin restricción

Esto se implementa mediante los parámetros `upper.limits` y `lower.limits` de `glmnet()`.

7.6 Relaxed lasso

El lasso ordinario usa la misma penalización para la selección de variables Y para la estimación de los coeficientes, lo que puede contraer en exceso hacia cero coeficientes importantes. El **relaxed lasso** separa estos pasos:

1. Primero, seleccionar las variables mediante el lasso (la penalización determina qué coeficientes sobreviven)
2. Después, reajustar las variables supervivientes con menos penalización o sin ella para obtener estimaciones menos sesgadas

El deslizador **Gamma** controla el grado de relajación:

Gamma	Efecto
0	Totalmente relajado (reajuste por MCO, sin contracción alguna)
1	Sin relajación (igual que el glmnet ordinario)
Intermedio	Combina los ajustes relajado y penalizado

Con validación cruzada, gamma se elige automáticamente para obtener un rendimiento óptimo.

7.7 Semilla aleatoria

Un campo de texto **Random seed**, precargado con un entero aleatorio, hace reproducible la asignación de pliegues de la validación cruzada — se llama a `set.seed()` antes del ajuste, y la semilla empleada se muestra en la pestaña Glmnet Output y en el mensaje de estado del ajuste. Después de cada ajuste se genera automáticamente una nueva semilla (usando la hora del sistema, no el generador de números aleatorios de R). Las últimas 5 semillas empleadas para el archivo actual se muestran como enlaces en los que se puede hacer clic para recuperarlas fácilmente; el historial de semillas se reinicia al cargar un nuevo archivo.

7.8 Matriz de interacciones

Una matriz triangular superior interactiva permite controlar qué pares de variables pueden interactuar. Cada celda contiene una casilla — marcada significa que el término de interacción $x1:x2$ se incluye en la matriz del modelo.

- **Allow All** — marcar todos los pares de interacción
- **Clear All** — desmarcar todos los pares (el valor por omisión para un archivo nuevo es todos desmarcados)
- **Hacer clic en el nombre de una variable** — alternar todas las interacciones de esa variable

Las selecciones de interacción se guardan en el localStorage del navegador para cada archivo de entrada y se restauran al volver a cargar el mismo archivo.

Cuando hay alguna interacción activada, el informe exportado incluye una matriz **Enabled Interactions** (los predictores en ambos ejes y una marca de verificación donde se incluyó el término producto $x1:x2$), renderizada en apaisado y análoga a la matriz Allowed Interactions de earthUI. Tenga en cuenta que el lasso puede haber contraído a cero el coeficiente de un término activado.

***Consejo:** Con muchos predictores, el lasso anulará automáticamente los términos de interacción innecesarios. Empezar con "Allow All" y dejar que el modelo seleccione es un enfoque razonable.*

Advertencia

Activar las interacciones de glmnet producirá ajustes que pueden resultar difíciles de interpretar y explicar ante un tribunal o en una auditoría. Use las interacciones con cautela y verifique que los ajustes resultantes sean razonables.

7.9 Bloquear el efecto principal (solo interacción)

Haga **clic con el botón derecho** sobre el nombre de una variable en la matriz de interacciones para impedir que entre en el modelo como efecto principal. Aparece un indicador “1” en negrita después del nombre de la variable. Cuando está bloqueada:

- La variable se **elimina** de los términos principales de la fórmula
- La variable **sí puede** aparecer en términos de interacción ($x1:x2$)
- Vuelva a hacer clic con el botón derecho para desbloquearla
- El estado persiste mediante localStorage para cada archivo de entrada

Un caso de uso típico: al modelar ajustes temporales que deben escalar con el tamaño de la propiedad, bloquee **sale_age** como efecto principal y permítalo solo en una interacción con **living_area**. Esto crea un término de interacción **sale_age:living_area** en lugar de un coeficiente **sale_age** plano — la tasa de apreciación temporal varía según el tamaño de la vivienda.

7.10 Parámetros avanzados

Una sección plegable “Advanced” al final de glmnet Call Parameters expone ajustes adicionales. Estos usan valores por omisión razonables, pero son visibles para que toda la configuración del modelo pueda documentarse ante un tribunal o en una auditoría.

- **Lambda Min Ratio** — razón entre el lambda más pequeño y el más grande de la trayectoria de regularización. Por omisión: 0,00001. Para conjuntos de datos pequeños (<50 observaciones), un valor mayor (p. ej., 0,01) puede mejorar el rendimiento.

- **Number of Lambda Values** — número de valores de lambda en la trayectoria. Por omisión: 100.
- **CV Loss Metric** — función de pérdida para la validación cruzada: MSE (por omisión para gaussian), MAE (robusta frente a valores atípicos) o desviación.
- **Convergence Threshold** — umbral de convergencia del descenso por coordenadas. Por omisión: 1e-07.
- **Max Iterations** — número máximo de iteraciones del descenso por coordenadas. Por omisión: 100.000. Auméntelo si aparecen avisos de convergencia.
- **Fit Intercept** — incluir un término de intercepto (constante). Por omisión: activado. El intercepto es el “valor base” en el RCA de tasación.

7.11 Valores predeterminados de los ajustes

Tres opciones controlan cómo se inicializan los parámetros al cargar un archivo:

- **Use last settings for input file** — restaurar exactamente los ajustes empleados la última vez con este archivo
- **Use default settings** — aplicar los valores predeterminados globales guardados
- **glmnet defaults** — restablecer los valores de fábrica (alpha = 1, CV lambda.1se, 10 pliegues, gaussian, standardize activado)

Haga clic en **Save current as default** para almacenar todos los ajustes actuales como valor predeterminado global para archivos futuros.

8 Ajuste del modelo

8.1 El botón de ajuste

La Sección 6 de la barra lateral contiene el botón **Fit Glmnet Model**. Al hacer clic en él se ejecuta el modelo con su configuración actual.

8.2 Qué ocurre durante el ajuste

Cuando hace clic en Fit, glmnetUI:

1. **Prepara los datos** — elimina las filas con NA, aplica los pesos (si se han designado), excluye la fila del sujeto (en modo de tasación) y codifica las variables de factor como columnas indicadoras.
2. **Construye la matriz del modelo** — crea la matriz de diseño incluyendo los términos de interacción permitidos por la matriz de interacciones.
3. **Aplica las restricciones** — si la imposición de signos está activada, fija `upper.limits` y `lower.limits` sobre los coeficientes.
4. **Ejecuta la validación cruzada** — si se ha seleccionado CV, llama a `cv.glmnet()` para hallar el λ óptimo. Si el relaxed lasso está activado, llama a `cv.glmnet(..., relax = TRUE)`.
5. **Ajusta el modelo final** — almacena el modelo ajustado, los coeficientes, las predicciones y los residuos.
6. **Actualiza todas las pestañas de resultados** — se rellenan Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA y Diagnostics.

Un mensaje de estado bajo el botón Fit muestra el número de observaciones, las filas excluidas y el λ seleccionado.

9 Pestañas de resultados

9.1 Data Preview

Muestra los datos importados como una DataTable interactiva. En los modos de tasación y de mercado, la vista previa se divide en dos tablas: **Subject Property** (fila 1) y **Comparable Sales** (filas 2 y siguientes).

Cada celda se mantiene en una sola línea y se trunca al ancho de la columna, de modo que los campos anchos de texto libre (como las observaciones sobre la propiedad) no estiren una fila a lo largo de la página. **Haga doble clic en cualquier celda** para abrir una ventana emergente que muestre su contenido completo, sin truncar.

9.2 Equation

Muestra la ecuación del modelo ajustado renderizada en LaTeX mediante MathJax. Muestra todos los coeficientes no nulos con un formato matemático adecuado:

- Los coeficientes positivos se muestran con signo +
- Los coeficientes negativos se muestran con signo −
- Los términos de interacción se muestran como $x_1 \times x_2$
- Los números se formatean con separadores y con los decimales adecuados

9.3 Correlation

Una matriz en forma de mapa de calor que muestra las correlaciones de Pearson entre todos los predictores numéricos y la variable respuesta. Disponible inmediatamente después de la importación de los datos — no se requiere ajustar el modelo.

- Gradiente de color: azul (−1) → blanco (0) → rojo (+1)
- Los valores de correlación se imprimen en cada celda
- El tamaño del texto y de los ejes se adapta al número de variables
- Útil para identificar multicolinealidad antes del ajuste

9.4 Summary

Estadísticos de ajuste del modelo mostrados como tarjetas:

- **R²** — proporción de varianza explicada
- **R² ajustado** — ajustado por el número de predictores
- **GR²** — R² generalizado a partir de la razón de desviación de glmnet en el lambda seleccionado
- **R² de CV** — R² de validación cruzada (cuando se usa CV)
- **RMSE** — raíz del error cuadrático medio
- **MAE** — error absoluto medio

En los modos de tasación y de mercado, tarjetas adicionales muestran:

- **Median Ratio** — mediana de (predicho / real)
- **COD** — coeficiente de dispersión (cuanto menor, mejor)
- **PRD** — diferencial relacionado con el precio

Aparece un **aviso de sobreajuste** cuando el R² de entrenamiento supera al R² de CV en más de 0,1.

Debajo de las tarjetas, una tabla de coeficientes con avisos de signo duplica la pestaña Coefficients por comodidad.

9.5 Coefficients

Una tabla de todos los coeficientes no nulos que muestra:

- Nombre de la variable
- Valor del coeficiente
- Signo esperado (de la configuración de variables)
- Marca de aviso de signo si la dirección del coeficiente no coincide con lo esperado

9.6 Variable Importance

Magnitud del coeficiente estandarizado: $|\beta| \times \text{sd}(x)$ para cada predictor, agregada entre las columnas indicadoras en el caso de las variables de factor. Dos modos de visualización:

- **plotly interactivo** (cuando el paquete `plotly` está instalado): diagrama de barras horizontales con información emergente que muestra la importancia exacta, el coeficiente y el porcentaje relativo.
- **ggplot2 estático** (alternativa): diagrama de barras horizontales con etiquetas de eje formateadas con separadores de millares.

Debajo del gráfico, una DataTable muestra Variable, Importance, Coefficient y Relative %.

9.7 Contributions

Gráficos de efecto parcial por variable que muestran la contribución de cada predictor a las predicciones del modelo. Seleccione una variable en el desplegable.

Para predictores numéricos: un diagrama de dispersión de los valores x de la variable frente a su contribución ($\beta \times x$), con un segmento de línea rojo y una etiqueta de pendiente. La etiqueta de pendiente muestra el efecto marginal por unidad (por ejemplo, $+1,234.56/\text{unit}$), con unidades adaptativas para variables de rango pequeño como la latitud y la longitud (por ejemplo, $+0.12/0.001$). El subtítulo muestra el valor del intercepto (base).

Para variables de factor: un diagrama de caja que muestra la distribución de los valores de contribución por nivel del factor.

Para interacciones de dos variables: tres visualizaciones apiladas — un gráfico de superficie 3D (`plotly interactivo` cuando está disponible, o `persp()` estático en caso contrario) de la contribución conjunta, un diagrama de dispersión con los puntos coloreados según la contribución y un mapa de calor de la contribución media sobre los rangos de ambas variables.

9.8 ANOVA

Tabla de descomposición de la varianza que muestra, para cada predictor:

- **SS** — suma de cuadrados: $\sum (\text{contribution}_j - \overline{\text{contribution}_j})^2$
- **% de la SS del modelo** — porcentaje de la suma de cuadrados total del modelo
- **Coefficient** — el coeficiente dominante de ese predictor

Incluye una fila de intercepto y una fila TOTAL MODEL. Los términos de interacción (que contienen “:”) se muestran como grupos separados.

9.9 Diagnostics

Cuatro gráficos de diagnóstico con fuentes grandes (tamaño base 16 pt), 15 marcas de eje y etiquetas con separadores de millares (sin notación científica):

1. **Coefficient Path** (trayectoria de coeficientes) — muestra cómo cambian todos los coeficientes a medida que aumenta $\log(\lambda)$. Las líneas se colorean por variable. Ayuda a visualizar la trayectoria de regularización y la selección de variables.

2. **CV Error** — error de validación cruzada en función de $\log(\lambda)$. Los puntos muestran el error medio de CV; las barras de error muestran ± 1 error estándar. Las líneas verticales discontinuas marcan $\lambda_{\min}$ (azul) y λ_{1se} (verde).
3. **Actual vs Predicted** — diagrama de dispersión de los valores observados frente a los predichos. Los puntos deberían agruparse en torno a la línea de referencia discontinua a 45 grados. Una dispersión mayor indica más error de predicción.
4. **Residuals vs Fitted** — diagrama de dispersión de los residuos frente a los valores ajustados. Debería mostrar una dispersión aleatoria en torno a la línea del cero (discontinua). La presencia de patrones sugiere una especificación incorrecta del modelo (p. ej., no linealidad, heterocedasticidad).

Cada gráfico tiene un botón **Download PNG** para guardarlo a 150 DPI.

9.10 Glmnet Output

Salida bruta del modelo que muestra: la semilla aleatoria empleada, la impresión del modelo, el λ seleccionado, $\lambda_{\min}/\lambda_{1se}$, γ (si el relaxed lasso está activado) y el vector completo de coeficientes.

9.11 Report

Permite configurar los campos del informe (nombre del tasador, dirección de la propiedad, fecha del informe, número de expediente) y exportarlo a Word (.docx) o PDF directamente desde la pestaña. Para el flujo de trabajo completo de informes basado en Quarto (HTML, Word o PDF), véase el capítulo 13.

10 Descarga de datos

Tras el ajuste, el botón **Download Output (Excel)** (sección 7 de la barra lateral, disponible en todos los modos de uso) exporta un archivo de Excel con las predicciones y los diagnósticos a la carpeta de salida del proyecto. El encabezado de la sección dice “Download Estimated Sale Prices & Residuals” en los modos de tasación y de mercado, y “Download Estimated Target Variable(s) & Residuals” en el modo General.

10.1 Columnas de salida

Columna	Descripción
<code>est_<target></code>	Predicción del modelo (p. ej., <code>est_sale_price</code>)
<code>residual</code>	Real – predicho
<code>cqa</code>	Puntuación de condición-calidad-atractivo (escala 0–10)
<code>residual_sf</code>	Residuo / superficie habitable (si se ha designado <code>living_area</code>)
<code>cqa_sf</code>	CQA calculada a partir del orden según <code>residual_sf</code>
<code><var>_contribution</code>	Contribución de cada variable a la predicción
<code>basis</code>	Valor del intercepto (el mismo para todas las filas)

Una marca de verificación blanca aparece en el botón de descarga tras completarse correctamente.

10.2 Puntuaciones CQA

La CQA ordena el residuo de cada fila frente a todos los demás en una escala de 0 a 10:

- **CQA alta (~9–10):** se vendió por mucho más de lo predicho — probablemente condición, calidad o atractivo superiores
- **CQA baja (~0–1):** se vendió por mucho menos de lo predicho — probablemente condición inferior o venta forzada
- **CQA ~5:** cerca de la mediana

En los modos de tasación y de mercado, las filas se ordenan por `residual_sf` de forma descendente cuando se ha designado una columna `living_area`.

11 Cálculos RCA y descarga

El flujo de trabajo RCA (Residual Constraint Approach) está disponible en los modos **Appraisal** y **Market Area Analysis**, después de ajustar el modelo. La Sales Comparison Grid subsiguiente (capítulo 12) es exclusiva de la tasación.

11.1 Apertura del diálogo RCA

Haga clic en el botón **Calculate RCA Adjustments & Download** de la sección 8 de la barra lateral. Elija:

- El tipo de puntuación **CQA** o **CQA per SF**. Si se selecciona “CQA per SF”, el residuo del sujeto se escala por la superficie habitable.
- Introduzca la puntuación CQA del sujeto (0,00–9,99; por omisión 5,00)

11.2 Interpolación de la puntuación CQA

1. Se ordenan las puntuaciones CQA y los residuos de los comparables
2. Una interpolación lineal asocia su valor de CQA a un residuo
3. Valor del sujeto = predicción del modelo + residuo interpolado

11.3 Columnas de salida

Columna	Descripción
subject_value	Predicción del modelo + residuo interpolado
<var>_adjustment	Contribución del sujeto – contribución del comparable
residual_adjustment	Residuo del sujeto – residuo del comparable
net_adjustments	Suma de todos los ajustes
gross_adjustments	Suma de los valores absolutos de los ajustes
adjusted_sale_price	Precio de venta del comparable + ajustes netos

Una marca de verificación blanca aparece en el botón tras completarse correctamente el cálculo.

12 Sales Comparison Grid

La Sales Comparison Grid (cuadrícula de comparación de ventas) está disponible únicamente en el modo **Appraisal**, después de calcular los ajustes RCA (paso 8). Genera un libro de Excel con formato apto para incluirse en los informes de tasación.

12.1 Diálogo de selección de comparables

Al hacer clic en el botón **Generate Sales Grid & Download** de la sección 9 de la barra lateral se abre un diálogo modal que enumera todas las ventas comparables de la salida RCA. Los comparables se dividen en dos grupos:

- **Recommended comps** (marcados de antemano): ajuste bruto < 25 % del precio de venta, ordenados por edad de la venta (los más recientes primero).
- **Other comps** (desmarcados por omisión): los comparables restantes, ordenados por edad de la venta.

Seleccione hasta 30 comparables y haga clic en **Generate Sales Grid** para crear el libro.

12.2 Estructura del libro

El libro generado contiene hasta 10 hojas, cada una con 3 comparables uno al lado del otro:

- **Filas de encabezado:** direcciones del sujeto y de los comparables, precios de venta, fechas de venta y días en el mercado
- **Filas agrupadas:** localización (latitud, longitud, zona), tamaño del terreno (superficie del lote, dimensiones del sitio) y antigüedad (edad real, edad efectiva) con encabezados de grupo y subelementos
- **Filas de variables del modelo:** una fila por cada coeficiente no nulo del modelo que muestra la contribución del sujeto, las contribuciones de los comparables y los porcentajes de ajuste
- **Filas de características residuales:** celdas vacías y desbloqueadas para que el tasador introduzca características no captadas por el modelo (p. ej., condición, calidad, vistas)
- **Filas de resumen:** % de ajuste neto, % de ajuste bruto y precio de venta ajustado, con fórmulas activas de Excel

12.3 Protección de la hoja y celdas residuales

Cada hoja está protegida con una contraseña para evitar ediciones accidentales, pero las celdas de valor de las características residuales están explícitamente desbloqueadas. Esto permite a los tasadores introducir valores para características ausentes del modelo sin dejar de preservar los cálculos de ajuste basados en fórmulas.

13 Descarga de informes

13.1 El flujo de trabajo Quarto de dos pasos

La exportación de informes desde la barra lateral es un flujo de trabajo Quarto de dos pasos:

1. **Generate Quarto Report** (sección 10 de la barra lateral) — escribe un paquete Quarto autocontenido (el código fuente `.qmd` más los gráficos, `report_data.rds` y `reference.docx`) en la carpeta de salida de glmnet del proyecto.
2. **Convert Quarto Report** (sección 11 de la barra lateral) — renderiza un archivo `.qmd` a los formatos de salida elegidos. La ruta apunta por omisión al paquete generado más recientemente, pero puede navegar hasta cualquier `.qmd`.

13.2 Formatos de informe

En el paso Convert hay tres formatos disponibles:

- **HTML** — documento HTML autocontenido. Funciona en todas las plataformas sin más software que pandoc. Opción recomendada por omisión.
- **Word (.docx)** — documento con formato. Adecuado para su edición y distribución. Funciona en todas las plataformas.
- **PDF** — generado mediante Quarto con LaTeX. El tamaño de papel sigue la configuración regional (Letter o A4). **Requiere una instalación de LaTeX** (véase el apéndice de requisitos del sistema). Si no se detecta LaTeX, la opción de PDF se oculta automáticamente de las casillas de formato.

Los informes se generan con Quarto cuando está disponible, con rmarkdown como alternativa. La pestaña Report ofrece además botones directos de exportación a Word y a PDF.

13.3 Contenido del informe

Los informes incluyen el contenido de todas las pestañas:

- Descripción del conjunto de datos (nombre del archivo, observaciones, respuesta, predictores, propósito)
- Especificación del modelo (alpha, lambda, family, standardize, relaxed, coeficientes no nulos)
- Resumen de resultados (R^2 , R^2 ajustado, R^2 generalizado, R^2 de CV, RMSE, MAE)
- Ecuación del modelo (renderizada en LaTeX)
- Tabla de coeficientes
- Importancia de las variables (diagrama de barras y tabla)
- Gráficos de contribución por variable equivalentes a los de la pestaña Contributions:
 - **Predictores numéricos:** diagrama de dispersión con línea de ajuste lineal
 - **Variables de factor:** diagrama de caja por nivel
 - **Interacciones:** dispersión coloreada según la contribución, mapa de calor de la contribución media e instantánea estática de la superficie 3D (persp)
- Matriz Enabled Interactions (cuando hay alguna interacción activada) — los predictores en ambos ejes, con una marca de verificación donde se incluyó el término producto $x_1:x_2$, renderizada en apaisado
- Mapa de calor de la matriz de correlaciones
- Tabla de descomposición ANOVA
- Gráficos de diagnóstico:
 - Trayectoria de coeficientes

- Error de validación cruzada
 - Valores reales frente a predichos
 - Residuos frente a valores ajustados
 - Gráfico Q-Q normal
- Salida del modelo

Todas las etiquetas de eje y las leyendas de color usan números con separador de millares (sin notación científica).

Una marca de verificación blanca aparece en el botón del informe tras generarse correctamente, y se muestra un cronómetro mientras se renderiza el informe.

14 Comparación con earthUI

La aplicación earth y su Post Processing Module de glmnet forman un único conjunto de herramientas para el modelado de regresión. Comparten el mismo formato de datos, los mismos tipos de columna especiales, el mismo flujo de trabajo RCA y los mismos conjuntos de datos de demostración, pero usan motores de modelado distintos.

14.1 Diferencias principales

Característica	glmnetUI	earthUI
Método	Elastic net (glmnet)	MARS/Earth (earth)
Selección de variables	Automática mediante la penalización lasso	Automática mediante poda hacia adelante/atrás
Relaciones de los coeficientes	Solo efectos lineales	Lineales por tramos (funciones bisagra)
Interacciones	Por pares mediante la matriz de interacciones	Hasta grado 3
Signos de los coeficientes	Imponibles mediante restricciones	No imponibles directamente
Ajuste relajado	Relaxed lasso disponible	No aplicable
Ajuste de alpha	Búsqueda en malla disponible	No aplicable (no hay parámetro de mezcla)
Sales Comparison Grid	Disponible (paso 9)	Disponible (paso 8)
Gráficos de funciones g	No aplicable	Visualización de términos agrupados
Formatos de informe	HTML, Word, PDF	HTML, Word, PDF

14.2 Funciones compartidas

Ambas herramientas ofrecen:

- Importación de CSV y Excel con conversión de columnas a snake_case
- Compatibilidad con configuraciones regionales de 31 países, con separador de CSV, signo decimal y tamaño de papel
- Designaciones especiales de columnas (contract_date, living_area, area, latitude, longitude, etc.)
- Cálculo automático de sale_age a partir de la fecha efectiva
- Alternancia entre modo claro y oscuro
- Persistencia de los ajustes mediante localStorage
- Mapa de calor de correlaciones, importancia de las variables, contribuciones, ANOVA, diagnósticos
- Puntuación CQA y flujo de trabajo de ajustes RCA
- Sales Comparison Grid con selección de comparables y salida a Excel con formato
- Compatibilidad con los mismos conjuntos de datos de demostración (Appraisal_1.csv)

14.3 Cuál usar en cada caso

- Use **glmnetUI** cuando desee un modelo lineal sencillo e interpretable con selección automática de variables. El elastic net resulta especialmente adecuado cuando se dispone de muchos predictores y se desea que el modelo seleccione los más importantes.

- **Use earthUI** cuando espere relaciones no lineales (por ejemplo, que el efecto de la superficie habitable sobre el precio cambie según el tamaño) o necesite funciones bisagra para captar efectos de umbral. Los modelos earth también pueden captar interacciones de orden superior (grados 2 y 3). earthUI ofrece además una función experimental de extracción de observaciones basada en Prolog/DCG que no tiene equivalente en glmnetUI.
- **Use ambos** para comparar modelos lineales y no lineales sobre los mismos datos. Si ambos producen valores de R^2 similares, puede preferirse el modelo glmnet, más sencillo, por su interpretabilidad.

15 Conjunto de datos de demostración: Appraisal_1.csv

15.1 Descripción

El conjunto de datos de demostración se distribuye con glmnetUI (y se comparte con earthUI). Localícelo con:

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

El archivo contiene 1.502 ventas residenciales (más 1 propiedad sujeto en la fila 1) procedentes de una exportación simulada de un MLS. Los datos representan ventas de viviendas unifamiliares en un mercado con varias zonas, con una variedad de tamaños, antigüedades y ubicaciones.

No se trata de datos reales, sino de datos basados en un vecindario realista del norte de California. Toda la información identificativa ha sido alterada o eliminada.

15.2 Columnas

Columna	Tipo	Tipo especial	Descripción
weight	numeric	weight	Peso de la observación (0 = excluir del ajuste)
id	numeric	display_only	Identificador interno del registro
property_id	numeric	display_only	Identificador de propiedad del MLS
listing_id	character	display_only	Número de listado del MLS
parcel_number	character	display_only	Número de parcela del tasador del condado (APN)
street_address	character	display_only	Dirección de la propiedad
city_name	character	display_only	Ciudad
postal_code	character	display_only	Código postal
county_name	character	display_only	Condado
contract_date	Date	contract_date	Fecha del contrato de venta (calcula sale_age)
sale_age	numeric	—	Días desde la fecha de contrato hasta la fecha efectiva
sale_price	numeric	(objetivo)	Precio de venta — variable respuesta
living_sqft	numeric	living_area	Superficie habitable bruta en pies cuadrados
beds_total	integer	—	Número de dormitorios
baths_total	numeric	—	Recuento total de baños (p. ej., 2,5 = 2 completos + 1 medio)
lot_size	numeric	lot_size	Superficie del lote en pies cuadrados
area_id	integer	area	Identificador de zona del MLS
age	numeric	actual_age	Edad de la propiedad en años
latitude	numeric	latitude	Latitud (redondeada a 3 decimales para el modelo)
longitude	numeric	longitude	Longitud (redondeada a 3 decimales para el modelo)
garage_spaces	integer	—	Número de plazas de garaje
days_on_market	integer	dom	Días en el mercado
listing_date	Date	listing_date	Fecha de listado
sale_concessions	numeric	concessions	Concesiones del vendedor

15.3 Inicio rápido sugerido

1. Inicie la rutina de glmnet: `earthUI::launch_glmnet()`
2. Cree un proyecto nuevo (Sección 1) con el propósito **For Appraisal**
3. Coloque **Appraisal_1.csv** en la carpeta de entrada del proyecto y selecciónelo en la Sección 2 (haga clic en **Refresh** si es necesario)
4. Seleccione **sale_price** como objetivo
5. Asigne los tipos especiales como se muestra en la tabla anterior
6. Incluya los predictores: **sale_age**, **living_sqft**, **baths_total**, **lot_size**, **area_id** (como factor), **age**, **latitude**, **longitude**, **garage_spaces**
7. Mantenga $\alpha = 1$ (Lasso), CV con `lambda.1se`
8. Haga clic en **Fit Glmnet Model**
9. Revise las pestañas Summary, Coefficients y Diagnostics
10. Descargue la salida (paso 7) y revise el orden por CQA
11. Calcule los ajustes RCA (paso 8) con una puntuación CQA de ~5,00
12. Genere una Sales Comparison Grid (paso 9) con los comparables recomendados

16 Requisitos del sistema y resolución de problemas

16.1 Plataformas admitidas

glmnetUI funciona en **macOS**, **Windows** y **Linux**. Se recomienda encarecidamente RStudio Desktop (2023.06 o posterior) — incluye **pandoc**, necesario para los informes.

16.2 Dependencias opcionales

- **LaTeX** (TinyTeX, MiKTeX o MacTeX): solo para la generación de informes en PDF. Si no se detecta, la opción de PDF se oculta automáticamente. Instálelo con: `tinytex::install_tinytex()`
- **sysfonts + showtext**: fuente Roboto Condensed para los gráficos. Recurre automáticamente a la fuente sans-serif del sistema si no está disponible o no hay conexión.
- **earth** ($\geq 5.3.0$): solo necesario para la canalización de importación desde earthUI.

16.3 Notas por plataforma

macOS: la que menos problemas presenta. Instale TinyTeX para el PDF: `tinytex::install_tinytex()`

Windows: funciona bien con RStudio. Los equipos corporativos o con restricciones pueden tener limitaciones en el directorio temporal — la aplicación avisará en la consola pero continuará sin persistencia de los ajustes.

Linux: puede necesitar bibliotecas del sistema: `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

16.4 Degradación elegante

Componente ausente	Comportamiento
Sin LaTeX	La opción de PDF se oculta. HTML y Word siguen disponibles.
Sin internet / fallan las fuentes	Se usa la sans-serif del sistema. Se registra un mensaje en la consola.
Sin RSQLite / sistema de archivos de solo lectura	Los ajustes no persisten. La aplicación funciona con normalidad.
Directorio temporal sin permiso de escritura	La generación del informe falla con un error claro. Fije TMPDIR en una ubicación con permiso de escritura.

16.5 Resolución de problemas

“**PDF option not available**”: ejecute `tinytex::install_tinytex()` en R y reinicie la aplicación.

“**Settings will not persist**”: compruebe los permisos del directorio de datos del usuario (macOS: `~/Library/Application Support/R/glmnetUI/`, Linux: `~/.local/share/R/glmnetUI/`, Windows: `%APPDATA%/R/data/glmnetUI/`).

El puerto 7879 ya está en uso: ejecute `lsof -ti:7879 | xargs kill` (macOS/Linux) o `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell).

Referencias

- Friedman, Jerome, Trevor Hastie y Robert Tibshirani (2010). «Regularization Paths for Generalized Linear Models via Coordinate Descent». En: *Journal of Statistical Software* 33.1, págs. 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». En: *Journal of Statistical Software* 39.5, págs. 1-13. DOI: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).
- Tay, J. Kenneth, Balasubramanian Narasimhan y Trevor Hastie (2023). «Elastic Net Regularization Paths for All Generalized Linear Models». En: *Journal of Statistical Software* 106.1, págs. 1-31. DOI: [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).

mgcvUI: una interfaz interactiva para el paquete mgcv (modelos aditivos generalizados)

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

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

Este artículo de acceso abierto se distribuye bajo los términos de la [licencia Creative Commons Atribución 4.0](#).

Resumen. mgcvUI es una interfaz gráfica de usuario para el paquete de R mgcv, que ajusta modelos aditivos generalizados (GAM) mediante splines de regresión penalizada con selección automática de suavidad. Ofrece tres modos de uso — modelado predictivo general, tasación inmobiliaria y análisis de zona de mercado — y guía al usuario a través de la importación de datos, la especificación de términos suaves, el ajuste del modelo, los gráficos de diagnóstico y de efectos, y los informes descargables. Este artículo documenta los requisitos de formato de datos de mgcvUI, el flujo de trabajo de modelado, las pantallas de salida y la referencia completa de funciones.

Palabras clave: modelos aditivos generalizados, GAM, splines penalizados, suavizado, mgcv, R, interfaz gráfica de usuario, valuación inmobiliaria

Nota del editor (julio de 2026): Poco antes de la publicación, mgcvUI se consolidó dentro de earthUI (versión 0.11.0 y posteriores), que ahora ofrece el flujo de trabajo de GAM aquí descrito junto con los flujos de trabajo de MARS y elastic net. mgcvUI sigue disponible en su forma final publicada (0.3.0) en <https://github.com/wcraytor/mgcvUI>; el desarrollo futuro continúa en earthUI. Este artículo se refiere a mgcvUI tal como fue publicado.

Índice

1. Introducción	4
1.1. Historia y antecedentes	4
1.1.1. Modelos aditivos generalizados	4
1.1.2. El paquete mgcv	4
1.1.3. Comparación con MARS (earth) y elastic net (glmnet)	5
1.1.4. La cadena earthUI–mgcvUI	5
1.1.5. Flujo de trabajo recomendado	6
1.2. ¿Qué es mgcvUI?	6
1.3. Tres modos de uso	7
1.4. Funciones específicas del sector inmobiliario	7
1.5. Primeros pasos	7
2. Requisitos de los datos de entrada del MLS	9
2.1. Formato y estructura del archivo	9
2.2. Columnas requeridas en el modo de tasación	9
2.3. Convenciones de nombres de columnas especiales	10
2.4. Calidad e integridad de los datos	10
2.5. Ubicación de la fila del sujeto	10

3. Modo de uso general	11
3.1. Panorama general	11
3.2. El flujo de trabajo de la barra lateral	11
3.3. Importar desde earthUI (opcional)	12
3.4. Pestañas del panel principal	12
3.5. Persistencia de los ajustes	13
3.6. Modo oscuro	13
3.7. Configuración regional y de idioma	13
4. Modo de tasación	15
4.1. Tratamiento de la fila del sujeto	15
4.2. Fecha efectiva y antigüedad de la venta	15
4.3. Designaciones especiales de columnas	15
4.4. Panorama de los ajustes RCA	15
5. Modo de análisis de zona de mercado	16
5.1. Diferencias respecto del modo de tasación	16
5.2. Cuándo usar el modo de mercado	16
6. Selección de variables	17
6.1. Variable objetivo	17
6.2. Transformación de la respuesta	17
6.3. La tabla de predictores	17
6.4. Detección de tipos de datos y anulaciones	18
6.5. Referencia de tipos especiales de columna	18
7. Selección de parámetros	19
7.1. Preajustes de parámetros	19
7.2. Family	19
7.3. Method	19
7.4. Gamma (multiplicador de suavizado)	20
7.5. Validación cruzada	20
7.6. Select (contracción)	20
7.7. Base predeterminada	20
7.8. k predeterminada (dimensión de la base)	20
7.9. Tipo de tensor	21
7.10. Interacciones permitidas	21
7.11. Interacciones factor por término suave	21
7.12. Parámetros avanzados	22
8. Ajuste del modelo	23
8.1. El botón de ajuste	23
8.2. Qué ocurre durante el ajuste	23
8.3. Ajuste en segundo plano y progreso	23
9. Pestañas de resultados	24
9.1. Data	24
9.2. Equation	24
9.3. Summary	24
9.4. Variable Importance	24
9.5. Gráficos de contribución	25
9.5.1. Términos suaves univariantes — $s(x)$	25

9.5.2. Términos factor por término suave — $s(x, by = factor)$	25
9.5.3. Términos de interacción — $ti(x_1, x_2)$ y $te(x_1, x_2)$	25
9.5.4. Términos paramétricos: lineales y de factor	25
9.6. Correlation	25
9.7. Diagnostics	26
9.8. RCA Adjustments	26
9.9. Anova	26
9.10. Mgev Output	26
9.11. Sign Check	26
9.12. Concurvity	26
9.13. Basis Check	27
10.Descarga de datos	28
10.1. Columnas de salida	28
10.2. Puntuaciones CQA	28
11.Cálculos RCA y descarga	29
11.1. Abrir el diálogo RCA	29
11.2. Interpolación de la puntuación CQA	29
11.3. Columnas de salida	29
12.Cuadrícula de comparación de ventas	30
12.1. Diálogo de selección de comparables	30
12.2. Disposición del libro	30
12.3. Protección de las hojas y celdas residuales	30
13.Descarga de informes	31
13.1. Paso 1: generar el informe Quarto	31
13.2. Paso 2: convertir el informe Quarto	31
13.3. Contenido del informe	31
14.Exportación de las funciones del modelo	33
14.1. Funciones de R	33
14.2. Funciones de Python	33
14.3. Cabecera de C++	33
14.4. Base de datos SQLite	33
15.Comparación con earthUI y glmnetUI	34
15.1. Diferencias principales	34
15.2. Funciones compartidas	34
15.3. Flujo de trabajo recomendado	35
16.Conjunto de datos de demostración: Appraisal_1.csv	36
16.1. Descripción	36
16.2. Columnas	36
16.3. Inicio rápido sugerido	37
17.Requisitos del sistema y resolución de problemas	38
17.1. Plataformas admitidas	38
17.2. Versión de R	38
17.3. Paquetes de R requeridos	38
17.4. Dependencias opcionales	39
17.5. Notas específicas de cada plataforma	39

17.5.1. macOS	39
17.5.2. Windows	39
17.5.3. Linux	39
17.6. Degradación elegante	40
17.7. Resolución de problemas	40
17.7.1. “PDF option is not available”	40
17.7.2. “Settings will not persist”	40
17.7.3. “Diagnostics plot unavailable”	40
17.7.4. “Report error: number of items to replace...”	41
17.7.5. El puerto 7880 ya está en uso	41
17.7.6. La instalación de paquetes falla en Linux	41

1 Introducción

1.1 Historia y antecedentes

1.1.1. Modelos aditivos generalizados

Los **modelos aditivos generalizados** (GAM) fueron formalizados por Trevor Hastie y Robert Tibshirani en su monografía de 1990. Los GAM extienden el modelo lineal sustituyendo cada término lineal $\beta_j x_j$ por una función suave $f_j(x_j)$, de modo que el modelo queda así:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

Cada f_j se estima a partir de los datos mediante una spline de regresión penalizada. Las funciones suaves permiten que el modelo capte relaciones no lineales sin que el usuario tenga que especificar de antemano una forma funcional. La penalización controla la ondulación de cada término suave, lo que evita el sobreajuste sin dejar de permitir la flexibilidad necesaria para seguir los patrones reales.

1.1.2. El paquete mgcv

El paquete de R **mgcv** (Mixed GAM Computation Vehicle) fue desarrollado por Simon Wood en la Universidad de Bath (Wood 2017). Publicado por primera vez en 2001, es hoy la implementación estándar de los GAM en R y se distribuye con cada instalación de R. Sus métodos subyacentes se describen en una serie de artículos que abarcan la estimación estable de múltiples parámetros de suavizado (Wood 2004), las splines de regresión thin-plate (Wood 2003), la estimación REML rápida y estable (Wood 2011) y la selección general de modelos suaves (Wood, Pya y Säfken 2016). Entre sus características principales se incluyen:

- **Selección automática de suavidad:** el parámetro de suavizado de cada término se estima como parte del ajuste del modelo, mediante REML (máxima verosimilitud restringida), GCV (validación cruzada generalizada) o ML (máxima verosimilitud).
- **Múltiples tipos de base:** splines de regresión thin-plate (tp), splines de regresión cúbica (cr), P-splines (ps) y B-splines (bs), cada uno con propiedades distintas.
- **Términos suaves de producto tensorial:** `te()` y `ti()` permiten interacciones suaves entre variables sin incurrir en la maldición de la dimensionalidad.
- **Selección de variables:** la opción `select = TRUE` añade una penalización adicional que puede contraer términos suaves hasta cero, con lo que realiza de hecho una selección de variables.
- **Diagnósticos de concurvidad:** miden el análogo de la multicolinealidad para los términos suaves.

1.1.3. Comparación con MARS (earth) y elastic net (glmnet)

La aplicación earth y sus dos Post Processing Modules (módulos de posprocesamiento; glmnet y mgcv) utilizan motores de modelado distintos. La tabla siguiente resume las diferencias principales:

Característica	mgcv (GAM)	earth (MARS)	glmnet (Elastic Net)
Tipo de modelo	No lineal suave (splines penalizadas)	Lineal por tramos (splines adaptativas)	Lineal (con regularización)
No linealidad	Automática mediante funciones suaves (thin plate, cúbicas, etc.)	Automática mediante funciones bisagra con nudos guiados por los datos	Solo mediante interacciones o expansión de base
Selección de variables	Opcional mediante contracción de doble penalización (select=TRUE)	Incorporada mediante pasos hacia adelante/atrás con poda por GCV	Incorporada mediante penalización L_1 (Lasso)
Interacciones	Términos suaves de producto tensorial (te(), ti())	Productos basados en bisagras (interpretables, hasta grado 3)	Productos cruzados por pares (difíciles de interpretar)
Interpretabilidad	Curvas suaves de efecto parcial; muy intuitivas	Las funciones g muestran efectos parciales lineales por tramos para cada variable	Los coeficientes son pesos lineales; fáciles de explicar individualmente
Más indicado para	Relaciones no lineales suaves, modelado flexible	Detección automática de no linealidad e interacciones	Datos de alta dimensión, selección de variables, muestras pequeñas
Ajustes RCA	Ajustes suaves mediante efectos parciales	Ajustes lineales por tramos mediante funciones g	Ajustes lineales por variable
Control del sobreajuste	Estimación del parámetro de suavizado por REML/GCV	Poda basada en GCV	Lambda validada por validación cruzada

MARS (Multivariate Adaptive Regression Splines) fue introducido por Jerome Friedman en 1991. Construye modelos lineales por tramos seleccionando de forma adaptativa funciones bisagra y las posiciones de sus nudos a partir de los datos. La implementación en R es el paquete earth de Stephen Milborrow.

La regresión **elastic net**, implementada en el paquete glmnet de Friedman, Hastie y Tibshirani (2010), combina las penalizaciones Lasso (L_1) y ridge (L_2) para realizar simultáneamente la selección de variables y la contracción de los coeficientes.

1.1.4. La cadena earthUI–mgcvUI

Un flujo de trabajo potente combina el descubrimiento automático de nudos de earthUI con la estimación suave de mgcvUI. Cuando earthUI exporta un archivo de resultados .rds y mgcvUI lo importa:

1. **Earth descubre los nudos** — MARS encuentra posiciones de bisagra guiadas por los datos (puntos de cambio) para cada variable.
2. **mgcvUI las usa como anclas** — los nudos de earth se convierten en la base de las splines de regresión cúbica (base `cr`) del GAM, lo que da a los términos suaves una ventaja inicial con posiciones de nudo que reflejan la estructura real de los datos.
3. **mgcv refina el ajuste** — la estimación por splines penalizadas suaviza el ajuste lineal por tramos de earth y produce curvas suaves de efecto parcial, al tiempo que conserva la forma general descubierta por earth.

Esta cadena tiende un puente entre el modelado exploratorio (earth) y el modelado confirmatorio (GAM), y combina lo mejor de ambos enfoques.

1.1.5. Flujo de trabajo recomendado

Para la tasación inmobiliaria y aplicaciones similares que requieren modelos interpretables y defendibles:

1. **Empiece con earthUI** para descubrir las variables importantes, las relaciones no lineales y las interacciones presentes en sus datos. La selección automática de base de earth y sus funciones `g` ofrecen excelentes intuiciones iniciales.
2. **Refine con mgcvUI** si desea efectos parciales más suaves. Importe los nudos de earthUI en mgcvUI para pasar sin fricción de los modelos lineales por tramos a los modelos suaves.
3. **Use glmnetUI** cuando necesite una selección de variables agresiva (Lasso), cuando el número de predictores se aproxime o supere al número de observaciones, o cuando desee un modelo puramente lineal con regularización por motivos de defendibilidad.

1.2 ¿Qué es mgcvUI?

mgcvUI es una interfaz gráfica de usuario para el paquete `mgcv` de R. Se ejecuta como una aplicación Shiny local — no hay inicio de sesión, ni servidor, ni cuentas. Se inicia desde R, se importa un conjunto de datos (CSV o Excel), se configura el modelo y se ajusta un GAM de manera interactiva.

A partir de la versión actual, mgcvUI ya no es un paquete que se instale por separado: se distribuye *dentro* del paquete earthUI como su Post Processing Module de GAM, y se inicia con `earthUI::launch_mgcv()`. La aplicación en sí — y todo lo descrito en este artículo — no ha cambiado; solo difieren la instalación y el comando de inicio. Las tres rutinas (earth, glmnet, mgcv) comparten un mismo árbol de proyectos, una misma base de datos de configuración y una misma copia de la infraestructura de soporte.

La aplicación proporciona un flujo de trabajo completo: importación de datos, configuración de variables con especificación de términos suaves, ajuste del modelo con procesamiento en segundo plano, gráficos de diagnóstico, curvas suaves de efecto parcial e informes descargables en formato Word, PDF o HTML.

Los **modelos aditivos generalizados** sustituyen la relación lineal βx por una función suave $f(x)$ para cada predictor. Esto significa que:

- **No hace falta especificar formas funcionales** — son los datos los que determinan la forma de cada relación.
- **Efectos parciales suaves** — la contribución de cada predictor es una curva suave, no una recta, lo que facilita ver cómo una variable afecta a la respuesta en distintos valores.
- **Selección automática de suavidad** — mgcv determina cuán ondulada debe ser cada curva mediante REML o GCV, equilibrando el ajuste frente al sobreajuste.

1.3 Tres modos de uso

Todo proyecto tiene un **purpose** (propósito o modo de uso), que se elige al crearlo en la Sección 1 de la barra lateral (**Project**) y queda fijado a partir de entonces. El propósito del proyecto activo determina qué herramientas y elementos de la interfaz aparecen; para trabajar en otro modo, cree o abra un proyecto con ese propósito. Los tres modos son:

- **General** — regresión GAM para cualquier tipo de población o conjunto de datos. Este es el modo predeterminado. Ofrece el flujo de trabajo completo de GAM sin añadidos específicos de ningún dominio.
- **For Appraisal** (para tasación) — regresión GAM adaptada a la tasación inmobiliaria. Añade funciones específicas para la valuación de una sola propiedad, incluidos el tratamiento de la propiedad sujeto, las designaciones especiales de columnas y el Residual Constraint Approach (RCA).
- **Market Area Analysis** (análisis de zona de mercado) — regresión GAM adaptada a los estudios de zona de mercado. Añade funciones para analizar grupos de propiedades dentro de un mercado definido, con tratamiento opcional de la fila del sujeto.

En los tres modos el motor de modelado subyacente es idéntico — siempre se ajusta un GAM mediante `mgcv::gam()`. El ajuste del propósito controla qué herramientas y elementos de interfaz adicionales están disponibles.

1.4 Funciones específicas del sector inmobiliario

Cuando se selecciona **For Appraisal** o **Market Area Analysis**, mgcvUI activa varias funciones diseñadas para el análisis inmobiliario:

- **Designaciones especiales de columnas** — cada predictor puede etiquetarse con un papel especial como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` o `display_only`. Estas designaciones controlan cómo se trata la columna durante el ajuste y en las salidas.
- **Column Sale Age** — cuando una columna se designa como `contract_date` y se proporciona una fecha efectiva, mgcvUI calcula una columna `sale_age` (días entre la fecha de venta y la fecha efectiva) y la sustituye como predictor.
- **Cálculos RCA (solo en tasación)** — en el modo de tasación, después de ajustar el modelo mgcvUI puede calcular la salida del Residual Constraint Approach (RCA), que produce ajustes por comparable, resúmenes de ajuste neto/bruto y un precio de venta ajustado para la propiedad sujeto.

***Consejo:** El modo General funciona con cualquier conjunto de datos — financiero, científico, de ingeniería o inmobiliario. Los modos de tasación y de mercado simplemente añaden funciones de conveniencia para los profesionales del sector inmobiliario.*

1.5 Primeros pasos

Para usar mgcvUI:

1. **Instale earthUI** en R: `install.packages("earthUI")` — la rutina de mgcv está incluida. (Véase “Instalación de earthUI y vProlog” en el artículo sobre earthUI de este número para la versión de desarrollo y los componentes opcionales).
2. **Inicie la aplicación:** ejecute `earthUI::launch_mgcv()` en R. La aplicación se abre en su navegador web en el puerto 7880.
3. **Cree o abra un proyecto** en la Sección 1 de la barra lateral. Un proyecto fija el propósito (General, For Appraisal o Market Area Analysis) y la ubicación de trabajo (país, estado, condado, ciudad), así como sus carpetas de entrada y salida. Los proyectos se almacenan

bajo una raíz **regProj** compartida y se comparten con las aplicaciones hermanas earthUI y glmnetUI. Abrir un proyecto con un propósito distinto es la forma de cambiar de modo.

4. **Importe sus datos** en la Sección 2 seleccionando un archivo CSV o Excel de la carpeta de entrada del proyecto activo. Añada archivos a esa carpeta y haga clic en **Refresh** (actualizar) para listarlos.
5. **Importe desde earthUI (opcional)** — la Sección 3 permite importar un archivo **.rds** de resultados de earthUI para sembrar el GAM con los nudos descubiertos por earth. Omita este paso si no dispone de un resultado de earthUI.
6. **Configure las variables** — elija su variable objetivo y sus predictores, fije los tipos de datos, asigne los papeles especiales de columna y marque las variables como factor o lineal.
7. **Fije los parámetros de mgcv** — family, method, gamma, tipo de base, k, tipo de tensor, interacciones y otras opciones.
8. **Ajuste el modelo** — haga clic en “Fit Mgcvm GAM Model” (ajustar el modelo GAM de mgcv) y revise los resultados en el panel principal.
9. **Exporte** — descargue las predicciones en Excel, genere y convierta un informe Quarto o (en modo de tasación) calcule los ajustes RCA y una cuadrícula de comparación de ventas.

Los ajustes se guardan automáticamente en una base de datos SQLite y se restauran cuando vuelve a cargar el mismo archivo de entrada.

2 Requisitos de los datos de entrada del MLS

Para los flujos de trabajo de tasación inmobiliaria y análisis de mercado, sus datos de entrada provienen normalmente de una exportación de un Multiple Listing Service (MLS, servicio de listado múltiple). Este capítulo describe la estructura de archivo esperada y las columnas que mgcvUI puede utilizar.

2.1 Formato y estructura del archivo

mgcvUI acepta archivos **CSV** y **Excel** (.xlsx, .xls). Al importarlos, los nombres de las columnas se convierten automáticamente a **snake_case** — por ejemplo, “Living SqFt” se convierte en **living_sqft**, “Contract Date” en **contract_date** y “Sale Price” en **sale_price**. Esta normalización garantiza referencias de columna coherentes a lo largo de todo el flujo de trabajo. El separador de CSV y la marca decimal utilizados durante la importación vienen determinados por la configuración regional (véase el capítulo 3, “Configuración regional y de idioma”).

Su archivo de datos debe ser una tabla plana con una fila por propiedad y una columna por atributo. La primera fila del archivo debe contener los encabezados de columna.

2.2 Columnas requeridas en el modo de tasación

Aunque mgcvUI funciona con cualquier conjunto de columnas, el flujo de trabajo completo de tasación se beneficia de disponer de las siguientes columnas:

Columna	Tipo especial	Finalidad
Sale Price	(objetivo)	Variable de respuesta del modelo
Contract Date	contract_date	Se usa para calcular sale_age (días desde la fecha efectiva)
Listing Date	listing_date	Se usa junto con la fecha de contrato para calcular el DOM si no existe una columna DOM
Days on Market	dom	Días en el mercado; se muestra en las exportaciones
Concessions	concessions	Concesiones de venta; Net SP = Sale Price – Concessions
Living Area (SF)	living_area	Habilita los residuos por pie cuadrado (residual_sf , cqa_sf)
Lot Size	lot_size	Columna de superficie del terreno
Site Dimensions	site_dimensions	Agrupada con el tamaño del terreno
Latitude	latitude	Redondeada a 3 decimales
Longitude	longitude	Redondeada a 3 decimales
Area ID	area	Identificador de zona de mercado o vecindario
Actual Age	actual_age	Edad de la propiedad
Effective Age	effective_age	Edad efectiva de la propiedad
Address	display_only	Se muestra en las exportaciones; se excluye del modelo

Los nombres de las columnas de la hoja de cálculo pueden estar en otro idioma — los nombres “especiales” están en inglés para que el programa de R pueda darles un tratamiento especial. Por lo demás, los nombres de columna proporcionados aparecen en los modelos de regresión, en los gráficos y, si se realizan tasaciones, en los informes de salida.

No todas las columnas son obligatorias. mgcvUI se adapta — si falta una columna, la función correspondiente simplemente se omite. Sin embargo, para los modelos de precios inmobiliarios se recomiendan especialmente ciertas columnas:

1. **Sale Age** — el número de días transcurridos entre la fecha de contrato de venta y la fecha efectiva. Si se utiliza un historial de ventas de varios años, **sale_age** suele desempeñar un papel central.
2. **Living Area** — también recibe nombres como “Living Sqft” o “GLA” (gross living area, superficie habitable bruta).
3. **Total Bath Count** — el número total de baños completos, de un cuarto, de medio baño y de tres cuartos.
4. **Garage Bays** o **Garage Area** — plazas de garaje o superficie en pies cuadrados.
5. **Lot Size** — la superficie de terreno de la propiedad.
6. **Longitude**, **Latitude** y, si está disponible, **Area ID** para la variación geográfica de precios.

2.3 Convenciones de nombres de columnas especiales

mgcvUI identifica las columnas por su **designación de tipo especial**, no por su nombre de columna. Puede dar a sus columnas el nombre que desee en la exportación del MLS — lo que importa es que asigne el tipo especial correcto en la tabla de configuración de variables (capítulo 6).

2.4 Calidad e integridad de los datos

- **Valores faltantes (NA)**: las filas con valores NA en cualquier columna de predictor o de objetivo se eliminan automáticamente antes del ajuste. mgcvUI reduce los datos únicamente a las columnas utilizadas en el modelo, de modo que los NA en columnas no utilizadas no provocan la pérdida de filas.
- **Columnas numéricas**: el precio de venta, la superficie habitable, el tamaño del terreno y campos similares deben ser numéricos.
- **Columnas de factor**: las columnas de texto, TRUE/FALSE y de tipo factor de R se tratan automáticamente como categóricas — solo pueden entrar en el modelo como factores. Una columna **numérica** (por ejemplo, un código numérico de zona o de estilo) se convierte en factor únicamente cuando se marca su casilla **Factor**; mgcvUI nunca infiere el carácter categórico a partir del rango de valores de una columna numérica, de modo que predictores numéricos discretos como **bath_count** siguen siendo continuos (suavizados). El papel especial es independiente de la decisión sobre el factor.

***Consejo:** Revise la columna NA de la tabla de predictores después de la importación. Las columnas con muchos valores faltantes pueden provocar la eliminación de filas. Las columnas con $\geq 50\%$ de NA se marcan en rojo y se excluyen automáticamente.*

2.5 Ubicación de la fila del sujeto

En el modo **Appraisal**, la fila 1 debe ser la **propiedad sujeto**. Todas las filas restantes son ventas comparables. La fila del sujeto se excluye del ajuste del modelo. Después del ajuste, el modelo genera de todos modos predicciones para la fila del sujeto.

En el modo **Market Area Analysis**, colocar el sujeto en la fila 1 es opcional.

En el modo **General** no hay tratamiento especial de filas — todas las filas se tratan por igual.

3 Modo de uso general

3.1 Panorama general

El modo de uso general es el predeterminado al iniciar mgcvUI. Ofrece el flujo de trabajo completo de GAM para cualquier conjunto de datos — no solo inmobiliarios. Puede usar mgcvUI para datos científicos, análisis financiero, estudios de ingeniería o cualquier problema de regresión en el que quepa esperar relaciones no lineales suaves.

En el modo General, la interfaz omite las funciones específicas del sector inmobiliario (columnas especiales, antigüedad de la venta, RCA). La barra lateral se simplifica para centrarse en la selección de variables, la configuración de parámetros, el ajuste del modelo y la exportación.

3.2 El flujo de trabajo de la barra lateral

La barra lateral se organiza en secciones numeradas y plegables que le guían desde la selección del proyecto hasta la exportación del informe. La mayoría de las secciones aparecen solo una vez abierto un proyecto activo, y varias son específicas del propósito del proyecto. La carpeta de salida se deriva del proyecto activo — no hay un campo aparte para la carpeta de salida.

1. Project — crear o abrir un proyecto. Un proyecto fija el propósito y la ubicación de trabajo (país, estado, condado, ciudad) y sus carpetas de entrada (`<os>_in/`) y de salida. Los proyectos se almacenan bajo una raíz **regProj** compartida (por omisión `~/regProj`, modificable con la variable de entorno `REGPROJ_ROOT` o desde el desplegable Settings) y se comparten con las aplicaciones hermanas **earthUI** y **glmnetUI**, de modo que los modelos de las tres conviven lado a lado. Haga clic en **Close Project** (cerrar proyecto) para cambiar.

2. Import Data — seleccionar un archivo CSV o Excel de la carpeta de entrada del proyecto activo. Añada archivos a la carpeta mediante el Finder o el Explorador, y haga clic en **Refresh** para listarlos. Para los archivos de Excel con varias hojas aparece un selector de hoja. Los nombres de columna se convierten automáticamente a `snake_case`.

3. Import from earthUI (opcional) — importar un archivo `.rds` de resultados de **earthUI** para sembrar el GAM con las posiciones de nudo descubiertas por **earth** (y para sembrar las selecciones de variables). Este paso es opcional — omítalo si no dispone de un resultado de **earthUI** que importar.

4. Variable Configuration — selector de la variable objetivo, transformación de la respuesta (none, log, log10) y tabla de predictores con casillas para Include, Factor y Linear. La columna Special aparece solo en los modos Appraisal y Market. Véase el capítulo 6 para todos los detalles.

5. MgcV Call Parameters — toda la configuración del modelo: preajustes de parámetros, family, method, gamma, validación cruzada, select, tipo de base, k, tipo de tensor, matriz de interacciones y parámetros avanzados. Véase el capítulo 7 para la referencia completa de parámetros.

6. Fit MgcV GAM Model — el botón que ejecuta el modelo. Los resultados aparecen en las pestañas de la derecha.

7. Download Output — exporta las predicciones, los residuos, las puntuaciones CQA y las contribuciones por variable como archivo de Excel a la carpeta de salida del proyecto. Disponible en todos los modos de uso una vez ajustado el modelo.

8. Calculate RCA Adjustments & Download — calcula los ajustes RCA por variable para cada comparable e interpola el residuo del sujeto mediante CQA. Aparece solo en los modos Appraisal y Market, después del ajuste.

9. Generate Sales Grid & Download — genera la Sales Comparison Grid (cuadrícula de comparación de ventas), recomendando comparables según el porcentaje de ajuste bruto. Aparece solo en el modo Appraisal.

10. Generate Quarto Report — escribe un paquete de informe Quarto (`.qmd`) autocontenido — con todos los gráficos y tablas — en la carpeta de salida del proyecto. Disponible una vez ajustado el modelo.

11. Convert Quarto Report — renderiza el paquete `.qmd` generado a HTML, Word o PDF. El PDF requiere una instalación de LaTeX; si no se detecta ninguna, la opción de PDF queda oculta. Véase el capítulo 13.

Los pasos específicos de cada propósito se omiten (y los pasos restantes se renumeran) en los demás modos: en el modo Market falta el paso de la Sales Grid, y en el modo General faltan tanto el paso de RCA como el de la Sales Grid.

3.3 Importar desde earthUI (opcional)

La Sección 3 de la barra lateral — **Import from earthUI** — permite importar un archivo `.rds` de resultados de earthUI. Esto habilita la cadena `earth-mgcv`: las posiciones de nudo de earth, guiadas por los datos, se convierten en puntos de anclaje para los términos suaves del GAM.

Cuando se importa un resultado de earthUI, `mgcvUI`:

- Muestra un resumen con la variable objetivo, el R cuadrado, el número de términos y las posiciones de nudo por variable.
- Selecciona automáticamente el preajuste de parámetros “Earth Pipeline” (base de spline de regresión cúbica, $\gamma = 1.4$).
- Fija la variable objetivo para que coincida con la de earth.
- Marca previamente Include para las variables predictoras de earth.
- Marca previamente Factor para las variables categóricas de earth y Linear para los predictores lineales de earth.
- Rellena previamente las interacciones detectadas por earth (resaltadas en amarillo en la matriz de interacciones y bloqueadas frente a cambios manuales).
- Ofrece un botón **Export Knots CSV** para descargar los nudos de earth como archivo CSV.

La tabla de configuración de variables muestra entonces una nota que explica que las selecciones de variables (Include, Linear, Factor) y los nudos de las splines se *sembraron* a partir del modelo de earth. A diferencia de `glmnetUI`, donde el ajuste de `glmnet` queda ligado a la base de earth, `mgcvUI` ajusta el GAM a partir de sus selecciones actuales — earth solo proporciona una ventaja inicial, y todos los controles siguen siendo editables.

Un botón **Clear** elimina los datos importados de earth y restablece el modo autónomo.

3.4 Pestañas del panel principal

Tras la importación de los datos, el panel principal ofrece las siguientes pestañas (las que dependen del modelo se llenan después del ajuste):

Pestaña	Contenido
Data	DataTable interactiva con los datos importados. En los modos de tasación y de mercado se divide en tablas de propiedad sujeto y de ventas comparables.
Equation	Fórmula completa del GAM representada con MathJax, más una tabla con las definiciones de las funciones suaves.
Summary	Métricas clave (R^2 , R^2 de CV, desviación explicada, AIC, n, número de términos suaves) y tablas de términos suaves y paramétricos.
Variable Importance	Gráfico de barras horizontales de los estadísticos F (términos suaves) y de los valores $ t $ (paramétricos), más una tabla ordenada.
Contribution	Curvas suaves de efecto parcial interactivas en plotly, con bandas de confianza y marcadores de los nudos de earth.
Correlation	Mapa de calor de las correlaciones entre predictores numéricos (disponible antes del ajuste).
Diagnostics	Residuos frente a ajustados, gráfico Q-Q, histograma, respuesta frente a ajustados (mediante gratia). Diagrama de dispersión de valores reales frente a predichos.
RCA Adjustments	Histogramas del análisis RCA (modo de tasación o de mercado, tras el cálculo).
Anova	Tabla ANOVA combinada de los términos paramétricos y suaves.
Mgcv Output	Texto sin formato: tiempos, fórmula, resumen del modelo, familia, concurvidad.
Sign Check	Comprobación de la coherencia de dirección entre earth y el GAM (cuando se ha importado earth).
Concurvity	Matrices de concurvidad global y por pares con codificación por colores.
Basis Check	Pruebas de <code>gam.check()</code> sobre la adecuación de la dimensión de la base.

3.5 Persistencia de los ajustes

mgcvUI guarda automáticamente su configuración en una base de datos SQLite, indexada por el nombre del archivo de entrada. Cuando vuelve a cargar el mismo archivo se restauran todos los ajustes: selección del objetivo, casillas de los predictores, tipos de datos, parámetros de mgcv y matriz de interacciones. El propósito es una propiedad del proyecto activo y no un ajuste global; el último archivo de entrada utilizado se recuerda por proyecto y se carga automáticamente cuando el proyecto se vuelve a abrir.

3.6 Modo oscuro

Haga clic en el icono de luna/sol de la esquina superior derecha para alternar entre los temas Nord Light y Nord Dark. La preferencia de tema se guarda en el almacenamiento local y persiste entre sesiones. Todos los elementos de la interfaz se adaptan al tema seleccionado.

3.7 Configuración regional y de idioma

mgcvUI admite convenciones internacionales de formato numérico y de CSV mediante un sistema de configuración regional basado en el país. El desplegable **Settings** de la barra de título ofrece selectores **Country** y **Paper** para más de 30 países admitidos. Cada preajuste configura:

- **Separador de CSV** — coma (,) para EE. UU., Reino Unido y Japón, o punto y coma (;) para gran parte de Europa.
- **Signo decimal** — punto (.) o coma (,).
- **Separador de millares** — coma (EE. UU./Reino Unido/Japón), punto (Alemania/Italia/España), espacio (Finlandia/Francia/Polonia/Bálticos) o apóstrofo (Suiza).

- **Tamaño de papel** — Letter (EE. UU./Canadá/México) o A4 (en el resto de los casos).

Haga clic en **Save** para almacenar globalmente sus preferencias regionales. La sección Import Data dispone además de un selector **Locale** por archivo para importaciones puntuales que usen convenciones distintas. El mismo desplegable Settings contiene el campo de la **carpeta raíz regProj**, que reubica el árbol de proyectos compartido que usan mgcvUI, earthUI y glmnetUI.

***Consejo:** El formato numérico de los ejes de los gráficos y de las etiquetas de pendiente se adapta automáticamente a su configuración regional. La configuración alemana usa puntos para los millares (200.000), la finlandesa usa espacios (200 000) y la suiza usa apóstrofes (200'000). No se muestran símbolos de moneda — mgcvUI es agnóstico respecto de la moneda.*

4 Modo de tasación

Cuando el propósito del proyecto activo es **For Appraisal**, mgcvUI se configura para la valuación de una sola propiedad. Todas las funciones descritas en el capítulo 3 siguen disponibles; este capítulo trata únicamente los añadidos específicos de la tasación.

4.1 Tratamiento de la fila del sujeto

En el modo de tasación, **la fila 1 de su conjunto de datos es la propiedad sujeto** y todas las filas restantes son ventas comparables. Su archivo de entrada debe organizarse en consecuencia (véase el capítulo 2). El precio de venta del sujeto puede dejarse en blanco o fijarse en cualquier valor — mgcvUI lo trata automáticamente como NA durante el ajuste.

Tras el ajuste, el modelo genera predicciones para la fila del sujeto, que constituyen la base del flujo de trabajo de ajustes RCA.

4.2 Fecha efectiva y antigüedad de la venta

En los modos de tasación y de mercado aparece un campo **Effective Date** en la sección de configuración de variables (con la fecha de hoy por omisión). Si designa una columna como **contract_date** en el desplegable **Special**, mgcvUI calcula una columna **sale_age** — el número de días enteros entre la fecha de contrato de cada venta y la fecha efectiva. Esta columna se añade como predictor.

Cuando cambia la fecha efectiva, **sale_age** se recalcula automáticamente.

4.3 Designaciones especiales de columnas

En los modos de tasación y de mercado aparece un desplegable **Special** para cada predictor de la tabla de configuración de variables. Véase el capítulo 6 para la referencia completa de los tipos especiales y sus efectos.

4.4 Panorama de los ajustes RCA

El botón **Calculate RCA Adjustments & Download** (visible en los modos de tasación y de mercado tras el ajuste) calcula ajustes derivados del mercado para cada comparable respecto del sujeto. El flujo de trabajo RCA completo se describe en el capítulo 11.

***Consejo:** La puntuación CQA que asigne al sujeto controla qué parte de la distribución de residuos se le atribuye. Una puntuación de 5.00 sitúa al sujeto en la mediana de los comparables.*

5 Modo de análisis de zona de mercado

Cuando el propósito del proyecto activo es **Market Area Analysis**, mgcvUI ofrece las mismas funciones específicas del sector inmobiliario que el modo de tasación (columnas especiales, antigüedad de la venta), pero está orientado a analizar un grupo de propiedades en lugar de valorar un único sujeto.

5.1 Diferencias respecto del modo de tasación

- **El tratamiento de la fila del sujeto es flexible** — en el modo de mercado, la fila 1 no se trata automáticamente como propiedad sujeto.
- **Sin Sales Grid** — el paso “Generate Sales Grid & Download” no está disponible.

5.2 Cuándo usar el modo de mercado

El modo de análisis de zona de mercado es apropiado cuando usted:

- Construye un modelo GAM para un vecindario o zona de mercado con el fin de comprender los factores no lineales que determinan el valor
- Analiza cómo variables como la superficie, la antigüedad, el tamaño del terreno y la ubicación afectan a los precios de venta en un grupo de propiedades
- Prepara el respaldo para un análisis de condiciones de mercado o una delimitación de vecindario
- Trabaja con un conjunto de datos en el que desea las funciones de columnas especiales (antigüedad de la venta) pero no necesita el flujo de trabajo de la cuadrícula de ventas

***Consejo:** El modo de mercado también resulta útil para la regresión inmobiliaria general cuando se desean las funciones de columnas especiales pero no se necesita el flujo de trabajo de valuación de una sola propiedad.*

6 Selección de variables

La Sección 4 de la barra lateral — **Variable Configuration** — es donde se elige qué columnas participan en el modelo y cómo se tratan.

6.1 Variable objetivo

El desplegable **Target (response) variable** situado en la parte superior de la Sección 4 lista todas las columnas numéricas de su conjunto de datos. Seleccione una columna como variable de respuesta (por ejemplo, el precio de venta). La columna objetivo se excluye automáticamente de la lista de predictores.

6.2 Transformación de la respuesta

Debajo del selector de objetivo, un desplegable **Response Transform** ofrece tres opciones:

- **None (raw)** — usar la respuesta tal cual. Apropiado cuando la relación entre los predictores y la respuesta es aproximadamente aditiva en la escala original.
- **Log (natural)** — aplicar la transformación logarítmica natural antes del ajuste. Al deshacer la transformación, esto hace que los ajustes sean proporcionales (%) en lugar de absolutos (\$). Resulta especialmente útil para los precios de venta, donde un ajuste de \$10,000 significa cosas muy distintas a distintos niveles de precio.
- **Log10** — aplicar la transformación logarítmica en base 10.

***Consejo:** La transformación logarítmica hace que los ajustes temporales sean proporcionales (%) en lugar de absolutos (\$). En los modelos inmobiliarios que abarcan un amplio rango de precios, las transformaciones logarítmicas suelen producir residuos mejor comportados.*

Cuando se selecciona una transformación logarítmica, los valores ≤ 0 de la respuesta se filtran automáticamente. Todas las salidas (predicciones, contribuciones, residuos) se retransforman automáticamente a la escala original.

6.3 La tabla de predictores

Debajo del selector de objetivo, una tabla desplazable lista todas las columnas restantes. Las columnas de casillas usan encabezados verticales rotados para mayor compacidad. El orden de las columnas es:

Columna	Descripción
Variable	Nombre de la columna (fuente monoespaciada, truncado con puntos suspensivos). Celda con borde.
Type	Desplegable de tipo de dato: numeric, integer, character, factor, logical, Date, POSIXct. Desplegable con borde.
Include	Casilla — incluir esta columna como predictor en el modelo. Encabezado vertical.
Factor	Casilla — tratar como factor categórico (crea un coeficiente por nivel). Encabezado vertical.
Linear	Casilla — forzar una relación lineal (sin término suave, solo βx). Encabezado vertical.
Special	Desplegable (solo en los modos de tasación y de mercado) — véase más abajo la referencia de tipos especiales de columna. Desplegable con borde.
NAs	Recuento y porcentaje de valores faltantes, con código de color: rojo ($\geq 50\%$), naranja ($\geq 20\%$), gris (>0). Celda con borde.

Factor frente a suave frente a lineal: por omisión, las variables numéricas incluidas reciben un término suave $f(x)$. Marcar **Factor** crea un término categórico (un coeficiente por nivel). Marcar **Linear** crea un término lineal simple βx en lugar de un término suave. Las variables marcadas a la vez como Factor y Linear se tratan como Factor.

6.4 Detección de tipos de datos y anulaciones

mgcvUI detecta automáticamente los tipos de datos en la importación. Se reconocen las columnas numéricas, enteras, lógicas, de factor y de fecha. Puede anular cualquier detección cambiando el desplegable **Type**. Cambiar los tipos afecta a cómo se trata la columna en el modelo.

6.5 Referencia de tipos especiales de columna

En los modos de tasación y de mercado, el desplegable **Special** ofrece las siguientes opciones:

Ponderación:

- **weight** — columna de peso de las observaciones (solo se permite una; las filas con peso = 0 se excluyen del ajuste)

Tipos de fecha y hora:

- **contract_date** — activa el cálculo automático de **sale_age** a partir de la fecha efectiva
- **sale_age** — identifica una columna de antigüedad de venta ya calculada (días desde la fecha efectiva), que se usa tal cual
- **listing_date** — se usa como alternativa para calcular los días en el mercado
- **dom** — identifica la columna de días en el mercado

Tipo de registro:

- **sale_type** — identifica la columna de tipo de registro que distingue el sujeto de las ventas comparables

Tipos monetarios:

- **concessions** — identifica las concesiones de venta (créditos del vendedor, incentivos al comprador, etc.)

Tipos de tamaño y ubicación:

- **latitude** — los valores se redondean automáticamente a 3 decimales
- **longitude** — mismo tratamiento de redondeo que la latitud
- **area** — identificador de zona de mercado o vecindario
- **living_area** — habilita los cálculos de residuos por pie cuadrado (**residual_sf** y **cqa_sf**)
- **lot_size** — columna de superficie del terreno
- **site_dimensions** — agrupada con el tamaño del terreno

Tipos de edad:

- **actual_age** — columna de edad de la propiedad
- **effective_age** — edad efectiva de la propiedad

Tipos de visualización:

- **display_only** — la columna se incluye en las exportaciones a Excel pero se excluye por completo del ajuste del modelo. Úselo para campos de dirección, números de MLS, identificadores de parcela u otros datos de referencia.

***Consejo:** Las columnas designadas como **display_only** permanecen en su conjunto de datos y aparecen en la exportación a Excel, pero quedan excluidas por completo de la tabla de predictores. Úselo para columnas de identificador, direcciones u otros campos de referencia.*

7 Selección de parámetros

La Sección 5 de la barra lateral — **Mgcv Call Parameters** — da acceso a todas las opciones de configuración del GAM. Cada parámetro tiene un icono azul de ayuda (?) con una explicación emergente.

7.1 Preajustes de parámetros

Un desplegable situado en la parte superior ofrece dos preajustes que configuran valores predeterminados razonables para los flujos de trabajo habituales:

Preajuste	Base	Ajustes
Standalone (descubrimiento)	tp (thin plate)	k = auto (10), gamma = 1.2, select = TRUE
Earth Pipeline (refinamiento)	cr (regresión cúbica)	k = auto (10), gamma = 1.4, select = TRUE

El preajuste **Earth Pipeline** se selecciona automáticamente cuando se importan nudos de earthUI. La base de spline de regresión cúbica (cr) es necesaria para la integración de los nudos de earth porque permite especificar posiciones de nudo exactas.

7.2 Family

Elija la familia de distribución para su variable de respuesta:

Familia	Caso de uso
gaussian	Respuestas continuas (por ejemplo, el precio de venta). La más común.
Gamma	Respuestas continuas positivas con varianza proporcional a la media.
poisson	Datos de recuento (por ejemplo, el número de ventas).
binomial	Resultados binarios.
inverse.gaussian	Datos positivos con fuerte asimetría a la derecha.

7.3 Method

El método de estimación del parámetro de suavizado:

Método	Descripción
REML	Máxima verosimilitud restringida (predeterminado, recomendado). El más robusto frente al sobreajuste.
GCV.Cp	Validación cruzada generalizada. Tiende a seleccionar modelos algo más complejos.
ML	Máxima verosimilitud. Similar a REML, pero puede suavizar de menos.
P-REML	Variante REML de Pearson.
P-ML	Variante ML de Pearson.
GACV.Cp	Validación cruzada aproximada generalizada.

7.4 Gamma (multiplicador de suavizado)

El parámetro gamma multiplica los grados de libertad efectivos en el criterio de suavizado, lo que favorece ajustes más suaves (menos ondulados). El valor predeterminado es 1.2.

Valor	Efecto
1.0	Suavizado estándar
1.2–1.4	Algo más suave que el predeterminado (recomendado)
2.0+	Mucho más suave — adecuado para datos ruidosos o muestras pequeñas

Los valores más altos de gamma protegen frente al sobreajuste al penalizar la complejidad con mayor severidad. El preajuste Earth Pipeline usa 1.4 para suavizar más al refinar los nudos de earth.

7.5 Validación cruzada

Cuando está marcada (por omisión), mgcvUI ejecuta una validación cruzada de 10 pliegues tras el ajuste para calcular un R cuadrado de validación cruzada. Esto proporciona una estimación honesta de la capacidad predictiva fuera de la muestra.

7.6 Select (contracción)

Cuando está marcada (por omisión), añade una penalización adicional a cada término suave que puede contraerlo hasta cero. Esto permite la selección automática de variables — los términos suaves poco importantes quedan efectivamente eliminados del modelo. Se implementa mediante `select = TRUE` en `gam()`.

7.7 Base predeterminada

El tipo de función base de spline para los términos suaves:

Base	Nombre completo	Descripción
tp	Spline de regresión thin plate	Predeterminada. Isotrópica (no hace falta colocar nudos). Buena elección de propósito general.
cr	Spline de regresión cúbica	Permite colocar nudos de forma explícita. Necesaria para integrar los nudos de earth.
ps	P-spline	Base de B-splines con nudos equiespaciados y penalización de diferencias. Computacionalmente eficiente.
bs	B-spline	Base de B-splines flexible.

7.8 k predeterminada (dimensión de la base)

La dimensión de la base k controla la ondulación máxima de cada término suave. Un valor de 0 (el predeterminado) significa “automático” — mgcvUI usa $k = 10$ o el número de nudos de earth, según lo que corresponda.

- **k pequeña (3–5):** curvas muy suaves, casi lineales.
- **k media (8–15):** flexibilidad moderada (suele ser suficiente).

- **k grande (20+):** muy flexible — riesgo de sobreajuste sin una regularización fuerte.

Consejo: Los grados de libertad efectivos (*EDF*) de la pestaña *Summary* indican cuántos grados de libertad usa realmente cada término suave. Si el *EDF* se acerca a $k - 1$, la base puede ser demasiado pequeña — aumente k . Si el *EDF* es mucho menor que k , el término suave está bien penalizado.

7.9 Tipo de tensor

Para las interacciones entre variables continuas hay disponibles dos tipos de producto tensorial:

Tipo	Función	Descripción
ti	ti(x1, x2)	Interacción tensorial — modela solo el efecto de interacción, y los efectos principales se tratan por separado mediante términos s() univariantes. Preferible por su interpretabilidad.
te	te(x1, x2)	Producto tensorial — modela todo el efecto conjunto, incluidos los efectos principales. Más difícil de descomponer para los ajustes RCA.

7.10 Interacciones permitidas

Una sección plegable **Allowed Interactions** muestra una matriz triangular superior de casillas para todos los predictores suaves incluidos (no lineales y no factores). Cada casilla habilita un término suave de producto tensorial entre el par de variables correspondiente.

- **Allow All / Clear All** — botones de activación masiva.
- **Haga clic en el nombre de una variable** — alterna todas las interacciones de esa variable.
- **Haga clic con el botón derecho en el nombre de una variable** — “Block from main effect” — marca la variable para uso exclusivo en interacciones (sin término suave univariante).

Cuando se importan nudos de earthUI, las interacciones detectadas por earth quedan premarcadas y bloqueadas (resaltadas con fondo amarillo). Si el modelo de earth usó $\text{degree} = 1$ (sin interacciones), se muestra un mensaje informativo.

Advertencia

Habilitar interacciones tensoriales aumenta considerablemente la complejidad del modelo. Use interacciones cuando disponga de conocimiento del dominio o de evidencia detectada por earth que las respalde. Verifique que los ajustes resultantes sean razonables y explicables.

7.11 Interacciones factor por término suave

Aparece una matriz aparte de **Factor-by-Smooth Interactions** cuando se incluyen a la vez variables de factor y variables suaves. Cada casilla crea un término **s(x, by = factor)** — una curva suave distinta para cada nivel del factor.

7.12 Parámetros avanzados

Una sección plegable “Advanced” expone ajustes adicionales:

- **Optimizer** — algoritmo de estimación del parámetro de suavizado: `outer/newton` (predeterminado), `outer/bfgs` o `efs` (Fellner-Schall extendido).
- **Scale** — parámetro de escala. 0 (predeterminado) = estimar a partir de los datos. Fíjelo en un valor conocido para modelos de escala fija.
- **Discrete covariate method** — ajuste discretizado rápido para conjuntos de datos grandes (`discrete = TRUE`).
- **Threads** — número de hilos paralelos (1–32). Solo tiene efecto cuando `discrete = TRUE`.

8 Ajuste del modelo

8.1 El botón de ajuste

La Sección 6 de la barra lateral contiene el botón **Fit Mgcv GAM Model**. Al hacer clic en él se ejecuta el modelo con su configuración actual.

8.2 Qué ocurre durante el ajuste

Cuando hace clic en Fit, mgcvUI:

1. **Valida la configuración** — comprueba que haya al menos un predictor y al menos 10 filas completas. Advierte si menos del 50 % de las filas están completas (indicando las columnas responsables de la mayoría de los NA).
2. **Prepara los datos** — reduce los datos únicamente a las columnas relevantes para el modelo, elimina las filas con NA, aplica los pesos (si se han designado), excluye la fila del sujeto (en el modo de tasación) y aplica las transformaciones de la respuesta.
3. **Construye la fórmula del GAM** — construye términos suaves $s(x, bs="cr", k=10)$ para los predictores numéricos, términos de factor para los categóricos, términos lineales para las variables marcadas como Linear y productos tensoriales para las interacciones seleccionadas. Los nudos de earth se incorporan como posiciones de nudo explícitas en los términos con base **cr**.
4. **Trata los casos límite** — descarta las variables constantes, convierte los predictores binarios en factores, limita k al número de valores únicos menos 1 y completa los nudos de earth con cuantiles de los datos cuando hacen falta más nudos.
5. **Ajusta el GAM** — llama a `mgcv::gam()` con la fórmula construida y los parámetros family, method, gamma, select y demás.
6. **Ejecuta la validación cruzada** (si está habilitada) — CV de 10 pliegues que calcula el R cuadrado fuera de la muestra.
7. **Actualiza todas las pestañas de resultados** — se llenan Summary, Equation, los gráficos de contribución, Diagnostics y todas las demás pestañas.

8.3 Ajuste en segundo plano y progreso

Cuando el paquete **callr** está disponible, el ajuste del modelo se ejecuta en un proceso en segundo plano para que la aplicación siga respondiendo. Una superposición de progreso de estilo terminal y tema oscuro muestra:

- Un **cronómetro** con los segundos y minutos transcurridos.
- Un **registro de traza con código de color**: azul para el estado, amarillo para el progreso de los pliegues de CV, rojo para los errores y verde para los resultados.
- Un botón **Abort** para detener el proceso de ajuste si es necesario.
- Un botón **Close** que aparece al finalizar.

Si **callr** no está disponible, el ajuste se ejecuta de forma síncrona (la aplicación se congela hasta que termina).

Tras un ajuste correcto aparece una marca de verificación blanca en el botón Fit y una línea de estado muestra: “R-sq = X, CV R-sq = X, Dev = X %, AIC = X, n = X”.

9 Pestañas de resultados

9.1 Data

Muestra los datos importados en una DataTable interactiva con desplazamiento horizontal y 15 filas por página. En los modos de tasación y de mercado, la pestaña se divide en una tabla de **Subject Property** (fila 1) y una tabla de **Comparable Sales** (las filas restantes); en el modo General se muestra una sola tabla.

Cada celda se mantiene en una única línea y se trunca al ancho de la columna, de modo que los campos de texto libre extensos (como las observaciones sobre la propiedad) no estiren una fila a lo largo de la página. Un clic simple selecciona una fila; **haga doble clic en cualquier celda** para abrir una ventana emergente con su contenido completo sin truncar.

9.2 Equation

Muestra el modelo GAM ajustado en dos partes:

Ecuación del modelo (representada con MathJax):

- El lado izquierdo muestra la respuesta con la transformación correspondiente (log, log10) y la función de enlace.
- El lado derecho muestra el valor del intercepto, los términos suaves como $f_i(\text{variable})$, los términos de factor agrupados por variable con el recuento de niveles y los términos lineales con los valores de sus coeficientes.
- La familia y el método se muestran encima de la ecuación.

Tabla de **Smooth Function Definitions** (definiciones de las funciones suaves):

- Cadena con la fórmula del término (por ejemplo, `s(living_sqft, bs="cr", k=8)`)
- Nombre de la variable
- Tipo (s, te, ti o lineal)
- Tipo de función base
- Valor de k
- Posiciones de los nudos de earth (si se han importado)

9.3 Summary

Seis tarjetas de métricas mostradas en la parte superior:

- **R-squared** — proporción de desviación explicada
- **CV R-squared** — R cuadrado de validación cruzada (si la CV está habilitada)
- **Dev. Explained %** — porcentaje de desviación explicada
- **AIC** — criterio de información de Akaike
- **n** — número de observaciones utilizadas
- **Smooths** — número de términos suaves

Debajo de las tarjetas:

- **Tabla Smooth Terms** — término, EDF (grados de libertad efectivos), Ref.df, estadístico F, valor p
- **Tabla Parametric Terms** — término, estimación, error estándar, valor t, valor p

9.4 Variable Importance

Un gráfico de barras horizontales que muestra la importancia relativa de cada término del modelo:

- **Términos suaves:** ordenados por el estadístico F (barras azules)

- **Términos paramétricos:** ordenados por el valor $|t|$ (barras verdes)

Debajo del gráfico, una DataTable muestra el término, el tipo, el EDF, el estadístico y el valor p, ordenados por estadístico descendente.

9.5 Gráficos de contribución

Todo término del modelo que contribuye al valor predicho tiene un gráfico de contribución interactivo en **plotly**, mostrado en una tarjeta con borde. La pestaña muestra gráficos para todos los tipos de término:

9.5.1. Términos suaves univariantes — $s(x)$

- **Línea azul** con banda de confianza del 95 %
- **Puntos grises** que muestran los residuos parciales (en los modelos sin transformación logarítmica)
- **Líneas verticales rojas discontinuas** en las posiciones de los nudos de earth (si se ha importado earth)
- **Ventana emergente al pasar el cursor:** valor de la variable, contribución (\$), IC del 95 %, tasa (pendiente por unidad)

En los **modelos con transformación logarítmica**, el eje y se retransforma a contribuciones en dólares mediante la fórmula $\bar{y} \times (e^{f(x)} - 1)$, lo que hace que las curvas sean directamente interpretables en términos monetarios.

Para las variables de **latitud y longitud**, la pendiente se muestra por cada 0.001 grados en lugar de por grado, ya que ese es un incremento geográfico más significativo.

9.5.2. Términos factor por término suave — $s(x, \text{by} = \text{factor})$

Una línea de color por cada nivel del factor, superpuestas en un mismo gráfico. Útil para ver cómo varía el efecto de un predictor entre grupos (por ejemplo, en distintos vecindarios).

9.5.3. Términos de interacción — $ti(x_1, x_2)$ y $te(x_1, x_2)$

Mapa de calor que muestra la superficie de contribución en 2D. Escala de color: rojo (negativo) → blanco (cero) → azul (positivo). Al pasar el cursor se muestran los valores de ambas variables y el importe de la contribución.

9.5.4. Términos paramétricos: lineales y de factor

- **Términos lineales numéricos:** diagrama de dispersión de los datos con la recta ajustada superpuesta
- **Términos de factor:** gráfico de barras con la contribución media por nivel del factor

9.6 Correlation

Una matriz en mapa de calor que muestra las correlaciones de Pearson entre todos los predictores numéricos. Disponible inmediatamente después de importar los datos — no requiere ajustar el modelo.

- Gradiente de color: azul (-1) → blanco (0) → rojo ($+1$)
- Los valores de correlación se imprimen en cada celda
- Útil para identificar la concurvidad (el análogo suave de la multicolinealidad) antes del ajuste

9.7 Diagnostics

Dos conjuntos de visualizaciones de diagnóstico:

Panel de diagnóstico (mediante `gratia::appraise`): cuatro gráficos — residuos frente a ajustados, gráfico Q-Q, histograma de los residuos y respuesta frente a valores ajustados.

Diagrama de dispersión **Actual vs Predicted**: valores observados frente a predichos con una línea de referencia discontinua a 45 grados. En los modelos con transformación logarítmica, los valores se retransforman a la escala original.

9.8 RCA Adjustments

Tres histogramas que se muestran tras el cálculo del RCA (solo en los modos de tasación y de mercado):

- **Residual Adjustment %** — distribución de los ajustes por residuo como porcentaje del precio de venta
- **Net Adjustment %** — distribución de los ajustes netos
- **Gross Adjustment %** — distribución de los ajustes brutos

Cada histograma muestra la media, la mediana y la desviación típica en el subtítulo, con líneas de referencia discontinuas para la media y la mediana.

9.9 Anova

Una tabla ANOVA combinada que fusiona los términos paramétricos y suaves del resumen del GAM. Muestra el término, el tipo (paramétrico o suave) y todos los estadísticos asociados.

9.10 Mgcvc Output

Salida de texto sin formato que muestra:

- Los tiempos del ajuste
- La fórmula del GAM
- La salida de impresión del modelo
- La salida completa de `summary(model)`
- La familia y la función de enlace
- Las medidas globales de concurvidad

9.11 Sign Check

Cuando se importan nudos de earthUI, esta pestaña compara la dirección de cada variable en el modelo de earth con su dirección en el GAM:

- **earth_direction**: creciente, decreciente o mixta (según los signos de las funciones bisagra)
- **gam_direction**: creciente, decreciente o mixta (estimada a partir de la derivada del término suave sobre una malla de 100 puntos)
- **consistent**: TRUE/FALSE
- **warning**: descripción de cualquier discrepancia

Una línea de estado muestra “All directions consistent” (todas las direcciones son coherentes) o el número de variables con incoherencias.

9.12 Concurvity

La concurvidad es el análogo suave de la multicolinealidad. Esta pestaña muestra:

- **Tabla Overall Concurvity:** filas para las medidas worst, observed y estimate, y columnas para cada término suave. Con código de color: blanco (<0.5), amarillo ($0.5-0.8$), rojo (>0.8).
- **Tabla Pairwise Concurvity:** matriz por pares del peor caso, con el mismo código de color.

Los valores superiores a ~ 0.8 en la fila “worst” indican que un término suave puede aproximarse bien mediante los demás términos suaves, lo que sugiere redundancia.

9.13 Basis Check

Salida de `mgcv::gam.check()`: pruebas de adecuación de la dimensión de la base para cada término suave. Si la k de un término suave es demasiado pequeña, esta prueba lo señalará con un valor p bajo, lo que sugiere que debería aumentar k .

10 Descarga de datos

Tras el ajuste, el botón **Download Output (Excel)** (sección 7 de la barra lateral) exporta un archivo de Excel con las predicciones y los diagnósticos a la carpeta de salida de mgcv del proyecto activo.

10.1 Columnas de salida

Columna	Descripción
basis	Valor del intercepto retransformado (igual para todas las filas)
<var>_contribution	Contribución de cada término suave a la predicción. En los modelos logarítmicos, se asigna proporcionalmente a valores monetarios.
est_<target>	Predicción del modelo (por ejemplo, est_sale_price), retransformada desde la escala logarítmica si procede
residual	Valor real – predicho
calc_residual	Verificación: valor real – suma de las contribuciones (retransformada)
cqa	Puntuación de condición, calidad y atractivo (CQA), en escala 0–10
residual_sf	Residuo / superficie habitable (si se ha designado living_area)
cqa_sf	CQA calculada a partir de la ordenación por residual_sf

En el modo de tasación, la fila 1 (el sujeto) tiene residual, cqa y cqa_sf fijados en NA. Las filas se ordenan por residual_sf descendente cuando se ha designado una columna living_area.

Tras completarse correctamente aparece una marca de verificación blanca en el botón de descarga.

10.2 Puntuaciones CQA

La CQA ordena el residuo de cada fila frente a todos los demás en una escala de 0 a 10:

- **CQA alta (~9–10)**: se vendió por mucho más de lo predicho — probablemente condición, calidad o atractivo superiores
- **CQA baja (~0–1)**: se vendió por mucho menos de lo predicho — probablemente condición inferior o venta forzada
- **CQA ~5**: cerca de la mediana

11 Cálculos RCA y descarga

El flujo de trabajo RCA (Residual Constraint Approach) está disponible en los modos **Appraisal** y **Market Area Analysis**, después de ajustar el modelo. La propiedad sujeto debe estar en la fila 1.

11.1 Abrir el diálogo RCA

Haga clic en el botón **Calculate RCA Adjustments & Download** de la sección 8 de la barra lateral. Aparece un diálogo modal con:

- **Score type:** CQA o CQA per SF (si se ha designado `living_area`). La opción por pie cuadrado escala el residuo del sujeto según la superficie habitable.
- **Subject CQA Score:** admite valores de 0.00 a 10.00. El campo aparece inicialmente en blanco; cuando se ha importado un modelo de earthUI, el tipo de puntuación y el valor que usó en earthUI se trasladan desde los ajustes compartidos del proyecto.

Haga clic en **Generate** para calcular y descargar. El archivo de Excel se escribe en la carpeta de salida de mgcv del proyecto activo (`<os>_out_mgcvcv`).

11.2 Interpolación de la puntuación CQA

1. Se ordenan las puntuaciones CQA y los residuos de los comparables
2. Una interpolación lineal (`approx()`) asigna a su valor de CQA un residuo
3. Valor del sujeto = predicción del modelo + residuo interpolado
4. Las filas con peso 0 se tratan como si tuvieran el mismo residuo que el sujeto

11.3 Columnas de salida

Columna	Descripción
<code>subject_value</code>	Predicción del modelo + residuo interpolado
<code>subject_cqa</code>	Puntuación CQA introducida por el usuario
<code><var>_adjustment</code>	Contribución del sujeto – contribución del comparable (en dólares)
<code>residual_adjustment</code>	Residuo del sujeto – residuo del comparable
<code>net_adjustments</code>	Suma de todos los ajustes
<code>gross_adjustments</code>	Suma de los valores absolutos de los ajustes
<code>residual_adj_pct</code>	Ajuste por residuo como % del precio de venta
<code>net_adj_pct</code>	Ajustes netos como % del precio de venta
<code>gross_adj_pct</code>	Ajustes brutos como % del precio de venta
<code>adjusted_sale_price</code>	Precio de venta real + ajustes netos

Tras un cálculo correcto aparece una marca de verificación blanca en el botón.

12 Cuadrícula de comparación de ventas

La Sales Comparison Grid (cuadrícula de comparación de ventas) solo está disponible en el modo **Appraisal**, después de calcular los ajustes RCA. Genera un libro de Excel con formato apto para incluirlo en informes de tasación.

12.1 Diálogo de selección de comparables

Al hacer clic en el botón **Generate Sales Grid & Download** se abre un diálogo modal que lista todas las ventas comparables. Los comparables se dividen en dos grupos:

- **Recommended comps** (premarcados): ajuste bruto $< 25\%$ del precio de venta, ordenados por antigüedad de la venta (los más recientes primero). Máximo 30 comparables.
- **Other comps** (sin marcar por omisión): los comparables restantes, ordenados por porcentaje de ajuste bruto.

Seleccione sus comparables y haga clic en **Generate Sales Grid** para crear el libro de Excel.

Tipos especiales obligatorios: deben designarse `contract_date` y `living_area`. **Tipos especiales recomendados:** `latitude`, `longitude` y `lot_size` — la cuadrícula funcionará sin ellos, pero algunos campos quedarán en blanco.

12.2 Disposición del libro

El libro generado contiene varias hojas, cada una con 3 comparables uno al lado del otro:

- **Filas de encabezado:** direcciones del sujeto y de los comparables, precios de venta, fechas de venta y días en el mercado
- **Filas agrupadas:** ubicación (latitud, longitud, zona), tamaño del terreno (superficie del lote, dimensiones del sitio) y edad (edad real, edad efectiva), con encabezados de grupo y subelementos
- **Filas de variables del modelo:** una fila por término suave con la contribución del sujeto, las contribuciones de los comparables y los importes de ajuste
- **Filas de características residuales:** celdas vacías y desbloqueadas para que el tasador introduzca características no captadas por el modelo (por ejemplo, condición, calidad, vistas)
- **Filas de resumen:** % de ajuste neto, % de ajuste bruto y precio de venta ajustado, con fórmulas de Excel activas

12.3 Protección de las hojas y celdas residuales

Cada hoja está protegida para evitar ediciones accidentales, pero las celdas de valor de las características residuales están explícitamente desbloqueadas. Esto permite que los tasadores introduzcan valores para características ausentes del modelo sin alterar los cálculos de ajuste basados en fórmulas.

13 Descarga de informes

La generación de informes es un flujo de trabajo Quarto de dos pasos, que sustituye al informe de un solo paso de las versiones anteriores.

13.1 Paso 1: generar el informe Quarto

El botón **Generate Quarto Report** (disponible una vez que el modelo está ajustado) escribe un paquete `.qmd` autocontenido en la carpeta de salida de `mgcv` del proyecto activo. El paquete contiene el código fuente de Quarto, los archivos de gráficos prerrenderizados (PNG y PDF), una carga de datos `report_data.rds` con todas las tablas y los metadatos del modelo, y una plantilla de estilo `reference.docx` para la salida en Word. Como el paquete es autocontenido, puede editar el código fuente `.qmd` o combinarlo con otros documentos Quarto antes de renderizarlo.

Los gráficos se prerrenderizan al generar el paquete; si falla un gráfico concreto, se escribe una imagen de marcador de posición y el resto del paquete se genera con normalidad.

13.2 Paso 2: convertir el informe Quarto

La sección **Convert Quarto Report** renderiza un archivo `.qmd` a uno o varios formatos de salida mediante `quarto::quarto_render()`. Por omisión toma el último paquete generado, pero el botón **Browse** permite seleccionar cualquier archivo `.qmd`. Hay tres formatos disponibles:

- **HTML** — documento HTML autocontenido. Funciona en todas las plataformas sin software adicional. Opción predeterminada recomendada.
- **Word (.docx)** — renderizado con la plantilla de estilo `reference.docx` incluida. Adecuado para edición y distribución.
- **PDF** — **requiere una instalación de LaTeX** (véanse los requisitos del sistema más abajo). Si no se detecta LaTeX, la opción de PDF se oculta automáticamente de las casillas de formato.

Los renderizadores heredados de un solo paso, `render_gam_report()` y `export_gam_docx()`, siguen disponibles para uso programático.

13.3 Contenido del informe

El informe generado incluye:

- **Dataset & Model:** archivo de datos, tamaño de la muestra, familia, método e información sobre el ajuste
- **Results Summary:** R^2 , R^2 de CV, desviación explicada, AIC y otras métricas de ajuste relacionadas
- **Model Equation:** la ecuación del modelo con notación f_i para los términos suaves y coeficientes explícitos para los términos lineales, además de las definiciones de las funciones suaves. Las ecuaciones largas se dividen en varias líneas en el PDF.
- **Tablas de resumen:** términos suaves (EDF, F, valor p) y términos paramétricos (estimación, error estándar, t, valor p)
- **Interacciones habilitadas:** una matriz de los términos de interacción tensorial y de factor por término suave incluidos en el modelo (se muestra solo cuando hay interacciones)
- **Variable Importance:** gráfico de barras ordenado por el estadístico F (términos suaves) o por $|t|$ (paramétricos)
- **Smooth Effects** (todos los tipos de término):
 - Gráficos de términos suaves univariantes (versiones estáticas en `ggplot2` de los gráficos interactivos)

- Gráficos multilínea de factor por término suave
- Mapas de calor de interacción (términos ti y te)
- Gráficos de términos paramétricos (dispersión para los numéricos, barras para los de factor)
- Mapa de calor de **correlaciones entre predictores**
- **Diagnostics**: diagnósticos de residuos y diagrama de dispersión de valores reales frente a predichos
- **mgcv Summary**: la salida de texto completa del modelo ajustado
- **Apéndice de notación**: una sección final sin numerar que explica la notación de las funciones de contribución al valor empleada en el informe

Tras una generación correcta aparece una marca de verificación blanca en el botón Generate Quarto Report.

14 Exportación de las funciones del modelo

mgcvUI puede exportar las funciones suaves del GAM ajustado como código autónomo en varios lenguajes. Cada término suave se evalúa sobre una malla de 200 puntos y se exporta como tabla de consulta con interpolación lineal, lo que permite usar el modelo fuera de R sin ninguna dependencia de mgcv.

En la versión actual, la exportación de funciones es programática. Las funciones auxiliares exportadas toman un objeto de resultado ajustado (producido por `fit_gam()`) y escriben los archivos que se describen a continuación:

- `generate_r_code()`, `generate_python_code()` y `generate_cpp_code()` — archivos de código fuente por lenguaje
- `export_functions_sqlite()` — base de datos SQLite con tablas de consulta
- `export_functions_zip()` — agrupa cualquier combinación de los cuatro formatos en un archivo zip

14.1 Funciones de R

Archivo: `gam_functions.R`

- Comentarios de cabecera con la variable de respuesta, el R^2 , el R^2 de CV y la fórmula.
- Una constante `INTERCEPT`.
- Una `approxfun()` por cada variable suave, con vectores `x`/`y` precalculados.
- Los coeficientes de los términos lineales como constantes con nombre.
- Uso: sumar el intercepto, cada `g_var(value)` y cada `coef_var * value`.

14.2 Funciones de Python

Archivo: `gam_functions.py`

- Importación de `numpy`.
- Constante `INTERCEPT`.
- Por variable: arreglos `_x_var` y `_fx_var`, y función `g_var(x)` que usa `np.interp()`.
- Los coeficientes de los términos lineales como constantes `COEF_*`.

14.3 Cabecera de C++

Archivo: `gam_functions.hpp`

- Solo cabecera, con `#pragma once`.
- Espacio de nombres `gam_functions`.
- Función auxiliar `interp()` basada en `std::lower_bound`.
- Por variable: función inline con arreglos `static const`.
- Los coeficientes de los términos lineales como `constexpr double`.

14.4 Base de datos SQLite

Archivo: `gam_functions.sqlite`

- Tabla `smooth_grids`: columnas (`variable`, `x`, `fx`) — las tablas de consulta de 200 puntos.
- Tabla `model_info`: `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.
- Tabla `linear_terms`: `variable`, `coeficiente`.
- Tabla `intercept`: valor del intercepto.

15 Comparación con earthUI y glmnetUI

La aplicación earth y sus Post Processing Modules forman un único conjunto de herramientas para el modelado de regresión. Comparten el mismo formato de datos, los mismos tipos especiales de columna, el mismo flujo de trabajo RCA y los mismos conjuntos de datos de demostración, pero usan motores de modelado distintos.

15.1 Diferencias principales

Característica	mgcvUI	earthUI	glmnetUI
Método	GAM (mgcv)	MARS/Earth (earth)	Elastic net (glmnet)
Relaciones	No lineales suaves	Lineales por tramos	Lineales
Selección de variables	Contracción con <code>select=TRUE</code>	Poda por pasos	Penalización Lasso
Interacciones	Productos tensoriales (te/ti)	Productos de bisagras (grado 2–3)	Productos por pares
Efectos parciales	Curvas suaves con IC	Funciones g (por tramos)	Pendientes lineales
Cadena con earth	Importa los nudos de earth	Fuente de los nudos	Importa la base de bisagras
Exportación de código	R, Python, C++, SQLite	R, Python, C++, SQLite	—
Formatos de informe	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

Conviene señalar un elemento diferenciador exclusivo de earthUI: earthUI incluye una función experimental de Prolog/DCG que extrae atributos estructurados de las propiedades a partir de las observaciones de texto libre del MLS.

15.2 Funciones compartidas

Las tres herramientas ofrecen:

- Un árbol de proyectos **regProj** compartido (por omisión `~/regProj`) — los proyectos, los archivos de entrada y las salidas de cada método se comparten entre las aplicaciones hermanas
- Importación de CSV y Excel con conversión de los nombres de columna a `snake_case`
- Compatibilidad con la configuración regional de más de 30 países, con separador de CSV, signo decimal y tamaño de papel
- Designaciones especiales de columnas (`contract_date`, `living_area`, `area`, `latitude`, `longitude`, etc.)
- Cálculo automático de `sale_age` a partir de la fecha efectiva
- Alternancia entre modo oscuro y claro (tema Nord)
- Persistencia de los ajustes (SQLite y/o almacenamiento local)
- Mapa de calor de correlaciones, importancia de las variables y diagnósticos
- Puntuación CQA y flujo de trabajo de ajustes RCA
- Sales Comparison Grid con selección de comparables y salida a Excel con formato
- Compatibilidad con los mismos conjuntos de datos de demostración (`Appraisal_1.csv`)

15.3 Flujo de trabajo recomendado

- **Empiece con earthUI** para descubrir las variables importantes, las relaciones no lineales y las interacciones. La construcción automática de modelos de earth ofrece el mejor punto de partida.
- **Refine con mgcvUI** si desea efectos parciales más suaves. Importe los nudos de earthUI para una transición sin fricción.
- **Use glmnetUI** cuando necesite un modelo puramente lineal con selección agresiva de variables o restricciones de signo en los coeficientes.
- **Compare las tres** sobre los mismos datos para valorar si está justificada la complejidad adicional de los modelos no lineales.

16 Conjunto de datos de demostración: Appraisal_1.csv

16.1 Descripción

El conjunto de datos de demostración se comparte con earthUI y glmnetUI. Se encuentra en la carpeta `demo_mls/` del código fuente de mgcvUI o, si earthUI está instalado:

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

El archivo contiene aproximadamente 1.500 registros de ventas residenciales procedentes de una exportación simulada de un MLS, con la propiedad sujeto en la fila 1 y las ventas comparables en las filas restantes. Los datos representan ventas de viviendas unifamiliares en un mercado con varias zonas y con una diversidad de tamaños, edades y ubicaciones de las propiedades.

No se trata de datos reales, sino de datos basados en un vecindario realista del norte de California. Toda la información de identificación ha sido alterada o eliminada.

16.2 Columnas

Columna	Tipo	Tipo especial	Descripción
<code>sale_type</code>	character	<code>sale_type</code>	Código de tipo de registro que distingue el sujeto de las ventas comparables
<code>weight</code>	numeric	<code>weight</code>	Peso de la observación (0 = excluir del ajuste)
<code>id</code>	numeric	<code>display_only</code>	Identificador interno del registro
<code>property_id</code>	numeric	<code>display_only</code>	Identificador de propiedad del MLS
<code>listing_id</code>	character	<code>display_only</code>	Número de listado del MLS
<code>parcel_number</code>	character	<code>display_only</code>	Número de parcela del catastro del condado (APN)
<code>street_address</code>	character	<code>display_only</code>	Dirección de la propiedad
<code>city_name</code>	character	<code>display_only</code>	Ciudad
<code>postal_code</code>	character	<code>display_only</code>	Código postal
<code>county_name</code>	character	<code>display_only</code>	Condado
<code>contract_date</code>	Date	<code>contract_date</code>	Fecha de contrato de venta (calcula <code>sale_age</code>)
<code>sale_age</code>	numeric	—	Días desde la fecha de contrato hasta la fecha efectiva (precalculado)
<code>coe_date</code>	Date	<code>display_only</code>	Fecha de cierre del depósito en garantía
<code>listing_status</code>	character	<code>display_only</code>	Estado del listado (por ejemplo, "Sold")
<code>sale_price</code>	numeric	(objetivo)	Precio de venta — variable de respuesta
<code>rent</code>	numeric	—	Renta mensual (para modelos con varios objetivos)
<code>list_price</code>	numeric	<code>display_only</code>	Precio de listado
<code>original_list_price</code>	numeric	<code>display_only</code>	Precio de listado original
<code>living_sqft</code>	numeric	<code>living_area</code>	Superficie habitable bruta en pies cuadrados
<code>ppsf</code>	numeric	<code>display_only</code>	Precio por pie cuadrado
<code>beds_total</code>	integer	—	Número de dormitorios
<code>baths_total</code>	numeric	—	Número total de baños (por ejemplo, 2.5 = 2 completos + 1 medio)
<code>lot_size</code>	numeric	<code>lot_size</code>	Tamaño del terreno en pies cuadrados
<code>area_id</code>	integer	<code>area</code>	Identificador de zona del MLS
<code>area_text</code>	character	<code>display_only</code>	Nombre de la zona
<code>age</code>	numeric	<code>actual_age</code>	Edad de la propiedad en años
<code>year_built</code>	integer	<code>display_only</code>	Año de construcción
<code>latitude</code>	numeric	<code>latitude</code>	Latitud (redondeada a 3 decimales para el modelo)
<code>longitude</code>	numeric	<code>longitude</code>	Longitud (redondeada a 3 decimales para el modelo)
<code>latitude6</code>	numeric	<code>display_only</code>	Latitud con precisión completa
<code>longitude6</code>	numeric	<code>display_only</code>	Longitud con precisión completa
<code>garage_spaces</code>	integer	—	Número de plazas de garaje
<code>fp_count</code>	integer	—	Número de chimeneas

Columna	Tipo	Tipo especial	Descripción
no_of_stories	numeric	—	Número de plantas
style	character	—	Estilo arquitectónico
view	character	—	Tipo de vistas (por ejemplo, “Neighborhood”, “Hills”)
days_on_market	integer	dom	Días en el mercado
listing_date	Date	listing_date	Fecha de listado
sale_concessions	numeric	concessions	Concesiones del vendedor

16.3 Inicio rápido sugerido

1. Inicie la rutina de mgcv: `earthUI::launch_mgcv()`
2. Cree un proyecto nuevo (Sección 1) con el propósito **For Appraisal**
3. Coloque **Appraisal_1.csv** en la carpeta de entrada del proyecto y selecciónelo en la Sección 2 (haga clic en **Refresh** si es necesario)
4. Seleccione **sale_price** como objetivo
5. Opcionalmente, fije Response Transform en **Log (natural)** para obtener ajustes proporcionales
6. Asigne los tipos especiales tal como se indica en la tabla anterior
7. Incluya los predictores **sale_age**, **living_sqft**, **baths_total**, **lot_size**, **area_id** (como factor), **age**, **latitude**, **longitude** y **garage_spaces**
8. Mantenga los valores predeterminados: REML, gamma = 1.2, base tp, k = auto, select = TRUE
9. Haga clic en **Fit Mgcv GAM Model**
10. Revise las pestañas Summary, Contribution (gráficos de términos suaves) y Diagnostics
11. Descargue la salida (paso 7) y revise la ordenación por CQA
12. Calcule los ajustes RCA (paso 8) con una puntuación CQA de ~5.00
13. Genere una Sales Comparison Grid (paso 9) con los comparables recomendados

17 Requisitos del sistema y resolución de problemas

17.1 Plataformas admitidas

mgcvUI funciona en **macOS**, **Windows** y **Linux**. La aplicación se desarrolla y se prueba principalmente en macOS con RStudio. Más abajo se ofrecen notas específicas de cada plataforma.

17.2 Versión de R

Se requiere $R \geq 4.1.0$. Se recomienda encarecidamente RStudio Desktop (2023.06 o posterior) — incluye **Quarto** y **pandoc** (necesarios para renderizar los informes) y ofrece un entorno cómodo para iniciar la aplicación.

17.3 Paquetes de R requeridos

Los siguientes paquetes se instalan automáticamente como dependencias:

Paquete	Finalidad
shiny	Marco de aplicaciones web
mgcv	Ajuste de los modelos GAM
valengrCore	Núcleo compartido de tasación (papeles especiales, traspaso desde earthUI)
ggplot2, plotly	Gráficos de contribución estáticos e interactivos
gratia	Extracción de términos suaves y diagnósticos
DT	Tablas de datos interactivas
quarto, rmarkdown	Generación y renderizado de informes Quarto
readxl	Importación de archivos de Excel
shinyFiles	Exploradores de archivos y carpetas
RSQLite, DBI	Bases de datos de ajustes y de proyectos (SQLite)
jsonlite	Serialización del estado de las variables

17.4 Dependencias opcionales

Componente	Cuándo se necesita
LaTeX (TinyTeX, MiKTeX o MacTeX)	Solo para generar informes en PDF. Si no se detecta ninguna instalación de LaTeX, la opción de PDF se oculta automáticamente del desplegable de formatos. Instálelo con: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	Fuente Roboto Condensed para los gráficos de ggplot2. Si no está disponible o se trabaja sin conexión, la aplicación recurre automáticamente a la fuente de palo seco del sistema.
earth ($\geq 5.3.0$)	Solo se necesita para la cadena de importación desde earthUI. No es necesario para el uso autónomo.
callr	Ajuste del modelo en segundo plano (mantiene la aplicación receptiva). Sin él, el ajuste se ejecuta de forma síncrona.
writexl, openxlsx	Descargas a Excel (salida del modelo, RCA, cuadrícula de ventas). La aplicación lo indica si faltan.
bslib	Temas de Bootstrap 5 (Nord Light/Dark).
officer	Exportación programática heredada a Word (<code>export_gam_docx()</code>).

17.5 Notas específicas de cada plataforma

17.5.1. macOS

La plataforma con menos problemas. Se recomienda Homebrew para las bibliotecas del sistema. Si se necesitan informes en PDF:

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2. Windows

Funciona bien con RStudio Desktop. Notas principales:

- **Equipos corporativos o restringidos:** si el directorio temporal o el directorio de datos del usuario están restringidos, la persistencia de los ajustes y la generación de informes pueden fallar. La aplicación mostrará una advertencia en la consola de R pero seguirá funcionando sin persistencia.
- **LaTeX:** instale TinyTeX desde R (el mismo comando que en macOS más arriba) o instale MiKTeX desde miktex.org.
- **Rutas de archivo largas:** Windows tiene un límite de 260 caracteres para las rutas. Mantenga corta la ruta de su directorio de trabajo.

17.5.3. Linux

La plataforma más variable entre distribuciones. Los usuarios de Ubuntu o Debian pueden necesitar bibliotecas del sistema antes de que los paquetes de R compilen:

```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

Para los informes en PDF: `tinytex::install_tinytex()` o instale `texlive-full` desde el gestor de paquetes del sistema.

En servidores sin interfaz gráfica, los gráficos interactivos de plotly se renderizan como alternativas estáticas en los informes. La propia aplicación Shiny requiere una conexión desde un navegador web.

17.6 Degradación elegante

mgcvUI está diseñado para funcionar incluso cuando faltan componentes opcionales:

Componente ausente	Comportamiento
Sin LaTeX	La opción de PDF se oculta del desplegable de formatos de informe. HTML y Word siguen disponibles.
Sin internet o fallo de las fuentes	La fuente Roboto Condensed se sustituye por la de palo seco del sistema. Se registra un mensaje en la consola.
Sin RSQLite o sistema de archivos de solo lectura	Los ajustes no persisten entre sesiones. Se muestra una advertencia en la consola. Por lo demás, la aplicación funciona con normalidad.
Fallo de <code>gratia::appraise</code>	La pestaña Diagnostics y los informes muestran un mensaje de marcador de posición. Use <code>mgcv::gam.check()</code> en la consola como alternativa.
Directorio temporal sin permiso de escritura	La generación del informe falla con un mensaje de error claro. Se resuelve fijando la variable de entorno <code>TMPDIR</code> en una ubicación con permiso de escritura.

17.7 Resolución de problemas

17.7.1. “PDF option is not available”

No se ha detectado ninguna instalación de LaTeX. Ejecute en R:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

Después reinicie la aplicación. La opción de PDF aparecerá.

17.7.2. “Settings will not persist”

No se puede crear el directorio de la base de datos SQLite ni escribir en él. Compruebe los permisos de:

- **macOS/Linux:** `~/Library/Application Support/R/mgcvUI/` o `~/local/share/R/mgcvUI/`
- **Windows:** `%APPDATA%/R/data/mgcvUI/`

17.7.3. “Diagnostics plot unavailable”

La función `gratia::appraise()` falla ocasionalmente con modelos complejos (muchas interacciones tensoriales, familias no estándar). Se trata de un problema conocido del paquete original. Solución alternativa: ejecute `mgcv::gam.check(model)` en la consola de R para ver directamente los gráficos de diagnóstico.

17.7.4. “Report error: number of items to replace...”

Este era un problema conocido de las versiones anteriores, relacionado con el paquete **gratia**. Asegúrese de tener instalada la última versión de **mgcvUI**. La versión actual captura este error y produce el informe con un marcador de posición en la sección de diagnósticos.

17.7.5. El puerto 7880 ya está en uso

Si la aplicación no arranca porque el puerto 7880 está ocupado (por una sesión anterior), ejecute:

```
# macOS/Linux:
lsof -ti:7880 | xargs kill

# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6. La instalación de paquetes falla en Linux

Faltan bibliotecas del sistema. Instale las cabeceras de desarrollo enumeradas en la sección de Linux más arriba y vuelva a intentar `install.packages()`.

Referencias

- Wood, S. N. (2003). «Thin-plate regression splines». En: *Journal of the Royal Statistical Society (B)* 65.1, págs. 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». En: *Journal of the American Statistical Association* 99.467, págs. 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». En: *Journal of the Royal Statistical Society (B)* 73.1, págs. 3-36. DOI: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- (2017). *Generalized Additive Models: An Introduction with R*. 2.^a ed. Chapman y Hall/CRC.
- Wood, S. N., N. Pya y B. Säfken (2016). «Smoothing parameter and model selection for general smooth models (with discussion)». En: *Journal of the American Statistical Association* 111, págs. 1548-1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).

Fundamentos

Panorama general

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor, editor • 1 de agosto de 2026

Versión 1.0.0

DOI: [10.67330/82vssv81](https://doi.org/10.67330/82vssv81)

Foundations publica tratamientos precisos y citables de los conceptos que subyacen a la metodología moderna de valuación. Las entradas toman términos que los tasadores en ejercicio encuentran — a veces con otros nombres, a veces sin reconocerlos — y los sitúan en su contexto formal: de dónde provienen, qué significan técnicamente, dónde aparecen en la práctica y por qué importan para la defendibilidad.

El número inaugural establece los seis conceptos centrales del marco hedónico, dispuestos de modo que cada entrada se apoya en el vocabulario de las anteriores. La secuencia va desde *Bien heterogéneo* (la propiedad de los bienes inmuebles que hace necesaria la metodología de valuación en primer lugar), pasando por *Espacio de características* y la *Función de precios hedónicos* (el aparato formal que organiza el conjunto de atributos que representa un inmueble), por *Precio implícito* y *Regresión hedónica* (cómo se recupera de la evidencia de mercado la contribución marginal de cada atributo), hasta llegar a *Variable latente* (qué representa realmente la señal residual no recuperada en cualquier modelo hedónico).

Un único ejemplo continuo — ocho comparables de un vecindario de Pacifica, California — recorre las seis entradas y constituye el hilo pedagógico del número.

Por qué estos seis, en este orden

La elección de las seis entradas no es arbitraria. Cada concepto es estructural para la metodología hedónica moderna, cada uno se encuentra genuinamente en la práctica cotidiana de la tasación inmobiliaria, y cada uno se malinterpreta o queda sin nombrar con la frecuencia suficiente como para que darle una forma citable tenga valor práctico.

El orden lo dicta la dependencia conceptual. *Bien heterogéneo* es la raíz: nombra la propiedad de los bienes inmuebles (y de los autos usados, los mercados laborales y algunos otros bienes) que hace necesario todo el marco hedónico. *Espacio de características* es el aparato formal para representar bienes heterogéneos como vectores; nada más puede decirse con precisión sin él. La *función de precios hedónicos* es el objeto de equilibrio que asigna precios a los puntos del espacio de características. Los *precios implícitos* son sus derivadas parciales — las contribuciones marginales que cada ajuste en un cuadro de comparación de ventas puede leerse como afirmando, nombre o no el tasador esos precios como tales. La *regresión hedónica* es el estimador empírico que recupera los precios implícitos a partir de las transacciones observadas; es la entrada donde la teoría se convierte en análisis de datos. Las *variables latentes* comienzan donde empiezan los residuos de la regresión: los atributos no observados cuyos efectos sobreviven en el residuo como señal estructurada y no como ruido descartable.

Un lector que avance en este orden debería salir con un hilo continuo desde “¿por qué existe siquiera la metodología de valuación?” hasta “¿qué debemos hacer con los residuos de un modelo ajustado?” — un recorrido que, según la experiencia del autor, la mayoría de los tasadores completa solo por partes a lo largo de muchos años de práctica.

El ejemplo continuo

Un único conjunto de datos de ocho comparables residenciales de Pacifica, California, recorre las seis entradas:

Comp.	GLA (sf)	Lote (sf)	Vista	Cond.	Precio de venta
A	1,650	9,000	sí	buena	\$1,425,000
B	1,650	7,200	no	buena	\$1,280,000
C	1,850	7,500	no	buena	\$1,407,500
D	1,450	7,000	no	buena	\$1,155,000
E	1,700	8,000	sí	excelente	\$1,470,000
F	1,550	7,100	no	media	\$1,177,500
G	1,700	10,500	no	buena	\$1,392,500
H	1,600	7,300	no	buena	\$1,312,500

El ejemplo es deliberadamente lo bastante pequeño para razonarse a mano y lo bastante rico para ejercitar cada concepto. La entrada 001 examina la diferencia de \$145,000 entre los comparables A y B como la contribución conjunta de dos características (vista y lote) entrelazadas en una sola observación de ventas pareadas. La entrada 002 sitúa los ocho comparables como puntos en un espacio de características de cuatro dimensiones y muestra que las métricas de distancia ingenuas los ordenan mal. La entrada 003 propone una forma funcional para la función de precios hedónicos de este vecindario. La entrada 004 recupera analíticamente los precios implícitos y descompone la diferencia original entre A y B en una contribución de \$127,039 proveniente de la vista y el lote, con \$18,000 sin explicar. La entrada 005 ajusta el modelo formalmente en R e informa la salida completa de `lm()`, incluidos los errores estándar y los residuos. La entrada 006 examina el único residuo que el modelo no puede explicar — una desviación de +\$46,455 en el comparable H — y lo conecta con el planteamiento de variables latentes que subyace a la metodología del Residual Constraint Approach desarrollada por el autor en otros trabajos.

El conjunto de datos está disponible en el repositorio del número en `shared/data/pacifica.comps.csv` para los lectores que deseen reproducir cualquiera de los cálculos.

Cómo leer este número

Un lector nuevo en la metodología hedónica debería leer las seis entradas en orden. Las dependencias son reales: cada entrada asume el vocabulario establecido en las anteriores, y saltarse pasos invita a confusiones que las propias entradas no pueden anticipar.

Un lector con exposición previa a la regresión hedónica — lo que incluye a la mayoría de los tasadores que han tomado un curso avanzado de estadística, trabajado con un AVM o leído la literatura de valuación masiva — puede usar el número como repaso y como obra de referencia. Las entradas están escritas para ser citables de manera atómica; un artículo publicado en otra parte de la revista que haga referencia a *precio implícito* puede enlazar a la entrada 004 por su DOI en lugar de volver a derivar el concepto.

Un lector interesado específicamente en la metodología del Residual Constraint Approach debería leer primero la entrada 006 para entender el planteamiento de variables latentes, volver luego a las entradas 003-005 para el aparato hedónico que el RCA extiende, y después a las entradas 001-002 para el vocabulario fundacional sobre el que todo descansa.

Un lector interesado en la cuestión de *qué hace defendible un ajuste de tasación* encontrará que la sección “Por qué importa para la defendibilidad” de cada entrada, leída en secuencia, equivale

a un breve ensayo sobre el estatus metodológico de la valuación basada en ajustes. Ese ensayo no está ensamblado explícitamente en ninguna parte del número, pero es el tema implícito de cada entrada.

Postura editorial

Foundations no es una sección de libro de texto. Cada entrada es lo bastante breve para leerse de una sentada, está escrita para un público ya versado en la práctica de la tasación inmobiliaria, y está diseñada para dar una formulación precisa de un concepto en lugar de una exposición exhaustiva. El objetivo es proporcionar un referente estable que otros artículos de la revista puedan citar.

Las entradas se revisan editorialmente y pueden llevar comentarios invitados de un experto en la materia. Se publican bajo una licencia Creative Commons Atribución y reciben DOI tanto a través de Crossref (como cita canónica) como de Zenodo (como archivo versionado). Cada entrada puede afinarse con el tiempo — las revisiones menores actualizan la versión de Zenodo, las revisiones mayores actualizan ambos DOI con una nota editorial que describe el cambio. El DOI de Crossref siempre resuelve a la versión más reciente; las versiones históricas específicas son direccionables a través de Zenodo.

Los números futuros extenderán el vocabulario de *Foundations* a áreas adyacentes: dependencia espacial, esperanza condicional, endogeneidad, sesgo por variables omitidas, identificación y el aparato formal que distingue las variables proxy de las variables latentes. Las sugerencias de conceptos que merezcan tratamiento son bienvenidas y pueden enviarse al editor.

Las seis entradas de este número:

1. **Bien heterogéneo** — por qué debe existir la metodología de valuación en primer lugar.
2. **Espacio de características** — la representación vectorial de los atributos del conjunto.
3. **Función de precios hedónicos** — la aplicación de equilibrio de las características a los precios.
4. **Precio implícito** — el precio marginal de mercado de cada característica.
5. **Regresión hedónica** — la estimación de la función a partir de transacciones observadas, con un ejemplo completo resuelto en R.
6. **Variable latente** — la recuperación de señal estructurada a partir de los residuos.

Referencias.

- 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

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (000-overview.pl) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```

%% 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).

```

Bien heterogéneo

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 1 de agosto de 2026

Versión 1.0.0

DOI: [10.67330/slscsv65](https://doi.org/10.67330/slscsv65)

Definición formal. Los bienes heterogéneos son bienes que no son intercambiables porque cada uno es un conjunto distinto de características que los compradores valoran por separado, y cuyas características no se comercian por separado.

Planteamiento intuitivo. Un bien homogéneo es aquel en el que cualquier unidad sirve: un barril de crudo WTI, un bushel de maíz amarillo No. 2, un kilovatio-hora de la red eléctrica, una acción de una empresa determinada. Los estándares de clasificación o las especificaciones que respaldan estos bienes existen precisamente para hacer intercambiables las unidades individuales, de modo que los precios se apliquen a la categoría y no a la unidad particular. Un bien heterogéneo es aquel en el que la unidad específica importa — *esta* casa, no *aquella* casa; *este* auto usado, no el que está al lado.

Dos rasgos hacen que un bien sea heterogéneo en el sentido técnico. Primero, los compradores valoran múltiples atributos distinguibles que varían entre las unidades del mercado. Segundo, esos atributos vienen agrupados por la tecnología de producción del bien; no pueden comprarse por separado. No se pueden comprar 200 pies cuadrados de superficie habitable (GLA) por sí solos. No se puede adquirir un distrito escolar con independencia de una parcela de terreno. Las características llegan agrupadas, y el mercado fija un precio sobre conjuntos completos, no sobre las características mismas.

Los bienes inmuebles son el bien heterogéneo canónico y exhiben esta propiedad con una intensidad inusual. Las características de localización son perfectamente no replicables: dos parcelas no pueden ocupar la misma ubicación, de modo que cada inmueble tiene al menos un atributo (sus coordenadas) que lo distingue de cualquier otro inmueble del mercado. Los atributos físicos, legales y ambientales añaden dimensiones adicionales de variación a un vector que, tomado en su conjunto, es único para cada inmueble.

Dónde lo encuentran los tasadores. Cada ajuste en el enfoque de comparación de ventas es un reconocimiento implícito de que el inmueble objeto de la tasación y los comparables son bienes heterogéneos. Si los inmuebles fueran homogéneos, los precios de venta ya serían directamente comparables; no haría falta cuadro alguno. El cuadro de comparación es un intento de conciliar los precios de conjuntos distintos ajustando por las diferencias en la composición del conjunto. Conviene señalar lo que el cuadro *no* hace: nunca descompone un precio de venta en contribuciones de valor por características. El cuadro tradicional no tiene columna para las contribuciones de valor — solo ajustes, estimados directamente, característica por característica.

El enfoque del costo maneja la heterogeneidad reconstruyendo el conjunto a partir de sus componentes — cimentación, estructura, acabados, depreciación — y valorando cada uno por separado. El enfoque de ingresos la maneja abstrayendo el conjunto en un único perfil de flujos de efectivo, con tasas de capitalización y de rendimiento que absorben la heterogeneidad del conjunto en un multiplicador derivado del mercado. Los tres enfoques de valuación son respuestas metodológicas al

hecho subyacente de que los bienes inmuebles son un bien heterogéneo. Difieren en cómo navegan ese hecho, no en si lo confrontan.

Los tasadores también encuentran la heterogeneidad de manera informal siempre que razonan sobre la selección de comparables, el análisis de pares comparables (*matched pairs*) o las comparaciones “de manzanas con manzanas”. El instinto de buscar comparables que se parezcan en tantos atributos como sea posible es la respuesta del profesional a la heterogeneidad: minimizar la distancia entre los conjuntos del inmueble objeto y del comparable, para que el componente no observado del conjunto contribuya menos a la diferencia entre los precios de venta.

Por qué importa para la defendibilidad. La heterogeneidad de los bienes inmuebles es la razón por la que la metodología de valuación debe existir como disciplina. Si los inmuebles fueran intercambiables, el valor de mercado sería directamente observable en los precios publicados; la valuación se reduciría a una consulta. Como no lo son, el valor debe inferirse a partir de los precios de inmuebles relacionados-pero-no-idénticos, y cada paso de inferencia es una elección metodológica que debe justificarse.

Reconocer formalmente la heterogeneidad aclara varias cuestiones prácticas de las que dependen la USPAP y la defendibilidad de las Standards Rules. La selección de comparables es la elección de elementos en el espacio de características lo bastante cercanos al inmueble objeto como para que los precios implícitos recuperados de ellos generalicen. La razonabilidad de un ajuste depende de si el precio implícito subyacente puede sustentarse con evidencia de mercado. La conciliación entre enfoques refleja el juicio del tasador sobre cuál de ellos navega con mayor credibilidad la heterogeneidad de *este* inmueble. Cada una de estas cuestiones de defendibilidad se vuelve más nítida una vez que se nombra “bien heterogéneo” como la condición subyacente.

Ejemplo de tasación resuelto. Dos ventas recientes en el mismo vecindario de Pacifica, California:

Atributo	Comp. A	Comp. B
GLA (sf)	1,650	1,650
Recámaras / Baños	3 / 2	3 / 2
Lote (sf)	9,000	7,200
Vista	Sí	No
Precio de venta	\$1,425,000	\$1,280,000

Bajo un planteamiento de bienes homogéneos, se trata del mismo inmueble: GLA idéntica, mismo número de habitaciones, mismo vecindario. Bajo el planteamiento de bienes heterogéneos, son conjuntos distintos, que difieren en dos características (lote y vista) que los compradores valoran por separado. La diferencia de precio de \$145,000 es la contribución *conjunta* de esos dos atributos — dos precios implícitos entrelazados en un solo par.

Recuperar los precios implícitos individualmente requiere más pares, idealmente variando un atributo a la vez. Tres o cuatro ventas adicionales que abarquen las combinaciones vista/sin-venta y lote-grande/lote-pequeño empiezan a identificar los precios implícitos por separado. Con suficiente variación entre suficientes comparables, la regresión hedónica realiza formalmente el desentrelazamiento; con variación limitada, el tasador queda restringido a un razonamiento acotado de pares comparables cuya fiabilidad depende explícitamente de cuán cercanos estén los conjuntos entre sí en el espacio de características.

El ejemplo ilustra por qué la heterogeneidad es la condición raíz: incluso dos casas “idénticas” no son idénticas, y el aparato metodológico de la tasación inmobiliaria existe precisamente para manejar la distancia entre ellas.

Referencias.

- 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.

Referencias cruzadas: espacio de características; función de precios hedónicos; precio implícito.

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (001-heterogeneous-good.pl) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 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).
```

Espacio de características

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 3 de agosto de 2026

Versión 1.0.0

DOI: [10.67330/3yymh010](https://doi.org/10.67330/3yymh010)

Definición formal. El espacio de características de una clase de bienes heterogéneos es el espacio vectorial cuyas coordenadas indexan los atributos que los compradores valoran, y en el cual cada bien se representa mediante el vector de sus niveles de atributos.

Planteamiento intuitivo. La entrada anterior estableció que un bien heterogéneo es un conjunto distinto de características. La siguiente pregunta natural es: ¿cómo debemos describir y comparar formalmente tales conjuntos? La respuesta, debida a Lancaster (1966), consiste en tratar cada bien como un punto en un espacio vectorial cuyos ejes son las características que interesan a los compradores.

En esta representación, un inmueble residencial de Pacífica ya no es “una casa en Park Pacífica” sino un vector con coordenadas para la superficie habitable (GLA), el tamaño del lote, la vista, el estado de conservación, el distrito escolar, la distancia al océano, etcétera. Dos inmuebles que compartieran todos estos valores de atributos ocuparían el mismo punto del espacio de características; los inmuebles que difieren en cualquier atributo ocupan puntos distintos. La geometría del espacio codifica cuán cercanos o cuán diferentes son dos inmuebles, de un modo que captura todas las dimensiones en las que los compradores hacen distinciones.

La dimensionalidad del espacio depende de qué atributos sean relevantes para el mercado en estudio. Para una tasación residencial típica, el espacio relevante podría tener una docena de dimensiones: físicas (GLA, lote, recámaras, baños, antigüedad, calidad, estado), de localización (distrito escolar, vista, proximidad a atractivos o a elementos desfavorables) y legales/económicas (zonificación, servidumbres). Para una tasación comercial, las dimensiones relevantes se desplazan hacia los atributos de flujo de efectivo (estructura de los arrendamientos, calidad de los inquilinos, ocupación), pero la idea geométrica subyacente es la misma.

Lo que hace útil al espacio de características es que convierte una intuición cualitativa del profesional (“estos comparables se parecen al inmueble objeto”) en una afirmación cuantitativa (“estos comparables están cerca del inmueble objeto en el espacio de características”). Una vez definida la distancia, cada decisión de comparación de ventas puede examinarse como una afirmación sobre la posición en el espacio.

Dónde lo encuentran los tasadores. Los tasadores trabajan constantemente en el espacio de características, por lo general sin nombrarlo. Cada renglón de un cuadro de comparación de ventas es un eje de coordenadas; cada comparable es un vector fila de valores de coordenadas; cada ajuste es un intento de moverse a lo largo de un eje manteniendo fijos los demás.

Los ocho comparables de Pacífica que usaremos a lo largo de este número ilustran la construcción directamente. Cada comparable es un punto en un espacio de características de cuatro dimensiones generado por la GLA, el tamaño del lote, la vista y el estado de conservación:

Comp.	GLA (sf)	Lote (sf)	Vista	Cond.	Precio de venta
A	1,650	9,000	sí	buena	\$1,425,000
B	1,650	7,200	no	buena	\$1,280,000
C	1,850	7,500	no	buena	\$1,407,500
D	1,450	7,000	no	buena	\$1,155,000
E	1,700	8,000	sí	excelente	\$1,470,000
F	1,550	7,100	no	media	\$1,177,500
G	1,700	10,500	no	buena	\$1,392,500
H	1,600	7,300	no	buena	\$1,312,500

Los comparables A y B de la entrada anterior son las dos primeras filas; los seis restantes amplían el ejemplo de maneras que explotaremos a lo largo del resto del número.

Cada comparable es un punto en $\mathbb{R}^4$:

- El comparable A está en (1650, 9000, 1, 2).
- El comparable B está en (1650, 7200, 0, 2).
- El comparable G está en (1700, 10500, 0, 2).

Las coordenadas de vista y estado de conservación se codifican numéricamente (la vista como $\{0, 1\}$, el estado como $\{1, 2, 3\}$); el precio *no* es una coordenada del espacio de características, sino una función de la posición dentro de él.

Por qué importa para la defendibilidad. El marco del espacio de características precisa qué es realmente la selección de comparables. Un comparable no es simplemente “un inmueble similar”, sino un punto de la región relevante del espacio de características cercano al inmueble objeto. Cuanto más cerca esté un comparable del inmueble objeto en el espacio, menores serán los ajustes requeridos y más confiable será el valor indicado.

Esta reformulación expone varias cuestiones prácticas que el cuadro informal oculta:

- *¿Qué dimensiones importan?* La elección de los ejes es una decisión de modelado. Incluir “vista” como eje solo tiene sentido si el mercado le asigna un precio distinto; incluir la “orientación Feng Shui” solo tiene sentido si un segmento del mercado la valora. Una tasación defendible exige defender los ejes, no solo los valores a lo largo de ellos.
- *¿Qué métrica de distancia aplica?* Dos comparables pueden diferir en \$200,000 de precio porque están alejados en el espacio de características, pero el cálculo de la distancia requiere un escalamiento para cada eje. La distancia euclidiana sin transformar trata un pie cuadrado de GLA como equivalente a un pie cuadrado de lote, cosa que el mercado no hace.
- *¿Están los comparables en una región donde la función de precios se comporta bien?* Si el inmueble objeto se sitúa en un rincón escasamente muestreado del espacio de características, ningún comparable queda cerca de él y la indicación de valor es necesariamente extrapolativa.

Ejemplo de tasación resuelto. Los ocho comparables de Pacífica ocupan una región relativamente compacta del espacio de características de cuatro dimensiones. Consideremos la pregunta de qué comparable es el “más cercano” a un inmueble objeto hipotético en (1,700, 7,500, 1, 2) — 1,700 sf, lote de 7,500 sf, con vista, estado de conservación bueno.

Si usamos ingenuamente la distancia euclidiana con ejes sin escalar, la dimensión del lote domina porque sus valores (7,000–10,500) son numéricamente mucho mayores que los de la vista (0–1) o el estado (1–3). Distancias al inmueble objeto:

Comp.	Distancia (euclidiana sin escalar)	Notas
A	1,501	la vista coincide; el lote domina
B	308	la vista difiere; lote cercano
C	230	la vista difiere; lote cercano
D	559	la vista difiere; GLA desviada
E	502	la vista coincide; el estado difiere
F	422	difieren la vista y el estado
G	3,000	la vista difiere; lote muy lejano
H	224	la vista difiere; el más cercano de todos

Por distancia bruta, el comparable H es el vecino más cercano. Pero el comparable H no tiene vista y el inmueble objeto sí. La métrica ingenua nos ha engañado: la escala numérica de la dimensión del lote desplazó a la dimensión categórica de la vista, que en realidad importa más para el precio.

Una métrica de distancia defendible escalaría cada eje por su precio implícito — a grandes rasgos, por cuántos dólares vale una unidad de variación en ese eje. Una vez que los ejes se escalan por el precio implícito, la vista (que vale ~\$100,000) ya no queda dominada por el lote (~\$25 por pie cuadrado). El comparable A, que coincide con el inmueble objeto en la vista, emerge entonces como el comparable primario apropiado.

Esto es precisamente lo que aporta la maquinaria de precios implícitos desarrollada en las entradas siguientes: un escalamiento con principios del espacio de características que convierte la “cercanía” de una conjetura geométrica en un cálculo derivado del mercado.

Referencias.

- 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.

Referencias cruzadas: bien heterogéneo; función de precios hedónicos; precio implícito.

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (002-characteristics-space.pl) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 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).

```

Función de precios hedónicos

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 1 de agosto de 2026

Versión 1.0.0

DOI: [10.67330/9x1a0489](https://doi.org/10.67330/9x1a0489)

Definición formal. La función de precios hedónicos es la aplicación de equilibrio que lleva los puntos del espacio de características de un bien heterogéneo a los precios de transacción, formada como la envolvente de las funciones de puja de los compradores y las funciones de oferta de los vendedores en un mercado competitivo.

Planteamiento intuitivo. Tenemos ya dos ingredientes: un bien heterogéneo (un conjunto de características) y un espacio de características (un espacio vectorial en el que cada conjunto es un punto). La función de precios hedónicos los combina. Es la función $p(\mathbf{z})$ que asigna un precio a cada punto $\mathbf{z}$ del espacio de características, diciéndonos qué paga el mercado por un conjunto con esa combinación particular de atributos.

La idea clave de Rosen (1974) es que $p(\mathbf{z})$ no la elige ningún comprador ni vendedor individual. Emerge del equilibrio de mercado. Cada comprador tiene una *función de puja* que expresa el máximo que pagaría por cualquier conjunto, dadas sus preferencias y su ingreso. Cada vendedor tiene una *función de oferta* que expresa el mínimo que aceptaría, dados sus costos y alternativas. La función de precios hedónicos es la envolvente de esas pujas y ofertas — la frontera a lo largo de la cual compradores y vendedores efectivamente transan.

Para los bienes inmuebles, $p(\mathbf{z})$ es la respuesta a la pregunta: si se especificara un conjunto completo de características, ¿qué precio produciría el mercado para él? La respuesta no es arbitraria. Refleja las preferencias agregadas de todos los compradores del mercado y los costos y alternativas agregados de todos los vendedores, mediados por el emparejamiento de equilibrio entre compradores y conjuntos.

La función de precios hedónicos es el objeto teórico central del análisis hedónico. Los precios implícitos son sus derivadas parciales. La regresión hedónica es una estimación empírica de ella. La defendibilidad de cualquier ajuste de comparación de ventas se reduce en última instancia a una afirmación sobre esta función.

Dónde lo encuentran los tasadores. Los tasadores razonan continuamente sobre $p(\mathbf{z})$, de nuevo por lo general sin nombrarla. Varias actividades de la práctica son manipulaciones implícitas de la función de precios hedónicos:

Ajustes en el enfoque de comparación de ventas. Cuando un tasador ajusta el comparable B al alza en \$100,000 para reflejar que el inmueble objeto tiene vista y el comparable B no, el tasador está afirmando algo sobre $p(\mathbf{z})$: a saber, que manteniendo fijas todas las demás características, la función aumenta en \$100,000 cuando la característica de vista pasa de 0 a 1. Es una afirmación sobre la pendiente local de p en la dimensión de la vista.

Conciliación entre enfoques. El enfoque del costo construye una indicación de valor desde abajo, sumando los costos de los componentes de un conjunto replicado. El enfoque de ingresos deriva el valor de los flujos de efectivo, que en equilibrio reflejan el precio de conjuntos que producen flujos similares. El enfoque de comparación de ventas lee el precio directamente de transacciones

observadas de conjuntos similares. En un mercado que funciona bien, los tres enfoques deberían converger en el mismo punto de $p(\mathbf{z})$. Cuando divergen, el tasador debe juzgar qué enfoque ha medido la función con mayor fiabilidad para la posición del inmueble objeto en el espacio de características.

Análisis de mercado y ajustes por tiempo. La función de precios hedónicos se desplaza con el tiempo a medida que cambian la oferta, la demanda, el financiamiento y las expectativas. Una tasa de apreciación anual del 6 por ciento es una afirmación sobre cómo se ha desplazado $p(\mathbf{z})$ en su conjunto; el hallazgo de que los inmuebles con prima por vista se han apreciado más rápido que los que carecen de vista es una afirmación de que la función se ha desplazado de manera desigual a lo largo del espacio de características.

Por qué importa para la defendibilidad. Nombrar explícitamente la función de precios hedónicos aclara varias cuestiones fundacionales de defendibilidad:

¿Existe la función? El marco de Rosen supone un mercado competitivo con suficientes transacciones como para que una función de equilibrio tenga sentido. En mercados delgados, en mercados que cambian rápidamente o en submercados con grupos de compradores idiosincrásicos, la función puede estar mal definida o ser inestable. Un tasador que trabaje en un mercado así debería reconocer que la maquinaria inferencial opera sobre un fundamento teórico frágil.

¿Es la función la misma en todas partes? El marco de Rosen produce una función por segmento de mercado, no una función global. Pacífica y San Francisco son mercados distintos con funciones hedónicas distintas; las viviendas unifamiliares y los condominios son submercados distintos incluso dentro de una misma ciudad. Un tasador que confunde segmentos está asumiendo implícitamente que una sola función abarca ambos, lo cual suele ser falso.

¿Qué forma funcional adopta p ? La función no tiene por qué ser lineal, log-lineal ni de ninguna otra forma específica. Rosen demostró únicamente que p es la envolvente de pujas y ofertas; no especificó su forma. Empíricamente, p exhibe no linealidades (el valor marginal de una recámara adicional suele no ser constante en la GLA), interacciones (la prima por vista depende de la orientación del lote) y umbrales (una cocina por debajo de cierto nivel de calidad puede atraer un descuento desproporcionado respecto de su puntuación de estado). Una estimación defendible requiere o bien una forma funcional sustentable a partir del mercado, o bien un estimador flexible capaz de recuperar la forma a partir de los datos.

Ejemplo de tasación resuelto. Para el vecindario de Pacífica en estudio, trabajaremos con una forma funcional aditiva simple que la próxima entrada ejercitará con mayor detalle:

$$p(\mathbf{z}) = \beta_0 + \beta_{\text{GLA}} z_{\text{GLA}} + \beta_{\text{lot}} z_{\text{lot}} + \beta_{\text{view}} z_{\text{view}} + \beta_{\text{cond}} z_{\text{cond}}.$$

Esta forma incorpora dos supuestos fuertes: que la contribución de cada característica es lineal en su nivel, y que las características no interactúan. Ambos supuestos son verificables y se violan con frecuencia; los adoptamos aquí como punto de partida.

Para los ocho comparables del ejemplo continuo de este número, nuestro análisis en las próximas entradas estimará valores de β que producen una función semejante a:

Parámetro	Valor estimado
β_0 (intercepto)	\$80,777
β_{GLA} (por sf)	\$575
β_{lot} (por sf)	\$22.70
β_{view} (sí/no)	\$86,179
β_{cond} (por escalón)	\$49,824

Sustituida en la forma aditiva, esta función predice un precio para cualquier punto del espacio de características de cuatro dimensiones. Sustituir las coordenadas del comparable A produce un precio predicho cercano a su valor observado de \$1,425,000; sustituir el inmueble objeto hipotético en (1700, 7500, 1, 2) produce una indicación de valor para ese inmueble.

El punto de la entrada no son estos números específicos, sino el movimiento conceptual: *la función existe, en principio, antes de que la estimemos*. La función es una propiedad del mercado. Nuestra estimación es un intento de recuperarla. La recuperación puede ser parcial, sesgada o ruidosa, y las tres entradas siguientes tratan de cómo hacer esa recuperación de manera responsable.

Referencias.

- 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.

Referencias cruzadas: bien heterogéneo; espacio de características; precio implícito; regresión hedónica.

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (003-hedonic-price-function.pl) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 003-hedonic-price-function.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 003
%% Title: Hedonic Price Function
%% Author: Bert Craytor
%% Concept: hedonic_price_function
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(hedonic_price_function, "Hedonic Price Function", economic_theory,
%     "The hedonic price function is the equilibrium mapping from
%     points in the characteristics space of a heterogeneous good to
%     transaction prices, formed as the envelope of buyer bid
%     functions and seller offer functions in a competitive market.",
%     published, 'dictionary-2026.1').
%
% term(bid_function, "Bid Function", economic_theory, ...).
% term(offer_function, "Offer Function", economic_theory, ...).
% term(sales_comparison_approach, ...).
% term(cost_approach, ...).
```

```

% term(income_approach, ...).
%
% formalized_by(hedonic_price_function, rosen_1974).
% depends_on(hedonic_price_function, heterogeneous_good).
% depends_on(hedonic_price_function, characteristics_space).
% -----

:- use_terms([ hedonic_price_function, heterogeneous_good,
               characteristics_space, implicit_price, hedonic_regression,
               bid_function, offer_function,
               sales_comparison_approach, cost_approach, income_approach,
               adjustment, reconciliation ]).
:- use_relations([related/2, depends_on/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_price_function).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.003').

% =====
% LOCAL VOCABULARY
% =====

local_term(equilibrium).
local_term(envelope).
local_term(functional_form).
local_term(market_segment).
local_term(linearity).
local_term(separability).
local_term(time_adjustment).
local_term(coefficient).
local_term(intercept).

% =====
% FORMAL ATTRIBUTION
% =====

attributed_formalization(hedonic_price_function, rosen_1974).

% =====
% CONSTRUCTION: ENVELOPE OF BID AND OFFER FUNCTIONS
% =====
%
% The Rosen (1974) construction:  $p(z)$  is the envelope of buyer bid
% functions and seller offer functions.

primitive_of_construction(bid_function, buyer).
primitive_of_construction(offer_function, seller).

constructed_as_envelope_of(hedonic_price_function,
                           [bid_function, offer_function]).

emerges_from(hedonic_price_function, market_equilibrium).

% Property that flows from the construction: the function is not
% chosen by any single market participant.

```

```

property(hedonic_price_function, not_set_by_individual_participant).
property(hedonic_price_function, equilibrium_object).
property(hedonic_price_function, defined_per_market_segment).

% =====
% APPRAISER ENCOUNTERS | IMPLICIT MANIPULATIONS OF p(z)
% =====

implicit_manipulation_of_hedonic(
    adjustment,
    sales_comparison_approach,
    "Asserts a local slope of p in the relevant characteristic dimension.").
implicit_manipulation_of_hedonic(
    reconciliation,
    multi_approach_valuation,
    "Three approaches should converge on the same point of p; divergence requires editorial
    ↪ judgment.").
implicit_manipulation_of_hedonic(
    time_adjustment,
    market_analysis,
    "p shifts over time; appreciation rates are statements about that shift.").

% =====
% DEFENSIBILITY QUESTIONS
% =====
%
% The article names three questions the hedonic-price-function framing
% sharpens.

defensibility_question(function_existence,
    "In thin or rapidly-changing markets, the equilibrium function may be poorly defined; the
    ↪ inferential machinery operates on a shaky foundation.").
defensibility_question(segment_specificity,
    "One function per market segment, not one function globally. Conflating segments is usually
    ↪ wrong.").
defensibility_question(functional_form,
    "Rosen showed only that p is an envelope, not its shape. Empirically nonlinear with
    ↪ interactions and thresholds; estimation requires defending the form.").

% =====
% WORKED EXAMPLE: ADDITIVE LINEAR SPECIFICATION
% =====
%
% For the Pacifica neighborhood, the article adopts the additive linear
% specification:
%  $p(z) = \beta_0 + \beta_{\text{gla}}z_{\text{gla}} + \beta_{\text{lot}}z_{\text{lot}} + \beta_{\text{view}}z_{\text{view}} + \beta_{\text{cond}}z_{\text{cond}}$ 
%
% The next entry (004) takes the coefficients as implicit prices; entry
% 005 estimates them by hedonic regression.

:- consult('../..kb/pacifica_comps').

specification_form(additive_linear).
specification_assumption(linearity_in_levels,
    "Contribution of each characteristic is linear in its level.").
specification_assumption(no_interactions,
    "Characteristics do not interact in the price function.").
specification_caveat(
    "Both assumptions are testable and frequently violated; adopted here as a starting point.").

```

```

% Coefficient values reported in the article (estimated in entry 005,
% asserted here so this entry's worked example can be queried).
coefficient(intercept,      80777.07).
coefficient(beta_gla,       574.93).
coefficient(beta_lot,       22.70).
coefficient(beta_view,     86179.03).
coefficient(beta_cond,     49824.33).

%! predicted_price(?Property, -Predicted) is det.
%
% The fitted additive-linear hedonic price function applied to a
% property's characteristics. Symbolic view/cond are mapped to {0,1}
% and {1,2,3} respectively.

predicted_price(P, Predicted) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond),
    coefficient(intercept, B0),
    coefficient(beta_gla, Bg),
    coefficient(beta_lot, B1),
    coefficient(beta_view, Bv),
    coefficient(beta_cond, Bc),
    Predicted is B0 + Bg * GLA + B1 * 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).

```

Precio implícito

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 1 de agosto de 2026

Versión 1.0.0

DOI: [10.67330/md6wke17](https://doi.org/10.67330/md6wke17)

Definición formal. El precio implícito de una característica es la derivada parcial de la función de precios hedónicos con respecto a esa característica, evaluada en un punto del espacio de características; representa el precio marginal de mercado de una unidad incremental de la característica, manteniendo fijas todas las demás características.

Planteamiento intuitivo. La función de precios hedónicos $p(\mathbf{z})$ asigna un precio a cada conjunto. La mayoría de las preguntas prácticas en la tasación inmobiliaria, sin embargo, no versan sobre el precio de un conjunto completo sino sobre el precio de un *componente* específico del conjunto. ¿Cuánto vale la vista? ¿Cuánto añaden cien pies cuadrados adicionales? ¿Cuánto contribuye la diferencia entre un estado de conservación “bueno” y uno “excelente”?

El precio implícito responde estas preguntas. El término se origina en Rosen (1974), quien mostró que los precios observados de productos diferenciados junto con sus características “definen un conjunto de precios implícitos o ‘hedónicos’”; Lancaster (1966) aportó la visión subyacente de los bienes como conjuntos de características. Para cada característica z_i , el precio implícito es la tasa a la que cambia $p(\mathbf{z})$ cuando z_i varía mientras todas las demás características se mantienen fijas:

$$\pi_i(\mathbf{z}) = \frac{\partial p(\mathbf{z})}{\partial z_i}.$$

El precio implícito tiene unidades de dólares por unidad de la característica — dólares por pie cuadrado de GLA, dólares por pie cuadrado de lote, dólares por escalón del estado de conservación. Para una característica binaria como la vista, el precio implícito es la prima completa pagada por pasar de sin-vista (0) a con-vista (1).

La palabra “implícito” captura el hecho central: estos precios *no* están publicados. Nadie ofrece vender “un pie cuadrado de GLA” por \$575. El precio está implícito en los precios de equilibrio de los conjuntos que *sí* se publican, y el aparato hedónico existe para extraerlo.

Los precios implícitos son el producto del marco hedónico relevante para la tasación. El ajuste en un cuadro de comparación de ventas es, en principio, una estimación de un precio implícito multiplicada por una diferencia de características. Cada ajuste es una apuesta sobre el valor de π_i en el punto relevante del espacio de características.

Dónde lo encuentran los tasadores. El cuadro de ajustes de comparación de ventas es una tabla de aplicaciones de precios implícitos, aun cuando los tasadores no los nombren así. Un renglón del cuadro que ajusta un comparable en “+\$10,000 por 200 sf menos de GLA que el inmueble objeto” está afirmando que el precio implícito de la GLA, en la posición relevante del espacio de características, es de \$50 por pie cuadrado. Un renglón que ajusta “\$100,000 por vista inferior” está afirmando un precio implícito de \$100,000 para la característica binaria de vista. Toda la defendibilidad del enfoque de comparación de ventas descansa en si estos precios implícitos pueden sustentarse a partir del mercado. En la práctica, sin embargo, los tasadores estiman los ajustes

directamente — a partir de pares comparables, de la experiencia o de la práctica de sus colegas — sin calcular nunca las contribuciones de valor subyacentes; el cuadro tradicional registra solamente los ajustes mismos. La lectura en términos de precios implícitos es una interpretación de lo que esos ajustes afirman, no una descripción de cómo se derivan.

Esta denominación también aclara por qué el análisis de pares comparables es la técnica fundamental para derivar ajustes. Dos comparables que difieren en exactamente una característica dan una estimación directa del precio implícito de esa característica: la diferencia de precio dividida por la diferencia en la característica. La fiabilidad del análisis de pares comparables depende de si los dos comparables están realmente lo bastante cerca en el espacio de características como para que el precio implícito recuperado de ellos generalice al inmueble objeto.

Tres propiedades de los precios implícitos que importan para la práctica:

Son locales. El precio implícito de la GLA en el punto de 1,650 sf no es necesariamente el mismo que en el punto de 2,500 sf. La función de precios hedónicos es en general no lineal, de modo que sus derivadas parciales varían con la posición en el espacio de características. Un ajuste derivado de un análisis de pares comparables a 1,500 sf no se aplica necesariamente a un inmueble objeto de 3,000 sf.

Son condicionales. El precio implícito de la vista en un submercado frente a la playa no es el mismo que en un submercado interior. La derivada parcial se toma manteniendo fijas las demás características, pero el valor de esa derivada depende de *dónde* se mantienen fijas las demás características.

Dependen del mercado. Los precios implícitos son precios de equilibrio, de modo que reflejan las preferencias del grupo de compradores en el momento de las transacciones. Un cambio del mercado puede alterar los precios implícitos aunque las características mismas no cambien. Los ajustes por tiempo son admisiones implícitas de que los precios implícitos de ayer pueden no ser los de hoy.

Por qué importa para la defendibilidad. El planteamiento en términos de precios implícitos expone cuestiones de defendibilidad de los ajustes que el cuadro informal puede ocultar:

¿En qué parte del espacio de características se encuentra usted? Un precio implícito es defendible solo si se estima cerca de la posición del inmueble objeto en el espacio de características. Una prima por vista derivada de comparables del interior y aplicada a un inmueble objeto frente a la playa está extrapolando a través de una región del espacio de características donde el comportamiento de la función es desconocido.

¿Cuál es la geometría local? Si dos características interactúan — por ejemplo, si la vista vale más en un lote más grande — entonces tratarlas como aditivas por separado es una especificación errónea. El precio implícito de la vista depende del tamaño del lote, y un ajuste que ignore la interacción subestima la estructura real de la función.

¿Cómo se recuperó el precio implícito? Un análisis de pares comparables, un coeficiente de regresión, una regla empírica del sector y el “juicio profesional” apuntan todos a la misma cantidad — el precio implícito — pero con un estatus epistémico muy distinto. La defendibilidad del tasador depende de declarar qué fuente produjo el precio implícito usado en cada ajuste.

Ejemplo de tasación resuelto. Para el ejemplo continuo de Pacífica con la especificación aditiva adoptada en la entrada anterior, los precios implícitos son simplemente los coeficientes de la función lineal:

$$\pi_{\text{GLA}} = \beta_{\text{GLA}}, \quad \pi_{\text{lote}} = \beta_{\text{lote}}, \quad \pi_{\text{vista}} = \beta_{\text{vista}}, \quad \pi_{\text{cond}} = \beta_{\text{cond}}.$$

Con los coeficientes estimados de la entrada 003:

Característica	Precio implícito	Unidades
GLA	\$575	por sf
Lote	\$22.70	por sf
Vista	\$86,179	por sí/no
Estado	\$49,824	por escalón

Estos precios implícitos implican un análisis de pares comparables. Volviendo a la diferencia entre los comparables A y B de la entrada 001:

- El comparable A tiene vista ($\Delta z_{\text{view}} = +1$); contribución +\$86,179.
- El comparable A tiene un lote 1,800 sf más grande ($\Delta z_{\text{lot}} = +1,800$); contribución $1,800 \times \$22.70 = +\$40,860$.
- Todas las demás características son idénticas.
- Diferencia de precio predicha: $\$86,179 + \$40,860 = \$127,039$.

La diferencia de precio observada fue de \$145,000. La descomposición por precios implícitos explica alrededor de \$127,000; los \$18,000 restantes reflejan la variabilidad muestral del pequeño conjunto de datos (y, más en general, el tipo de factores no observados que abordará la entrada sobre variables latentes).

Vale la pena señalar dos puntos operativos. Primero, el supuesto de linealidad está haciendo un trabajo real: bajo una $p(\mathbf{z})$ no lineal, la misma diferencia física de lote podría valer más en un lote pequeño que en uno grande, y la descomposición aditiva anterior estaría sesgada. Segundo, el precio implícito de \$22.70 por pie cuadrado de lote es el valor marginal de un pie cuadrado incremental, no el valor total del lote. La contribución total de un lote de 9,000 sf al precio del conjunto es $9,000 \times \$22.70 = \$204,300$, pero añadir un pie cuadrado más eleva el precio del conjunto en solo \$22.70.

Esta distinción — marginal frente a total — es la fuente de una confusión considerable en el lenguaje cotidiano de la tasación inmobiliaria y es una de las ganancias conceptuales más importantes de nombrar explícitamente el precio implícito.

Referencias.

- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Freeman, A. M., III (1979). Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues. *Scandinavian Journal of Economics*, 81(2), 154–173.
- Taylor, L. O. (2003). The Hedonic Method. In *A Primer on Nonmarket Valuation* (pp. 331–393). Kluwer Academic Publishers.
- Palmquist, R. B. (2005). Property Value Models. In *Handbook of Environmental Economics* (Vol. 2, pp. 763–819). North-Holland.
- Triplett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

Referencias cruzadas: bien heterogéneo; espacio de características; función de precios hedónicos; regresión hedónica.

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (004-implicit-price.pl) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 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).

```

Regresión hedónica

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 1 de agosto de 2026

Versión 1.0.0

DOI: [10.67330/ams1t139](https://doi.org/10.67330/ams1t139)

Definición formal. La regresión hedónica es la estimación de la función de precios hedónicos a partir de transacciones observadas en un mercado de bienes heterogéneos, que produce estimaciones empíricas de los precios implícitos de las características que entran en el modelo.

Planteamiento intuitivo. Las entradas anteriores describieron la función de precios hedónicos y sus precios implícitos como objetos teóricos. La función existe en principio, pero en la práctica observamos únicamente un conjunto finito de transacciones: conjuntos de características que se comerciaron y los precios a los que se comerciaron. La regresión hedónica es el puente empírico entre ambos.

El planteamiento es sencillo. Cada transacción aporta un par $(\mathbf{z}_i, p_i)$: el vector de características del conjunto transado y su precio de venta. La colección de estos pares es una muestra de la función de precios hedónicos (no observada). La regresión ajusta una forma funcional a esa muestra, estimando los parámetros que indexan la función. Una vez estimados los parámetros, las derivadas parciales de la función ajustada producen estimaciones de los precios implícitos.

Hay que tomar tres decisiones de modelado antes de que la regresión pueda ejecutarse:

1. *¿Qué características entran en el modelo?* La elección determina las dimensiones de la función estimada. Omitir una característica relevante hace que su efecto se filtre en los coeficientes de las características incluidas correlacionadas (sesgo por variables omitidas).
2. *¿Qué forma funcional?* Lineal, log-lineal, log-log, polinómica, con expansión de bases, semi-paramétrica, de aprendizaje automático. La elección restringe qué forma puede adoptar la función estimada y qué patrones de precios implícitos puede recuperar.
3. *¿Qué estimador?* MCO, mínimos cuadrados generalizados, M-estimadores robustos, regresión cuantílica, estimadores de econometría espacial, regresión penalizada. La elección determina qué optimiza el estimador y qué supuestos sustentan su inferencia.

Cada elección es metodológicamente sustantiva y cada una acarrea consecuencias de defendibilidad.

Dónde lo encuentran los tasadores. La regresión hedónica es el fundamento de la valuación masiva y el motor de la mayoría de los modelos de valuación automatizada (AVM). Cuando un asesor fiscal de condado revalúa 100,000 parcelas cada año, una regresión hedónica sobre ventas recientes está haciendo el trabajo. Cuando Zillow o Redfin producen una estimación automatizada, la está produciendo un modelo de estilo hedónico (a menudo un ensamble de árboles en lugar de MCO, pero conceptualmente un estimador hedónico de todos modos). Cuando el AVM de mercado secundario de un prestamista aprueba o rechaza un inmueble para un refinanciamiento, está involucrada la misma maquinaria.

La tasación de inmuebles individuales también recurre cada vez más a la regresión hedónica, incluso cuando la indicación final de valor proviene de la comparación de ventas. Los precios implícitos

estimados a partir de una regresión sobre una muestra del vecindario pueden sustentar las magnitudes de los ajustes de manera mucho más defendible que la intuición del tasador o los porcentajes de regla empírica. Los coeficientes de la regresión se convierten en evidencia, no en opinión, para los ajustes aplicados en el cuadro.

Por qué importa para la defendibilidad. Las estimaciones basadas en regresión tienen un perfil de defendibilidad distinto del análisis de pares comparables. Las ventajas: una regresión usa todas las ventas disponibles en lugar de dos a la vez; puede aislar el efecto de una característica mientras controla simultáneamente muchas otras; y produce errores estándar que cuantifican la incertidumbre de cada precio implícito. Las contrapartidas: una regresión impone una forma funcional que puede no coincidir con la verdadera función de precios; requiere más datos que un análisis de pares comparables; y sus resultados pueden ser sensibles a la elección de la muestra y de la especificación.

Varias cuestiones de defendibilidad se repiten en la tasación basada en regresión:

¿Es apropiada la muestra? La regresión estima la función hedónica tal como opera en el segmento de mercado y el periodo de la muestra. Una regresión sobre ventas del interior aplicada a un inmueble objeto frente a la playa es un fallo de defendibilidad por más limpias que luzcan las estadísticas.

¿Es apropiada la especificación? Una especificación lineal es un supuesto fuerte. Si la función verdadera exhibe rendimientos decrecientes de la GLA por encima de algún umbral, el ajuste lineal estimará mal los precios implícitos en ese rango. Los diagnósticos, los gráficos de residuos y las especificaciones alternativas forman parte del flujo de trabajo defendible.

¿Qué significa el error estándar? Un error estándar captura la variabilidad muestral bajo la especificación supuesta. No captura la especificación errónea del modelo, el sesgo por variables omitidas ni la selección de la muestra. Un error estándar pequeño sobre una especificación equivocada es más peligroso que un error estándar grande sobre una correcta.

¿Triangula el resultado con los demás enfoques? Un precio implícito basado en regresión y defendible debería ser coherente con lo que sugiere el análisis de pares comparables, con lo que implican los costos de los componentes del enfoque del costo y con lo que creen los tasadores que conocen el mercado. Un coeficiente de regresión en desacuerdo con los tres es una invitación a investigar, no a publicar.

Ejemplo de tasación resuelto. Podemos ahora ejercitar toda la maquinaria de la regresión hedónica sobre los ocho comparables de Pacifica. El conjunto de datos está disponible en [shared/data/pacifica.co](#)

Código en R:

```
comps <- read.csv("shared/data/pacifica_comps.csv")
fit <- lm(price ~ gla + lot + view + cond, data = comps)
summary(fit)
```

Salida:

```
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
```

Interpretación. El modelo ajustado estima la función de precios hedónicos como:

$$\hat{p}(\mathbf{z}) = 80,777 + 575 z_{\text{GLA}} + 22.70 z_{\text{lot}} + 86,179 z_{\text{view}} + 49,824 z_{\text{cond}}.$$

Cada coeficiente es el precio implícito estimado de la característica correspondiente. La GLA contribuye \$575 por pie cuadrado (con el error estándar más estrecho y la significancia más fuerte); la vista contribuye alrededor de \$86,000 a un inmueble de Pacífica; una unidad adicional de estado de conservación (p. ej., pasar de “bueno” a “excelente”) vale aproximadamente \$50,000.

Varias salvedades son inmediatamente visibles:

La muestra es diminuta. Ocho observaciones y cuatro predictores dejan tres grados de libertad residuales. Los errores estándar son correspondientemente amplios, y los valores p del lote, la vista y el estado no alcanzan el umbral convencional de 0.05 a pesar de su plausibilidad económica. Se trata de un artefacto de muestra pequeña, no de evidencia de que estas características no importen. Con 80 comparables en lugar de 8, los valores p se reducirían drásticamente.

El ajuste global es fuerte. Un R^2 ajustado = 0.93 y un valor p del modelo de 0.012 indican que las características incluidas explican conjuntamente la mayor parte de la variación de precios en la muestra. La combinación de un ajuste conjunto fuerte con una significancia individual débil es una señal de manual de multicolinealidad, que aquí es estructural: la GLA, el lote y el estado covarían en la muestra.

Un residuo sobresale. El residuo máximo es de +\$46,455 y corresponde al comparable H. Los otros siete residuos caen todos dentro de una banda de $\pm \$18,000$. Algo del comparable H no queda capturado por la función ajustada — el modelo está subprediciendo sistemáticamente su precio por una cantidad casi tres veces mayor que el siguiente residuo más grande.

Este residuo atípico no es ruido. Es la señal de que una característica no observada del comparable H está contribuyendo a su precio de una manera que las variables incluidas no pueden recuperar. La entrada sobre la variable latente aborda exactamente esta cuestión.

Validación frente al razonamiento de pares comparables. El análisis de pares comparables de la entrada 004 predijo una diferencia de \$127,039 entre los comparables A y B a partir de la vista y el lote únicamente. Los precios implícitos de la regresión son idénticos (puesto que son la fuente de ese cálculo). La regresión generaliza la lógica de los pares comparables a los ocho comparables simultáneamente, pero no produce nada fundamentalmente nuevo con este tamaño de muestra; confirma el análisis por pares a la vez que informa errores estándar que el razonamiento por pares no puede ofrecer.

Referencias.

- 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.

Referencias cruzadas: bien heterogéneo; espacio de características; función de precios hedónicos; precio implícito; variable latente.

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (005-hedonic-regression.pl) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 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\\pacificacomps').

% 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).

```

Variable latente

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 1 de agosto de 2026

Versión 1.0.0

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

Definición formal. Una variable latente es un atributo no observado que influye en los resultados observados, y cuyo valor debe inferirse a partir de sus efectos sobre las variables observables en lugar de medirse directamente.

Planteamiento intuitivo. Las primeras cinco entradas trataron el vector de características $\mathbf{z}$ como si pudiera medirse por completo: GLA, lote, vista, estado de conservación, etcétera, todo disponible como columnas de datos. En la práctica, ningún vector de características está jamás completo. Algunos atributos no están en los datos porque son difíciles de medir (la calidad de los acabados interiores, la fluidez de la distribución, la micro-orientación micro-locacional), porque están ausentes de los flujos de datos estándar (los registros de la MLS no capturan la calidad de la iluminación, la insonorización ni las mejoras recientes sin permiso), o porque todavía no se han reconocido como relevantes (rasgos a nivel de vecindario que resultan predecir el precio solo después de que el análisis los revela).

Estos atributos no-observados-pero-relevantes son *variables latentes*. Existen, influyen en el precio, pero no están en las columnas de nuestra regresión. Sus efectos no se desvanecen por no estar medidos; se acumulan en los residuos del modelo, donde aparecen como “ruido” desde la perspectiva de las variables incluidas pero son en realidad señal estructurada desde la perspectiva del proceso generador de datos subyacente.

Las variables latentes son el puente entre el marco hedónico idealizado y la realidad desordenada de los datos de tasación. Reconocerlas explícitamente es una precondition para razonar con honestidad sobre la incertidumbre del modelo, la estructura de los residuos y los límites de lo que las características observables pueden explicar.

Dónde lo encuentran los tasadores. Los tasadores encuentran variables latentes en tres contextos recurrentes.

Los ajustes por calidad y estado de conservación. La “calidad” y el “estado” como categorías de tasación son en sí mismos intentos de colapsar un constructo latente multidimensional en una sola escala ordinal. Una calificación de estado C3 en el UAD 3.6 pretende resumir el mantenimiento diferido de un inmueble, el desgaste de los acabados, la antigüedad de los sistemas mecánicos y la presentación — de cuatro a diez atributos subyacentes condensados en un solo número. El resumen de un solo número es un proxy de un constructo latente; el oficio del tasador incluye el juicio sobre cuán bien captura el proxy la verdad subyacente en cada caso.

El residuo sin explicar en el cuadro. Cuando tres comparables ajustados indican valor pero discrepan más de lo esperado, la discrepancia es informativa. Rara vez es toda ruido muestral; más a menudo refleja un atributo que el tasador aún no ha identificado ni ajustado. Investigar la discrepancia residual — “¿por qué este comparable indica \$50,000 por encima de los demás?” — es la versión informal del tasador de la inferencia de variables latentes.

La mejora reciente que nadie puede ver. Un inmueble puede haber pasado por una remodelación de cocina, un reemplazo de techo o la construcción de una ADU con permiso que aún

no se refleja en la MLS ni en los datos del catastro. Desde la perspectiva del tasador, la mejora es observable en persona pero no en los datos. Desde la perspectiva de la regresión, la mejora es latente. El tasador debe conciliar la observación en persona con la indicación basada en datos, y ambas deberían converger en la verdadera contribución del atributo latente.

Por qué importa para la defendibilidad. Las variables latentes introducen varias consideraciones de defendibilidad que el modelo lineal aditivo de la entrada 005 no aborda directamente.

Sesgo por variables omitidas. Si una variable latente está correlacionada con una variable observada, su efecto se filtra en el coeficiente de la variable observada. Si las casas de Pacífica vendidas recientemente tienden a tener cocinas remodeladas (un atributo latente) y la fecha del contrato de la venta está en los datos, entonces el coeficiente de la antigüedad de la venta (alias `sale_age` o “días desde la venta”) absorberá parte del efecto de la remodelación de cocina. La contribución estimada del precio implícito de `sale_age` queda sesgada — refleja `sale_age` más una porción del estado no medido de la remodelación de cocina. Una tasación defendible reconoce que los coeficientes de las variables observadas están condicionados a lo que falta en el modelo.

Interpretación de los residuos. La teoría estándar de la regresión trata los residuos como ruido gaussiano de media cero. La visión de variables latentes trata los residuos como una mezcla de ruido y señal: una parte es variación muestral, otra parte es el efecto acumulado de los atributos no observados. Identificar cuál es cuál es metodológicamente sustantivo. Un residuo que se correlaciona con la ubicación, con la fecha de venta, con la identidad del tasador o con cualquier otro observable sugiere que la variable latente correspondiente puede recuperarse parcialmente.

Selección del modelo. Añadir más variables observadas es el remedio obvio para el sesgo por variables omitidas, pero las variables latentes que más importan suelen ser las más difíciles de añadir al modelo (el grano fino del estado, la calidad de las mejoras recientes, los microefectos de vecindario). La respuesta defendible a veces consiste en reconocer que el modelo nunca las capturará y asignar la varianza sin explicar al residuo, en lugar de forzar el ajuste de variables adicionales.

Ejemplo de tasación resuelto. La regresión de la entrada 005 produjo un residuo inusual para el comparable H: el modelo subpredijo el precio del comparable H en \$46,455 mientras que los otros siete residuos cayeron dentro de una banda de $\pm \$18,000$. Las características observadas del comparable H — GLA de 1,600, lote de 7,300, sin vista, estado “bueno” — lo sitúan firmemente en el centro de la muestra; nada de su perfil observable predice un precio inusualmente alto. Y sin embargo se vendió por \$46,455 más de lo que el modelo dice que debería.

La hipótesis natural es que el comparable H tiene un atributo latente que contribuye a su precio pero no aparece en los datos. En la construcción del conjunto de datos de este número, ese es exactamente el caso: el comparable H pasó por una remodelación reciente de cocina que añade aproximadamente \$60,000 al valor de su conjunto. La remodelación no estaba en la lista de variables de la regresión, de modo que la contribución se acumuló en el residuo y no en un coeficiente. El residuo estimado (\$46,455) es ligeramente menor que la verdadera contribución latente (\$60,000) porque la regresión reparte pequeñas porciones del efecto latente entre los demás coeficientes, sesgándolos al alza en las dimensiones correlacionadas.

Se siguen tres consecuencias operativas:

Trate el residuo atípico como evidencia, no como error. La primera respuesta apropiada al residuo del comparable H no es descartarlo como atípico, sino investigar qué tiene el inmueble que explique la subpredicción. En este caso, una inspección en persona o una búsqueda en los registros de permisos revelaría la remodelación. El fallo del modelo tiene valor diagnóstico.

Reconozca que los demás coeficientes están ligeramente sesgados. Los coeficientes de la vista, el estado y el lote han absorbido pequeñas fracciones del efecto de la remodelación a través de correlaciones con las características observadas del comparable H. El sesgo es pequeño con $n = 8$

y se reduciría aún más con más comparables y más covariables, pero no es cero.

Los residuos en su conjunto llevan estructura recuperable. Esta es la afirmación sustantiva de la metodología del Residual Constraint Approach (RCA) desarrollada por el autor (Craytor, 2025): en los datos de tasación residencial, los residuos sistemáticamente no son ruido gaussiano iid. Reflejan la contribución estructurada de los atributos no observados — el grano fino del estado, el grano fino de la calidad, la utilidad funcional y otros constructos latentes — que puede recuperarse parcialmente tratando el residuo mismo como una señal de variables latentes que debe modelarse en lugar de descartarse.

El tratamiento que el RCA da al estado y a la calidad hace operativa esta estructura de dos niveles. Las calificaciones ordinales del tasador — los proxies de grano grueso discutidos arriba, que pueden etiquetarse condición-1 (grano grueso del estado) y calidad-1 (grano grueso de la calidad) — son lo bastante burdas como para que bien pudieran tratarse ellas mismas como variables latentes, pero aun así pueden ingresarse como regresores para obtener contribuciones aproximadas de precios implícitos. El protocolo traslada entonces esas contribuciones estimadas fuera de los atributos medidos en la sección de regresión del cuadro de ventas y de vuelta al residuo, desde el cual se reasignan a las variables latentes de grano fino condición-2 (grano fino del estado) y calidad-2 (grano fino de la calidad) y se refinan contra evidencia adicional, como los comentarios públicos de la MLS. La calificación de grano grueso proporciona una asignación de valor inicial; la variable latente de grano fino lleva la estructura recuperable. Al final, la contribución de valor de grano fino reemplaza a la contribución de valor de grano grueso.

La ilustración con un solo comparable presentada aquí es el caso más simple de la intuición del RCA. Con más datos y un modelado más rico, la estructura del residuo puede particionarse en factores latentes reconocibles, cada uno de los cuales contribuye al precio de mercado del conjunto de maneras que la especificación hedónica estándar no puede alcanzar. Las variables latentes no son una molestia que deba controlarse y eliminarse; son la próxima frontera del modelado de valuación defendible.

Referencias.

- 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)

Referencias cruzadas: espacio de características; función de precios hedónicos; precio implícito; regresión hedónica.

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (`006-latent-variable.pl`) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 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).
```

Resoluciones judiciales recientes

Casos judiciales recientes — Estados Unidos

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Recent Court Decisions • The Valuation Engineer

William Bert Craytor • Vol. 1, Número 1 • Verano 2026

DOI: 10.67330/fz34f711

Resúmenes de casos por jurisdicción

California

The Retail Property Trust v. Orange County Assessment Appeals Board No. 1. California Court of Appeal (Tribunal de Apelación de California), resuelto en abril de 2026.

El Court of Appeal abordó la disponibilidad de la desgravación del impuesto predial prevista en la sección 170 del California Revenue and Taxation Code (Código de Rentas e Impuestos de California) respecto de las restricciones de acceso impuestas por órdenes gubernamentales dictadas durante la pandemia de COVID-19. El centro comercial demandante sostuvo que el acceso restringido durante la pandemia constituía un “daño” (*damage*) al inmueble en el sentido de la sección 170, lo que le daría derecho a la desgravación del avalúo fiscal.

Implicación para los tasadores. La interpretación que la opinión hace del “daño” conforme a la sección 170 condicionará las futuras reclamaciones de desgravación basadas en disminuciones de la utilidad del inmueble causadas externamente, ya provengan de órdenes de salud pública, de proyectos de infraestructura o de otras actuaciones gubernamentales que afecten al acceso. Los tasadores a quienes se pida dictaminar sobre el efecto de las restricciones de acceso en el valor de un inmueble comercial deberían revisar el razonamiento del tribunal acerca de si la utilidad restringida — a diferencia del daño físico — constituye un daño indemnizable a efectos de desgravación fiscal.

Colorado

Cheroutes v. Rocket Mortgage & Solidifi. U.S. District Court, D. Colorado (Tribunal Federal de Distrito para el Distrito de Colorado); en trámite, con actualizaciones de estado en 2026.

La demanda subyacente alega sesgo en la tasación al amparo de la Fair Housing Act (Ley de Vivienda Justa) en relación con una tasación realizada en 2021 de un dúplex de Denver valorado en aproximadamente \$640,000 — más de \$220,000 por debajo del valor tasado el año anterior. Un juez federal denegó previamente la moción de sobreseimiento de los demandados, al sostener que Rocket Mortgage y la empresa de gestión de tasaciones Solidifi deben defenderse de las reclamaciones en lugar de declinar toda responsabilidad sobre la base de que los prestamistas y las AMC no pueden responder por la conducta de tasadores independientes contratados por honorarios. El caso se encuentra en fase de prueba (*discovery*) en 2026.

Implicación para los tasadores. El caso es el principal vehículo activo para poner a prueba la responsabilidad de las AMC y de los prestamistas conforme a la Fair Housing Act por la conducta de los tasadores contratados por honorarios. Si el tribunal termina declarando responsables en el fondo a la AMC y al prestamista, se ampliaría sustancialmente el universo de partes expuestas a

reclamaciones de vivienda justa derivadas de tasaciones, y ello se trasladaría probablemente a los contratos de encargo entre AMC y tasador.

Florida

Acción colectiva en Florida que impugna las comisiones de las AMC cobradas a prestatarios residenciales. Presentada el 2 de enero de 2026.

La demanda impugna las comisiones de las empresas de gestión de tasaciones cobradas a los prestatarios hipotecarios residenciales, y alega que el margen entre el honorario pagado al tasador y la comisión cobrada al prestatario en el cierre excede los límites establecidos por la legislación federal y se divulga de forma insuficiente. El caso se sitúa en la intersección entre el requisito de honorarios “habituales y razonables” de la Dodd-Frank Act (12 U.S.C. § 1639e) y los regímenes de divulgación al consumidor de la Truth in Lending Act (Ley de Veracidad en los Préstamos) y de la Real Estate Settlement Procedures Act (Ley de Procedimientos de Cierre de Operaciones Inmobiliarias).

Implicación para los tasadores. El litigio pone a prueba la estructura de la retribución de las AMC y, de certificarse como acción colectiva y resolverse en el fondo, podría reconfigurar las prácticas de divulgación de todo el sector. El resultado puede afectar indirectamente al entorno de fijación de honorarios de los tasadores que trabajan a través de AMC, al someter a escrutinio judicial el diferencial entre lo que la AMC paga y lo que se cobra al prestatario.

Georgia

Casos de servidumbres de conservación Green Rock. United States Tax Court (Tribunal Fiscal de los Estados Unidos), opiniones comentadas en *Tax Notes* el 2 de marzo de 2026.

El Tax Court recorrió cada uno de los elementos del test del mayor y mejor uso (*highest and best use*) al evaluar servidumbres de conservación constituidas sobre parcelas colindantes (la Green Rock Property y Green Rock South) en la provincia geológica del Piedmont, en Georgia. Los peritos de los recurrentes afirmaron que el mayor y mejor uso de los inmuebles antes de la constitución de la servidumbre era la explotación de una cantera de granito, y emplearon un análisis de flujos de caja descontados para estimar un valor de mercado previo a la servidumbre de aproximadamente \$13.75 millones frente a un valor posterior a la servidumbre de \$180,000.

El tribunal manifestó reservas sobre cada elemento del test. En cuanto a la admisibilidad jurídica, no quedó convencido de que fuera razonablemente probable una recalificación urbanística que permitiera la explotación de canteras, y señaló en particular que el cambio de clasificación de zonificación de la Green Rock Property se había obtenido poco después de solicitarse una nueva tasación de la servidumbre de conservación — una secuencia que al tribunal le resultó difícil de dar por buena. En cuanto a la posibilidad física, el tribunal consideró que las pruebas geológicas de los recurrentes eran insuficientes para acreditar que pudiera desarrollarse una cantera viable en la parcela de Harman Road, pese a la presencia general conocida de granito explotable en la provincia del Piedmont. El tribunal sostuvo además que, incluso si se dieran por supuestas la admisibilidad jurídica y la posibilidad física, ello no sería determinante sin satisfacer los restantes elementos del análisis.

Implicación para los tasadores. La opinión es un ejemplo trabajado de aplicación secuencial y disciplinada del test de cuatro elementos del mayor y mejor uso. Refuerza la línea establecida en *Ranch Springs, LLC v. Commissioner*, 164 T.C. No. 6 (2025), y *Beaverdam Creek Holdings, LLC v. Commissioner*, T.C. Memo. 2025-53, según la cual el Tax Court no aceptará afirmaciones especulativas sobre el mayor y mejor uso ni siquiera cuando la viabilidad general del uso propuesto esté

documentada en la región circundante. Los tasadores que preparen valoraciones de servidumbres deberían desarrollar pruebas específicas de la parcela para cada elemento, en lugar de apoyarse en una viabilidad regional o sustentada por analogía, y deberían ser especialmente cautelosos con los patrones temporales — como los cambios de zonificación obtenidos de forma simultánea al encargo de la tasación — que pueden suscitar escepticismo judicial.

Maryland

Adquisición del Pimlico Race Course por Baltimore. Procedimiento de expropiación forzosa (*eminent domain*), en curso 2025–2026.

La adquisición del Pimlico Race Course por la ciudad de Baltimore mediante la potestad expropiatoria plantea cuestiones de justa indemnización relativas a la relación entre el valor de la empresa en funcionamiento y el valor del inmueble en instalaciones de uso especializado. Dado que la explotación empresarial de Pimlico es funcionalmente inseparable de sus instalaciones físicas, el análisis convencional de ventas comparables de inmuebles tiene una utilidad limitada, y las valoraciones por el enfoque de ingresos dependen en gran medida de la continuidad supuesta de la actividad de carreras.

Implicación para los tasadores. El procedimiento merece seguimiento por parte de cualquier tasador al que se encargue valorar una instalación de propósito único en la que la explotación empresarial y el inmueble sean funcionalmente inseparables, como hipódromos, estudios cinematográficos, parques temáticos, puertos deportivos y determinados activos sanitarios u hoteleros. La metodología que surja de los procedimientos litigiosos influirá probablemente en la forma en que los cálculos de justa indemnización traten los incrementos por empresa en funcionamiento en futuros expedientes expropiatorios.

Reclamaciones federales contra tasadores al amparo de la Fair Housing Act

Braxton v. Bank of America. Presentada en abril de 2026.

Una propietaria de vivienda ha demandado a Bank of America y a un tasador radicado en Virginia, alegando discriminación racial con infracción de la Fair Housing Act. La demandante sostiene que ella y sus inquilinos estaban presentes durante la inspección de tasación y que la conducta del tasador evidenció animadversión racial. El caso continúa la pauta establecida en litigios anteriores sobre sesgo en la tasación, entre ellos el asunto *Connolly/Mott v. LoanDepot* en el Distrito de Maryland y el asunto *Austin v. Miller* en el condado de Marin, California, de reclamaciones individuales al amparo de la Fair Housing Act contra tasadores, AMC y prestamistas.

Implicación para los tasadores. El ritmo sostenido de presentación de demandas por sesgo en la tasación contra tasadores individuales, unido al repliegue federal en la aplicación de la doctrina del impacto dispar (*disparate impact*), indica que el campo de exposición a responsabilidad se ha desplazado: los demandantes privados y los organismos estatales de vivienda justa son ahora las principales fuentes de responsabilidad por sesgo en la tasación, más que las agencias federales. Los tasadores deberían revisar sus prácticas de documentación para asegurar que la selección de comparables, la derivación de los ajustes y la reconciliación del valor final sean defendibles de forma autónoma frente a un escrutinio posterior, y deberían ser conscientes de que las manifestaciones o la conducta durante la propia inspección son con frecuencia el fundamento fáctico de las reclamaciones de los demandantes.

Decisiones judiciales recientes — Unión Europea

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Recent Court Decisions • The Valuation Engineer

William Bert Craytor • Vol. 1, Número 1 • Verano 2026

DOI: 10.67330/r1yn1y02

Alcance y fuentes

Este estudio se ha elaborado a partir de fuentes primarias de acceso público cuando ha sido posible — señaladamente la base de datos oficial de jurisprudencia del Tribunal de Justicia de la Unión Europea (curia.europa.eu), HUDOC para el Tribunal Europeo de Derechos Humanos, las resoluciones publicadas del Bundesfinanzhof (bundesfinanzhof.de), el Tribunal Constitucional Federal alemán (bverfg.de), Légifrance para la jurisprudencia francesa y el Upper Tribunal (Lands Chamber) para el Reino Unido — y de comentarios secundarios en publicaciones profesionales y académicas. Las resoluciones dictadas en lenguas nacionales se han leído en su idioma original y se describen en inglés en la edición original; los encabezamientos de los casos conservan su grafía y sus signos diacríticos originales.

No se utilizan bases de datos jurídicas de suscripción (juris.de, beck-online, Westlaw UK, Lamy, Aranzadi). Las resoluciones de tribunales nacionales por debajo del nivel del tribunal supremo se recogen de forma selectiva, dando prioridad a las de clara relevancia jurisprudencial o metodológica. Las citas de casos contenidas aquí deben verificarse frente al repertorio oficial del tribunal antes de basarse en ellas para la práctica o para una publicación ulterior.

Esta entrega de referencia trata a Alemania, Francia y el Reino Unido como las principales jurisdicciones nacionales. Italia, España, los Países Bajos y los países nórdicos no se estudian por separado en esta entrega, pero se añadirán en entregas trimestrales sucesivas conforme lo justifique el material. Suiza, aunque situada fuera de la Unión Europea, se estudiará en entregas futuras por la fuerte influencia metodológica de las normas de valoración suizas sobre la práctica europea en general.

Contexto regulatorio europeo

Dos desarrollos regulatorios de ámbito comunitario afectan materialmente al entorno de trabajo de los tasadores, aunque ninguno de ellos sea en sí mismo una resolución judicial:

El EU AI Act y los modelos automatizados de valuación. El Reglamento (UE) 2024/1689 (EU AI Act, Reglamento de Inteligencia Artificial de la UE) entró en vigor en 2024, con disposiciones aplicables a los sistemas de IA de alto riesgo que surten efecto de forma escalonada hasta agosto de 2026. El 19 de mayo de 2026 la Comisión Europea publicó un proyecto de directrices sobre la clasificación de los sistemas de IA de alto riesgo para consulta con las partes interesadas. El anexo III del AI Act incluye la evaluación de la solvencia de personas físicas entre las ocho categorías de usos presuntamente de alto riesgo, y el proyecto de directrices de la Comisión desarrolla esta categoría en detalle. Dado que los modelos automatizados de valuación (AVM) empleados en la

originación hipotecaria, en el seguimiento de carteras hipotecarias y en la suscripción de seguros son funcionalmente inseparables de la evaluación de la solvencia de los préstamos subyacentes, los AVM operados por entidades de crédito o aseguradoras dentro de la Unión Europea se clasificarán, en la mayoría de las configuraciones, como sistemas de IA de alto riesgo. Las consecuencias en materia de cumplimiento comprenden sistemas obligatorios de gestión de riesgos, controles de gobernanza de datos, registro automático de eventos con conservación de los registros durante al menos seis meses, mecanismos de supervisión humana que permitan a los operadores comprender y anular las decisiones del sistema, y obligaciones de seguimiento posterior a la comercialización.

Implicación para los tasadores. Los tasadores europeos deben esperar que los flujos de trabajo asistidos por AVM que operan sus clientes prestamistas queden sujetos a requisitos de documentación y explicabilidad sustancialmente mayores. Cuando los tasadores aporten variables de entrada al modelo, datos de calibración o revisión humana de las salidas del AVM, esas actividades pasan a formar parte de la posición de cumplimiento en materia de IA de alto riesgo del propio prestamista.

Energy Performance of Buildings Directive (refundida). La Directiva (UE) 2024/1275, versión refundida de la Energy Performance of Buildings Directive (EPBD, Directiva relativa a la eficiencia energética de los edificios), se publicó en el Diario Oficial el 8 de mayo de 2024 y entró en vigor el 28 de mayo de 2024. El plazo de transposición para los Estados miembros es el 29 de mayo de 2026. Entre otras disposiciones, la EPBD refundida exige a los Estados miembros priorizar la renovación del dieciséis por ciento del parque de edificios no residenciales con peor rendimiento y del cuarenta y tres por ciento del parque residencial con peor rendimiento; fija normas mínimas de eficiencia energética (MEPS); exige edificios de cero emisiones para la nueva construcción pública a partir de 2028 y para toda la nueva construcción a partir de 2030; y ordena la evaluación del carbono de ciclo de vida completo para los edificios nuevos a partir de 2030.

Implicación para los tasadores. La EPBD refundida afectará a las valoraciones inmobiliarias por varias vías: los umbrales mínimos de eficiencia energética crean una nueva categoría de obsolescencia derivada del cumplimiento normativo; las calificaciones del certificado de eficiencia energética (EPC) pasan a ser un determinante material del precio tanto del parque residencial como del comercial; y los costes de renovación necesarios para cumplir deben reflejarse en los enfoques de coste y en la deducción de la vida económica remanente. Los tasadores deben anticipar además que los prestamistas sujetos a la Mortgage Credit Directive (Directiva de Crédito Hipotecario) exigirán cada vez más un tratamiento explícito de las calificaciones EPC y de las necesidades de inversión de capital vinculadas a las MEPS en las valoraciones de garantías.

Notas para entregas sucesivas

Las entregas trimestrales posteriores a esta de referencia se centrarán en los cambios: nuevas resoluciones, resolución de los asuntos pendientes aquí identificados (señaladamente el tratamiento que el Tribunal Constitucional Federal alemán dé a *1 BvR 472/26*) y legislación de desarrollo de los Estados miembros que transponga la EPBD refundida al derecho nacional. Se añadirán, conforme lo justifique el material, el marco jurídico suizo, el tratamiento de las resoluciones de los tribunales supremos italiano y español, y los desarrollos ante el *Conseil constitutionnel* francés.

Resúmenes de casos por jurisdicción

Tribunal de Justicia de la Unión Europea

GR Real (asunto C-351/23). Tribunal de Justicia de la Unión Europea, sentencia de 24 de junio de 2025.

El Tribunal de Justicia se pronunció sobre una cuestión prejudicial planteada por un tribunal eslovaco a raíz de un procedimiento extrajudicial de ejecución hipotecaria en el que unos consumidores perdieron su vivienda familiar en subasta mientras seguía pendiente el proceso judicial en el que se impugnaban cláusulas contractuales potencialmente abusivas. El Tribunal declaró que las garantías de protección de los consumidores derivadas de la Directiva 93/13/CEE exigen que las normas procesales nacionales permitan un control judicial efectivo del carácter abusivo de las cláusulas antes de que una vivienda hipotecada pueda venderse mediante ejecución extrajudicial, y que los procedimientos que permiten que la venta se consuma de forma irreversible antes de completarse ese control son incompatibles con la Directiva.

Implicación para los tasadores. La resolución no es directamente un caso de valoración, pero reconfigura el entorno procesal en el que se encargan y se utilizan las valoraciones de garantías hipotecarias. En varios Estados miembros con procedimientos de ejecución extrajudicial (señaladamente Eslovaquia, la República Checa, Hungría y los países bálticos), el efecto práctico es ralentizar el proceso de subasta ejecutiva y prolongar el periodo durante el cual se utiliza la valoración de la garantía del prestamista. Los tasadores que preparen valoraciones de garantías para contextos de ejecución en esas jurisdicciones deben anticipar que los informes de valoración podrán estar en uso durante periodos más largos que antes y podrán ser contrastados con los movimientos del mercado entre la fecha original de valoración y la venta finalmente producida.

Tribunal Europeo de Derechos Humanos

Demolición de un edificio en Bakú (Azerbaiyán), sentencia de julio de 2025. Tribunal Europeo de Derechos Humanos, artículo 1 del Protocolo n.º 1.

El Tribunal Europeo de Derechos Humanos apreció una vulneración del derecho al respeto de los bienes en relación con la demolición y expropiación de un edificio de Bakú que había sido ordenada tras invocar las autoridades el riesgo sísmico. El tribunal de Estrasburgo consideró que las autoridades internas, al determinar la indemnización, no habían aportado razones suficientes para rechazar el informe de valoración presentado por los demandantes y habían adoptado injustificadamente solo el importe del dictamen pericial encargado por el propio tribunal. El Tribunal concedió un total de €75,000 por daño patrimonial, €3,000 por daño moral y €1,055 en concepto de costas procesales.

Implicación para los tasadores. La resolución refuerza una línea jurisprudencial del artículo 1 del Protocolo 1 que exige que los tribunales internos ponderen de forma motivada las pruebas de valoración contrapuestas en los procedimientos expropiatorios, y que el rechazo del informe de valoración del demandante en favor de un dictamen pericial encargado por el tribunal esté a su vez motivado. Para los tasadores europeos que trabajan en contextos de expropiación y justa indemnización, el caso subraya que la defendibilidad documental de un informe de valoración — el razonamiento explícito que sustenta la selección de comparables, la derivación de los ajustes y la reconciliación del valor final — es en sí misma una cuestión de derechos fundamentales cuando el informe se presenta en apoyo de una reclamación frente al Estado.

Francia

Cour de cassation, Chambre commerciale, 19 de junio de 2024, nº 22-24.169. (Comentada en la literatura profesional a lo largo de 2025.)

La Cour de cassation (Tribunal de Casación francés) declaró, en el contexto de los litigios sobre la *valeur vénale* (valor de mercado) a efectos del impuesto sucesorio, que una venta del bien heredado producida con posterioridad al hecho imponible (el fallecimiento del *de cujus*) no puede ser utilizada por los herederos para revisar a la baja la *valeur vénale* originalmente declarada, ni siquiera cuando el precio de la venta posterior es significativamente inferior y la venta se produce a los pocos meses del fallecimiento. La resolución aplica el principio de que la *valeur vénale* a efectos del impuesto sucesorio se determina a la fecha del hecho imponible y no se corrige por referencia a transacciones posteriores sobre el mismo bien.

Implicación para los tasadores. La resolución confirma una disciplina estricta de fecha de valoración en la práctica francesa del impuesto sucesorio y limita la relevancia de las ventas posteriores como prueba retrospectiva de valor. Los tasadores que preparen valoraciones francesas a efectos del impuesto sucesorio deberían desarrollar pruebas exhaustivas de ventas comparables basadas en transacciones *en la fecha de valoración o antes de ella*, y no deberían presentar ventas posteriores del propio bien objeto de valoración como prueba confirmatoria. Es probable que el mismo razonamiento se extienda al impuesto sobre donaciones (*droits de donation*) y a los contextos del impuesto sobre el patrimonio (*IFI*).

Cour de cassation, 3^e Chambre civile, 8 de enero de 2026, nº 23-22.803. Dictámenes periciales amistosos (no judiciales).

La Tercera Sala de lo Civil de la Cour de cassation dictó una resolución que redefine el valor jurídico de los dictámenes periciales amistosos (no judiciales) en el proceso civil francés. La resolución es consecuencia de la reforma procesal más amplia operada por el *décret* de 18 de julio de 2025, que convirtió la mediación y la resolución amistosa de controversias en la regla por defecto del procedimiento civil. El Tribunal aclaró que los dictámenes periciales amistosos elaborados por encargo contractual de una sola parte pueden no obstante admitirse como prueba y recibir valor probatorio, con sujeción al principio de contradicción y a la apreciación por el tribunal de la independencia y la metodología del perito.

Implicación para los tasadores. La resolución refuerza la posición probatoria de las *expertises immobilières* elaboradas fuera del marco de los peritos judiciales designados por el tribunal (*experts judiciaires*), siempre que el informe observe los estándares metodológicos de la profesión y se documente la independencia del perito. Para los tasadores franceses que trabajan en valoraciones controvertidas en materia sucesoria, matrimonial o mercantil, la resolución crea una base más sólida para que los informes encargados por el cliente puedan citarse e invocarse en litigios posteriores. Cobra, en consecuencia, mayor importancia el cuidado particular en documentar la independencia (ausencia de subordinación, de interés económico o de relación familiar con el cliente) y en aplicar los métodos prescritos por las normas profesionales.

Alemania

Bundesfinanzhof, sentencias de 12 de noviembre de 2025 (II R 25/24, II R 31/24, II R 3/25). La trilogía del *Grundsteuer Bundesmodell*: constitucionalidad del sistema reformado de valoración a efectos del impuesto predial. La Segunda Sala del Tribunal

Fiscal Federal alemán (*Bundesfinanzhof*, BFH) resolvió en tres procedimientos paralelos de revisión que las disposiciones del método de capitalización de rentas (*Ertragswertverfahren*) aplicado bajo el denominado *Bundesmodell* para la valoración de inmuebles residenciales en once estados federados alemanes a efectos del impuesto predial reformado (*Grundsteuer*), vigente desde el 1 de enero de 2025, son conformes a la Constitución. Los demandantes eran propietarios de viviendas en Renania del Norte-Westfalia, Sajonia y Berlín, que impugnaron sus liquidaciones alegando que las reglas de valoración reglamentadas vulneran el principio general de igualdad (artículo 3.1 de la Ley Fundamental, *Grundgesetz*) y producen valores sistemáticamente inexactos.

El BFH sostuvo que el legislador había creado un sistema de valoración *conceptualmente orientado al valor de mercado* (*konzeptionell einer Verkehrswertorientierung folgt*) y diseñado para aproximar, en promedio sobre el conjunto de los objetos valorados, el “valor inmobiliario objetivamente realista” dentro de una horquilla del valor venal (*gemeiner Wert*). El Tribunal consideró que el legislador no había excedido el margen que le corresponde al ponderar los objetivos del esquema de valoración frente a las desigualdades necesariamente inherentes a un esquema de ese tipo, y que el peso significativo atribuido a evitar un nuevo “atasco de valoraciones” mediante actualización periódica automatizada era constitucionalmente admisible. El BFH subrayó asimismo que los contribuyentes conservan la opción, establecida por anteriores *Beschlüsse* del BFH de 27 de mayo de 2024 (II B 78/23 y II B 79/23), de acreditar un valor venal (*gemeiner Wert*) inferior mediante una tasación cualificada al amparo del § 198 BewG, con la limitación de que el valor inferior debe sustentarse en un informe de tasación emitido por un tasador inmobiliario designado públicamente que aplique métodos de valoración reconocidos.

Un recurso de amparo constitucional (*Verfassungsbeschwerde*) apoyado por *Bund der Steuerzahler Deutschland* y *Haus & Grund Deutschland* se encuentra ahora pendiente ante el Tribunal Constitucional Federal (*Bundesverfassungsgericht*) con el número de expediente 1 BvR 472/26.

Implicación para los tasadores. La trilogía es la jurisprudencia alemana de valoración inmobiliaria más relevante del periodo estudiado. Para los tasadores, la consecuencia práctica principal es el aumento de la demanda de tasaciones de “valor inferior” del § 198 BewG (*Verkehrswertgutachten zum Nachweis des niedrigeren gemeinen Werts*) en los casos en que el *Grundsteuerwert* determinado conforme al *Bundesmodell* parece exceder el valor de mercado real del inmueble. El respaldo explícito del BFH a esta vía, unido a su deferencia general hacia el *Bundesmodell*, significa que el informe de tasación es hoy el mecanismo operativo para obtener alivio individual frente a una sobrevaloración sistemática. Los tasadores deben esperar un incremento sostenido de los encargos de este tipo de informes y asegurar el cumplimiento de la metodología de la ImmoWertV 2021 y de los requisitos del § 198 BewG. El recurso de amparo pendiente ante el BVerfG (1 BvR 472/26) será relevante para el seguimiento de cualquier cambio ulterior de este marco y se prevé que sea uno de los asuntos principales de la próxima entrega trimestral.

Bundesfinanzhof, sentencia de 7 de octubre de 2025, IX R 26/24. Reparto del precio de compra entre suelo y edificación (*Kaufpreisaufteilung*). En un asunto de impuesto sobre la renta relativo a la base de amortización (*Absetzung für Abnutzung*, AfA) de un inmueble residencial en alquiler, la Novena Sala del Tribunal Fiscal Federal abordó la metodología para repartir el precio contractual de compra entre el componente de suelo no amortizable (*Grund und Boden*) y el componente de edificación amortizable (*Gebäude*). Los demandantes habían cuestionado que el tribunal de instancia se apoyara en los *Bodenrichtwerte* (valores de referencia del suelo) alegando que la base de datos subyacente no estaba documentada de forma fiable para la ubicación concreta del inmueble.

Implicación para los tasadores. La resolución continúa una línea jurisprudencial del BFH (señaladamente BFH II R 25/24 de 12 de noviembre de 2025, citada más arriba, y la anterior BFH IX R 12/21 de 20 de septiembre de 2022) que confirma que la ImmoWertV es la fuente de los principios de valoración reconocidos para el reparto del precio de compra. El caso ilustra además la importancia de la base de datos que sustenta los *Bodenrichtwerte*: cuando los *Gutachterausschüsse* (comités periciales de tasación) no disponen de datos suficientes de ventas comparables para la ubicación concreta, el *Bodenrichtwert* puede impugnarse por motivos de calidad de los datos. Para los tasadores que preparen repartos de precio de compra o dictámenes sobre la base de amortización, esto significa que la documentación de la base de datos del *Bodenrichtwert* — incluida la *Kaufpreissammlung* subyacente — pasa a formar parte del análisis defendible.

Reino Unido

Mehrunnisa Alikhan Ahmed Khan v Wolverhampton City Council. Upper Tribunal (Lands Chamber), 2025.

El Upper Tribunal (Lands Chamber) confirmó que la fecha de valoración procedente para determinar la indemnización por expropiación forzosa (*compulsory purchase*) es aquella en que el inmueble se transmite a la autoridad expropiante en virtud de una general vesting declaration (GVD, declaración general de transmisión). El Tribunal abordó asimismo la relación entre valoraciones contemporáneas y retrospectivas: en el caso enjuiciado, la valoración inicial de la corporación local realizada inmediatamente después de la GVD fue de £150,000, mientras que una segunda valoración elaborada en 2025 durante el procedimiento ante el tribunal, referida a la fecha de transmisión supuesta tres meses después de la notificación de la propietaria, fue de £120,000. Pese a confirmar la fecha legal de valoración, el Tribunal prefirió la cifra calculada en un momento más próximo a la fecha en que se emitió la GVD, lo que se tradujo en un pago superior a la reclamante.

Implicación para los tasadores. La resolución resulta metodológicamente instructiva para los tasadores del Reino Unido que trabajan en contextos de expropiación forzosa. Aunque la fecha legal de valoración es la fecha de transmisión, la preferencia del Tribunal por la valoración contemporánea frente a una valoración retrospectiva elaborada años después sobre la misma fecha legal de valoración ilustra que las pruebas de mercado contemporáneas conservan peso práctico incluso cuando formalmente no serían admisibles a la fecha de transmisión. Los tasadores deberían asegurarse de que los informes de valoración elaborados en fechas próximas a una GVD estén documentados con pruebas de apoyo suficientes para servir como registro principal, anticipando que procedimientos posteriores ante el tribunal pueden apoyarse en el informe contemporáneo con preferencia a una reconstrucción retrospectiva.

Decisiones judiciales recientes — América Latina

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Recent Court Decisions • The Valuation Engineer

William Bert Craytor • Vol. 1, Número 1 • Verano 2026

DOI: 10.67330/ywgzhg47

Alcance y fuentes

Este estudio se ha elaborado a partir de fuentes primarias de acceso público cuando ha sido posible — señaladamente el repositorio de casos de la Corte Interamericana de Derechos Humanos (cor-teidh.or.cr), los portales del Supremo Tribunal Federal (STF) y del Superior Tribunal de Justiça (STJ) de Brasil (portal.stf.jus.br, stj.jus.br), el *Semanario Judicial de la Federación* de México (sjf-semanal.scjn.gob.mx), la *Corte Suprema de Justicia de la Nación* argentina (csjn.gov.ar), el portal del *Poder Judicial* de Chile (poderjudicial.cl) y *Diario Constitucional*, y la *Corte Constitucional* de Colombia (corteconstitucional.gov.co) — y de comentarios secundarios en publicaciones profesionales y académicas. Las resoluciones dictadas en español y portugués se han leído en su idioma original y se describen en inglés en la edición original; los encabezamientos de los casos conservan su grafía y sus signos diacríticos originales.

No se utilizan bases de datos jurídicas de suscripción (vLex, La Ley, Thomson Reuters Proview Latin America). Las resoluciones de tribunales nacionales por debajo del nivel del tribunal supremo se recogen de forma selectiva, dando prioridad a las de clara relevancia jurisprudencial o metodológica. Las citas de casos contenidas aquí deben verificarse frente al repertorio oficial del tribunal antes de basarse en ellas para la práctica o para una publicación ulterior.

Esta entrega de referencia trata a Argentina, Brasil, Chile, Colombia y México como las principales jurisdicciones nacionales. Perú, Venezuela, Ecuador, Bolivia, Paraguay, Uruguay y las jurisdicciones centroamericanas y caribeñas no se estudian por separado en esta entrega, pero se añadirán en entregas trimestrales sucesivas conforme lo justifique el material. Los trabajos de valoración vinculados a la restitución de bienes en Venezuela y a las sanciones relativas a Cuba se estudiarán en una entrega futura por su singularidad metodológica y su relevancia para la práctica transfronteriza.

Contexto regional

A diferencia del estudio de la Unión Europea, este estudio latinoamericano no cuenta con un análogo directo de la capa regulatoria supranacional representada por el EU AI Act y por la Energy Performance of Buildings Directive refundida. La regulación de la tasación inmobiliaria en América Latina es abrumadoramente nacional; las principales influencias supranacionales son la jurisprudencia de derechos fundamentales de la Corte Interamericana de Derechos Humanos conforme al artículo 21 de la Convención Americana (derecho a la propiedad privada) y la adopción de las Normas Internacionales de Información Financiera (en particular la IFRS 13, Medición del Valor Razonable) por la mayoría de las jurisdicciones de la región a efectos de información financiera. No obstante, dos temas metodológicos recorren la jurisprudencia regional y merecen mención aquí:

Valoración de la propiedad comunal indígena. El reto metodológico más distintivo de la práctica tasadora latinoamericana es la valoración de la propiedad comunal indígena y afrodescendiente en contextos de restitución, expropiación e indemnización. La jurisprudencia de la Corte Interamericana ha sostenido de forma constante que la propiedad comunal indígena está protegida constitucionalmente en pie de igualdad con la propiedad privada individual, pero los métodos por los que dicha propiedad se valora — allí donde normalmente no existen transacciones comparables en condiciones de mercado, donde el bien se posee de forma comunal e inalienable y donde el significado cultural y espiritual del territorio es parte integrante de su valor — siguen estando poco desarrollados. A finales de 2025 seguían pendientes o solo parcialmente cumplidas veinticuatro sentencias de la Corte Interamericana favorables a pueblos indígenas, varias de ellas con cálculos indemnizatorios sin resolver que dependen de una metodología de valoración aún no fijada.

Inflación, moneda e indexación en la indemnización expropiatoria. Varias jurisdicciones de la región, señaladamente Argentina, exigen que las metodologías de valoración y los cálculos indemnizatorios lidien con una inflación elevada, con tasaciones en moneda extranjera y con regímenes de indexación complejos. La Corte Suprema argentina ha sostenido que la indemnización expropiatoria originalmente expresada en moneda extranjera no puede convertirse a moneda local a una paridad artificial, y que la valoración debe reflejar el valor económico real del bien al momento de la indemnización. Preocupaciones similares se plantean en Venezuela (en relación con las valoraciones de restitución y las vinculadas a sanciones) y periódicamente en Brasil durante los periodos de inflación elevada. Para los tasadores que ejercen en varias jurisdicciones de la región, estas consideraciones de inflación y moneda son parte integrante de la metodología de valoración de un modo que carece de paralelo europeo o estadounidense.

Notas para entregas sucesivas

Las entregas trimestrales posteriores a esta de referencia se centrarán en los cambios: nuevas decisiones de la Corte Interamericana, resolución de los asuntos pendientes aquí identificados, avances en la implementación de la Jurisdicción Agraria y Rural de Colombia hacia su puesta en marcha operativa en 2027, y desarrollos nacionales en la práctica de medición del valor razonable conforme a la IFRS 13. Perú, Venezuela, Ecuador y las jurisdicciones centroamericanas se añadirán conforme lo justifique el material. Se prestará particular atención al desarrollo metodológico de la valoración de la propiedad comunal indígena, ámbito en el que los pronunciamientos sustantivos de la Corte Interamericana han ido por delante de la metodología de valoración práctica disponible para tasadores y autoridades indemnizatorias.

Resúmenes de casos por jurisdicción

Corte Interamericana de Derechos Humanos

Maya Q'eqchi' Agua Caliente vs. Guatemala y Comunidad Garífuna de San Juan y sus Miembros vs. Honduras. Corte Interamericana de Derechos Humanos, sentencias de 2024, comentadas en la jurisprudencia regional a lo largo de 2025.

En dos sentencias contemporáneas la Corte Interamericana se pronunció sobre las obligaciones de Guatemala y Honduras de demarcar, titular e impedir el despojo de los territorios ancestrales pertenecientes a comunidades indígenas y afrodescendientes. La comunidad Maya Q'eqchi' Agua Caliente había obtenido un título de propiedad definitivo sobre su territorio solo tras cuarenta y cinco años

de actuaciones administrativas y judiciales; la comunidad Garífuna de San Juan, en Honduras, se enfrentaba al despojo continuado de tierras costeras ancestrales. Ambas sentencias abordaron las obligaciones del Estado conforme al artículo 21 (derecho a la propiedad privada) de la Convención Americana de reconocer, demarcar y proteger la propiedad comunal indígena. Comentaristas regionales han observado que el marco de restitución de la Corte Interamericana en estas decisiones es algo más restrictivo que en sentencias anteriores, en particular en su tratamiento de los “terceros inocentes” (*terceros inocentes*) que adquirieron derechos superpuestos sin conocimiento de la reclamación indígena.

Implicación para los tasadores. Las decisiones refuerzan que la propiedad comunal indígena y afrodescendiente está protegida constitucionalmente en todo el sistema interamericano y que la valoración de dicha propiedad en contextos de restitución o indemnización no puede apoyarse en ventas comparables convencionales en condiciones de mercado (que normalmente no existen) sin un complemento metodológico. Los tasadores contratados en contextos de restitución indígena deben anticipar que la valoración ha de integrar tres componentes distintos: el valor de mercado subyacente de tierras con características físicas comparables en la región; los costes de oportunidad y las mejoras atribuibles al uso indígena; y el significado cultural y espiritual no económico del territorio, que la jurisprudencia de la Corte Interamericana trata como una dimensión indemnizable del derecho de propiedad. La metodología para este tercer componente sigue poco desarrollada y es objeto continuado del razonamiento de la Corte Interamericana.

Mapuche / Consejo de Todas las Tierras vs. Chile. Corte Interamericana de Derechos Humanos, sentencia de 2024 (comentada en la jurisprudencia de 2025).

La Corte Interamericana condenó al Estado chileno por violaciones de derechos humanos contra 135 miembros del pueblo mapuche en relación con las acciones de reivindicación territorial del Consejo de Todas las Tierras (1989–1992) y con la aplicación de la legislación antiterrorista a las reclamaciones territoriales indígenas. La sentencia amplía la doctrina de *Norín Catrimán y otros vs. Chile* (2014) y reconoce expresamente el derecho de los pueblos indígenas a la libre determinación en relación con las reclamaciones territoriales.

Implicación para los tasadores. Aunque los pronunciamientos principales del caso se refieren a garantías procesales penales más que directamente a la metodología de valoración, el reconocimiento de la libre determinación indígena en relación con el territorio crea una capa procesal que afecta al trabajo de valoración en la restitución territorial indígena chilena y en contextos afines. Los tasadores contratados en tales contextos deberían asegurarse de que la metodología incorpora procesos de consulta adecuados y reconoce la libre determinación de la comunidad sobre el modo en que se caracterizan los valores de su territorio.

Argentina

Doctrina argentina sobre el Tribunal de Tasaciones de la Nación. Consolidada a través de la jurisprudencia de 2025.

La Corte Suprema de Justicia de la Nación (CSJN) y los tribunales federales inferiores han continuado desarrollando la doctrina relativa a las tasaciones realizadas por el Tribunal de Tasaciones de la Nación (TTN) en materia de expropiación y de *expropiación inversa*. La doctrina consolidada de la CSJN (Fallos 313:82, aplicada reiteradamente a lo largo de 2024–2025) sostiene que las tasaciones del TTN gozan de peso presuntivo en razón de la capacidad técnica de los miembros del tribunal, de la base probatoria en que se fundan y de la uniformidad de su práctica; apartarse de una tasación

del TTN exige acreditar error u omisión manifiestos. Los tribunales federales que aplicaron esta doctrina en 2024–2025 — entre ellos el Juzgado Federal de Oberá en *Sosa c. Entidad Binacional Yacyretá* (junio de 2024, *expropiación inversa* derivada del embalse hidroeléctrico de Yacyretá) — siguen apoyándose en las tasaciones del TTN como anclaje operativo de la indemnización.

Implicación para los tasadores. Para los tasadores argentinos, la consecuencia práctica es que los informes de tasación privados presentados en oposición a las tasaciones del TTN afrontan una carga probatoria sustancial. Para desplazar o modificar una tasación del TTN, el informe privado debe demostrar error u omisión metodológicos en el trabajo del TTN, y no meramente una conclusión de valor alternativa. Los tasadores que preparen tales informes deberían centrar la crítica metodológica en la selección de ventas comparables del TTN, en la derivación de sus ajustes y en su tratamiento de las condiciones del mercado local, más que en la magnitud de la diferencia de valor por sí sola.

Tasaciones en moneda extranjera y paridad indemnizatoria. La doctrina de la CSJN según la cual la indemnización expropiatoria originalmente expresada en moneda extranjera no puede convertirse a pesos argentinos a una paridad artificial de uno a uno sigue vigente a lo largo de 2025. La doctrina protege a los propietarios expropiados frente a indemnizaciones que, por efecto de la inflación o de la política cambiaria, no reflejen el valor económico real del bien al momento de la indemnización. El principio ha sido reafirmado periódicamente en 2024–2025 en relación con procedimientos expropiatorios de larga duración.

Implicación para los tasadores. Los tasadores argentinos que preparen valoraciones para expedientes expropiatorios que abarquen periodos de inflación significativa o de cambios de política cambiaria deberían documentar explícitamente la base monetaria de la valoración y considerar la metodología de indexación adecuada entre la fecha de valoración y la fecha prevista de pago. Es improbable que los informes que no aborden el riesgo de inflación y de moneda resulten persuasivos para los tribunales que aplican la doctrina indemnizatoria de la CSJN.

Brasil

Supremo Tribunal Federal, RE 922.144 (Tema 865 de repercusión general). Supremo Tribunal Federal de Brasil, sentencia de 2024 con aplicación continuada en 2025.

El STF sostuvo que la indemnización en los procedimientos de *desapropriação* es constitucionalmente compatible con el régimen federal de *precatório* establecido por el artículo 100 de la Constitución Federal. En concreto, la diferencia entre el valor inicialmente depositado por el poder público (a efectos de la *imissão provisória na posse*) y el valor finalmente determinado por la pericia judicial debe abonarse a través del sistema de *precatório* y no como un depósito complementario inmediato. La decisión pondera el mandato constitucional de una indemnización *justa, prévia e em dinheiro* (artículo 5, XXIV) frente a las exigencias de pago ordenado del sistema de *precatório*.

Implicación para los tasadores. La decisión tiene consecuencias prácticas directas para la valoración del depósito inicial que preparan los tasadores que actúan para las autoridades expropiantes. Dado que cualquier subestimación del depósito inicial se abonará más tarde por la vía del *precatório* (que puede implicar una demora considerable), existe un incentivo estructural para que los demandantes impugnen las valoraciones iniciales por considerarlas subestimadas. Los tasadores que preparen valoraciones de depósito inicial deberían documentar la metodología con especial cuidado para defender la valoración frente a la pericia judicial posterior, anticipando que la brecha entre el depósito inicial y la determinación judicial final quedará sujeta a intereses compensatorios (*juros compensatórios*) conforme a la jurisprudencia del STF y al marco del Decreto-Lei 3.365/1941.

Superior Tribunal de Justiça, REsp 2.129.162 (Tema repetitivo). Superior Tribunal de Justicia de Brasil, sentencia de mayo de 2025.

El STJ resolvió una controversia repetitiva relativa a los honorarios de abogado en los casos en que la autoridad expropiante desiste de la acción de *desapropriação*. El Tribunal estableció que los honorarios en los casos de desistimiento deben seguir el marco del artículo 27, § 1, del Decreto-Lei 3.365/1941 y no las reglas procesales generales del *Código de Processo Civil*. La decisión trae causa de una acción de Cemig (Companhia Energética de Minas Gerais) por la que se solicitaba una servidumbre administrativa para una línea de distribución eléctrica.

Implicación para los tasadores. La decisión es de naturaleza procesal más que metodológica, pero tiene implicaciones para la estructura del encargo al tasador en materia de servidumbres administrativas y de *desapropriação*. Los tasadores contratados como peritos (*peritos judiciais*) en tales procedimientos deberían asegurarse de que los acuerdos de honorarios se documenten con referencia al marco del decreto-ley aplicable y no a las reglas procesales generales.

Superior Tribunal de Justiça, doctrina sobre los *juros compensatórios*. Aplicación continuada en 2025.

El STJ ha mantenido su doctrina según la cual los intereses compensatorios (*juros compensatórios*) en materia de *desapropriação* solo se devengan tras la decisión judicial sobre el título del bien expropiado. Los intereses se aplican a la diferencia entre el depósito inicial y el valor final de la indemnización, al tipo históricamente establecido del doce por ciento anual (Súmula 618 del STF), y sirven para compensar al propietario expropiado por la pérdida anticipada de la posesión antes de que se abone la indemnización final.

Implicación para los tasadores. Los tasadores brasileños deben anticipar que la brecha de valoración que identifican en la pericia judicial — entre el depósito inicial del poder público y el valor de mercado tasado — se convierte en la base del cálculo de los intereses compensatorios. La relevancia económica de esa brecha es, por tanto, mayor de lo que sugiere la diferencia nominal, en particular en procedimientos largos.

Chile

Corte Suprema, Rol N° 31.423-2025. Corte Suprema de Chile, sentencia de septiembre de 2025.

La Corte Suprema de Chile rechazó un recurso de casación interpuesto por una inmobiliaria contra el *Fisco de Chile* en un procedimiento expropiatorio especial del artículo 12 del Decreto Ley N° 2.186. La demandante había alegado que la valoración fijada por la *Comisión de Peritos* no reflejaba el mayor y mejor uso del inmueble (*mayor y mejor uso*) ni el valor de las construcciones existentes y de las obras complementarias, y había solicitado además indemnización por lucro cesante (*lucro cesante*) sobre la base de un contrato de arrendamiento vigente que habría generado rentas futuras. La Corte sostuvo que los argumentos de la demandante equivalían a una discrepancia con la ponderación de la prueba efectuada por el tribunal de instancia, y no a una verdadera infracción de las normas legales que rigen la valoración, y que el razonamiento del tribunal de instancia — que se apoyó en la prueba de la *Comisión de Peritos* junto con prueba documental y testimonial — estaba adecuadamente fundado.

Implicación para los tasadores. La decisión ilustra la deferencia sustancial que la doctrina de la Corte Suprema de Chile otorga a la valoración de la *Comisión de Peritos* en los procedimientos

expropiatorios, y el alcance limitado del control casacional sobre las conclusiones de valor por el solo motivo de discrepancia metodológica. Los tasadores que preparen valoraciones contrapuestas en expedientes expropiatorios chilenos deberían centrarse en demostrar el error metodológico en el trabajo de la Comisión más que en un razonamiento de valoración alternativa, y deben ser conscientes de que las reclamaciones de *lucro cesante* basadas en contratos de arrendamiento vigentes afrontan exigencias probatorias sustanciales.

Corte Suprema (expropiación de Coquimbo). Corte Suprema de Chile, sentencia de 17 de julio de 2025.

La Corte Suprema de Chile ratificó una decisión de apelación que elevaba la indemnización en un expediente expropiatorio de Coquimbo a UF 2,969.75 (sobre la base del informe de un tasador privado) frente al importe defendido por el Fisco de UF 2,349.07 (basado en ventas inscritas del *Conservador de Bienes Raíces*). La Corte subrayó que el apoyo del tribunal de apelación en el informe del tasador privado, considerado junto con prueba testimonial y un informe de tasación adicional sobre un inmueble cercano, constituía un ejercicio admisible de la *sana crítica* (ponderación razonada de la prueba) y que el control casacional no alcanza a revisar esa ponderación probatoria.

Implicación para los tasadores. La decisión de Coquimbo es la contrapartida del Rol N° 31.423-2025: si la valoración de la Comisión prevalece cuando la impugna una inmobiliaria, también puede prevalecer la valoración superior de un tasador privado cuando la sustentan un razonamiento metodológico adecuado y prueba corroborante. La lección práctica para los tasadores chilenos es que la metodología y el soporte documental de una tasación privada son decisivos, y que la prueba de ventas inscritas por sí sola no basta para desplazar una tasación privada bien fundada que se apoya en comparables adicionales y en prueba contextual.

Tribunal Constitucional, octubre de 2025. Tasación de *bienes raíces* conforme al Código Tributario.

El Tribunal Constitucional de Chile rechazó una impugnación constitucional de las disposiciones del Código Tributario que rigen la tasación de los inmuebles (*bienes raíces*) a efectos del impuesto territorial. Los ministros discrepantes sostuvieron que las normas impugnadas ofrecían mecanismos insuficientes para que la parte ejecutada objetara la tasación, comprometiendo la garantía constitucional de un procedimiento racional y justo.

Implicación para los tasadores. La decisión mantiene el marco administrativo vigente de tasación para el impuesto inmobiliario chileno, pero el razonamiento discrepante identifica una vulnerabilidad del marco actual que podría revisarse en futuras impugnaciones constitucionales o contencioso-administrativas. Los tasadores chilenos que preparen valoraciones en relación con controversias de impuesto inmobiliario deberían vigilar el entorno normativo posterior a la decisión, ya que parecen plausibles reformas administrativas destinadas a mejorar los mecanismos de objeción del contribuyente.

Colombia

Corte Constitucional, aprobación de la Jurisdicción Agraria y Rural. Corte Constitucional de Colombia, decisión de 2025.

La Corte Constitucional de Colombia aprobó el control constitucional de la legislación por la que se crea la nueva Jurisdicción Agraria y Rural, una jurisdicción especializada en tierras rurales den-

tro de la rama judicial colombiana. La Corte avaló los artículos clave que definen el objeto de la reforma, su integración en la rama judicial, la creación de tribunales y juzgados especializados y los aspectos presupuestarios. Está previsto que la nueva jurisdicción comience a operar en 2027 y está diseñada para resolver controversias de restitución de tierras, disputas de propiedad agraria y asuntos derivados de la implementación del acuerdo de paz de 2016. A la fecha de la aprobación, se encontraban pendientes en el sistema administrativo y judicial existentes controversias agrarias que afectan a más de 61.000 hectáreas.

Implicación para los tasadores. La nueva jurisdicción se convertirá en el foro principal de las controversias de valoración derivadas de los procesos de restitución de tierras y de los asuntos de propiedad rural en Colombia. Los tasadores colombianos que ejercen en contextos de restitución (incluidos quienes trabajan para la Unidad de Restitución de Tierras y el Instituto Geográfico Agustín Codazzi) deben anticipar que los nuevos tribunales desarrollarán su propia jurisprudencia de valoración y que los estándares metodológicos podrán evolucionar desde el marco actual, de impulso administrativo. Los profesionales deberían seguir la puesta en marcha operativa en 2027 y el desarrollo de la carga inicial de asuntos de los tribunales.

Marco vigente de valoración basado en el IGAC para la restitución de tierras. Aplicación continuada en 2025.

La práctica colombiana de restitución de tierras sigue apoyándose en el *avalúo comercial* elaborado por el Instituto Geográfico Agustín Codazzi (IGAC) como base principal de la compensación en contextos de restitución y sustitución, complementado por avalúos elaborados por Lonjas de Propiedad Raíz cuando se requiere. El marco está articulado en la sentencia SU-034 de 2018 de la Corte Constitucional (sobre los requisitos procedimentales de cumplimiento) y en decisiones posteriores, entre ellas la T-120 de 2024 (sobre la relación entre la valoración del IGAC y las preferencias individuales del beneficiario en situaciones de compensación en especie).

Implicación para los tasadores. Para los tasadores que trabajan en la restitución de tierras en Colombia, el marco institucional sitúa las valoraciones del IGAC en el centro operativo del cálculo indemnizatorio, con los avalúos de la Lonja de Propiedad Raíz como mecanismo complementario o de resolución de controversias. La futura Jurisdicción Agraria y Rural puede alterar esta arquitectura institucional, y el peso relativo de las valoraciones del IGAC frente a las de la Lonja es una de las cuestiones que probablemente desarrollen los nuevos tribunales.

México

Suprema Corte de Justicia de la Nación, causación del ISR sobre el diferencial entre avalúo y precio. Suprema Corte de México, decisión anunciada a principios de 2026.

El Pleno de la Suprema Corte de Justicia de la Nación (SCJN) declaró constitucionales los artículos 125, 130(IV) y 132 de la Ley del Impuesto sobre la Renta (ISR), que disponen que, cuando el avalúo de una adquisición inmobiliaria excede el precio contractual de compra en más de un diez por ciento, la diferencia se trata como ingreso acumulable para el adquirente y queda por tanto sujeta al ISR. La Corte sostuvo que la regla no vulnera los principios constitucionales de proporcionalidad y equidad tributarias, razonando que la adquisición de un bien a un precio materialmente inferior al valor de mercado representa un incremento objetivo del patrimonio del adquirente y constituye por ello, con propiedad, un beneficio económico gravable.

Implicación para los tasadores. La decisión tiene consecuencias prácticas sustanciales para las tran-

sacciones inmobiliarias mexicanas y para la práctica tasadora que las sustenta. Los avalúos realizados por *corredores públicos* u otros profesionales autorizados tienen ahora consecuencias fiscales directas para los adquirentes cuando el avalúo excede el precio contractual en más de un diez por ciento. Los tasadores deben anticipar un mayor escrutinio de la metodología en los avalúos transaccionales y una mayor demanda de avalúos que documenten con particular rigor la base de ventas comparables. La decisión eleva además implícitamente la trascendencia de la elección metodológica entre enfoques de valoración alternativos, puesto que esa elección puede afectar ahora a la deuda tributaria del adquirente.

Suprema Corte de Justicia de la Nación, Amparo en revisión 391/2024. Segunda Sala de la Suprema Corte de México, sentencia de 25 de junio de 2025.

La Segunda Sala de la Suprema Corte de México resolvió una impugnación constitucional planteada por una empresa minera contra un Decreto que afectaba al marco jurídico de la expropiación de lotes mineros y de los derechos accesorios conexos (ocupación temporal, servidumbres, preferencias en concesiones de agua y disposiciones sobre prórroga de arrendamientos). La Corte sostuvo que el Decreto no vulneraba el principio de irretroactividad porque las disposiciones afectadas se referían a meras expectativas de derechos futuros y no a derechos adquiridos.

Implicación para los tasadores. La decisión es relevante para la valoración de derechos mineros y de propiedades mineras en México. La caracterización por la Corte de los derechos de prórroga de lotes mineros y de las disposiciones conexas como “meras expectativas” y no como derechos adquiridos afecta a la valoración de las concesiones mineras y a la apreciación del riesgo regulatorio en dicha valoración. Los tasadores que preparen valoraciones de propiedades mineras o de derechos conexos en México deberían abordar explícitamente la dimensión de riesgo regulatorio de la valoración y evitar tratar los valores contingentes o dependientes de una prórroga como si estuvieran plenamente consolidados.