

# Valuation Engineer Journal

*L'ingénierie de la valeur : mesurée, rigoureuse, reproductible.*

Volume 1 · Numéro 1 · Été 2026



## DANS CE NUMÉRO

L'écart entre doctrine et pratique dans l'évaluation immobilière

Residual Constraint Approach (RCA) : cadre et protocole

earthUI, glmnetUI & mgcvUI : interfaces de modélisation ouvertes

**FONDEMENTS (F0–F6)** Bien hétérogène · Espace des caractéristiques ·

Fonction de prix hédonique · Prix implicite · Régression hédonique · Variable latente

**DÉCISIONS DE JUSTICE RÉCENTES** États-Unis · Union européenne · Amérique latine

# Valuation Engineer Journal

Volume 1 · Numéro 1

Été 2026

ISSN 3143-5033

**Rédacteur en chef** William Bert Craytor, SRA

## Sommaire

|                                                                                                    |            |
|----------------------------------------------------------------------------------------------------|------------|
| <b>Mission</b>                                                                                     | <b>2</b>   |
| <b>Éditorial</b>                                                                                   | <b>3</b>   |
| Bienvenue dans le numéro inaugural . . . . .                                                       | 4          |
| <b>Lettres à la rédaction</b>                                                                      | <b>6</b>   |
| Une invitation à correspondre . . . . .                                                            | 7          |
| <b>Articles</b>                                                                                    | <b>9</b>   |
| L'écart entre doctrine et pratique dans l'évaluation immobilière . .                               | 10         |
| Residual Constraint Approach (RCA) : cadre et protocole . . . . .                                  | 42         |
| earthUI : une interface interactive pour le paquet earth (MARS) .                                  | 101        |
| glmnetUI : une interface interactive pour le paquet glmnet (lasso<br>et elastic net) . . . . .     | 149        |
| mgcvUI : une interface interactive pour le paquet mgcv (modèles<br>additifs généralisés) . . . . . | 187        |
| <b>Fondements</b>                                                                                  | <b>230</b> |
| Vue d'ensemble (F0) . . . . .                                                                      | 231        |
| Bien hétérogène (F1) . . . . .                                                                     | 237        |
| Espace des caractéristiques (F2) . . . . .                                                         | 245        |
| Fonction de prix hédonique (F3) . . . . .                                                          | 252        |
| Prix implicite (F4) . . . . .                                                                      | 258        |
| Régression hédonique (F5) . . . . .                                                                | 265        |
| Variable latente (F6) . . . . .                                                                    | 273        |
| <b>Décisions de justice récentes</b>                                                               | <b>280</b> |
| Décisions de justice récentes – États-Unis (R1) . . . . .                                          | 281        |
| Décisions de justice récentes – Union européenne (R2) . . . . .                                    | 284        |
| Décisions de justice récentes – Amérique latine (R3) . . . . .                                     | 290        |

## Mission

La *Valuation Engineer Journal* publie des travaux de recherche et des études de cas rigoureux et tournés vers la pratique sur l'ingénierie de la valeur : méthodes, modèles et mesures reliant les décisions techniques aux résultats économiques. Nous privilégions la reproductibilité, la transparence des hypothèses et la clarté de l'exposition.

# Éditorial

RUBRIQUE : ÉDITORIAL

## Bienvenue dans le numéro inaugural

*Traduction française, préparée avec l'assistance d'une IA et revue par la  
rédaction ; la version anglaise fait foi.*

William Bert Craytor, SRA — Rédacteur en chef

Volume 1, Numéro 1 · Été 2026

Publié le 1<sup>er</sup> juillet 2026

DOI: 10.67330/0tymfa32

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

Cet article en libre accès est diffusé selon les termes de la licence Creative Commons  
Attribution 4.0.

**Résumé.** Le Message des rédacteurs du volume 1, numéro 1 : pourquoi cette revue existe, ce qu'apportent les cinq articles et les trois rubriques, et une invitation à y prendre part.

**Mots-clés :** éditorial, ingénierie de l'évaluation, numéro inaugural

Bienvenue dans le numéro inaugural de *The Valuation Engineer Journal*.

Cette revue existe à cause d'un écart. La doctrine de l'évaluation immobilière décrit depuis longtemps l'évaluation comme la mesure disciplinée du comportement du marché ; pourtant, la pratique quotidienne repose encore largement sur une sélection manuelle des biens comparables et sur des ajustements estimés au seul jugement. Pendant ce temps, les outils statistiques capables de mesurer ce que les marchés paient réellement pour les caractéristiques d'un bien — régression hédonique, modèles de splines modernes, apprentissage automatique validé de façon croisée — ont mûri pendant des décennies sans guère être adoptés dans l'évaluation en exercice. L'*ingénieur en évaluation* (*valuation engineer*) est le nom que nous donnons au praticien qui comble cet écart : la compétence de l'expert agréé, jointe à l'ingénierie des données, à la statistique et à un protocole reproductible.

Les cinq articles de ce numéro posent les fondations de travail. L'article d'ouverture examine l'écart entre doctrine et pratique lui-même et soutient que les normes mêmes de la profession exigent déjà davantage de mesure que ses méthodes quotidiennes n'en délivrent. En deuxième lieu — et au cœur de tout ce que cette revue entend faire — l'approche par contrainte résiduelle (*Residual Constraint Approach*) est présentée intégralement : un protocole qui laisse le canal de régression mesurer ce qui peut l'être, discipline ce qui ne le peut pas en une décomposition résiduelle explicite, et contraint le total de telle sorte que le jugement puisse affiner la valeur sans jamais l'inventer. Suivent trois articles logiciels, qui documentent earthUI et ses interfaces

compagnes pour la régression pénalisée et les modèles additifs généralisés — l’outillage ouvert sur lequel cette revue s’appuie — regroupés peu avant la mise sous presse en une seule application, comme l’exposent les notes de l’éditeur dans chaque article.

Trois rubriques accompagnent les articles. *Foundations* (Fondements) s’ouvre sur sept courtes entrées (F0–F6) retraçant la chaîne conceptuelle qui va des biens hétérogènes à l’espace des caractéristiques, à la fonction de prix hédonique, aux prix implicites et à la régression hédonique, jusqu’aux variables latentes qui résident dans les résidus. Chaque entrée cite sa filiation savante, porte son propre DOI et se termine par une formalisation Prolog lisible par machine — le commencement d’une base de connaissances que nous entendons faire croître, numéro après numéro, jusqu’à constituer un vocabulaire commun et précis pour le domaine. *Recent Court Decisions* (Décisions de justice récentes) passe en revue les décisions rendues aux États-Unis, dans l’Union européenne et en Amérique latine qui touchent à la pratique de l’évaluation, car l’environnement juridique fait partie du problème d’ingénierie. *Letters to the Editor* (Courrier des lecteurs) ouvre la colonne de correspondance : les méthodes destinées à être utilisées sont mises à l’épreuve d’une manière que leurs auteurs n’avaient pas prévue, et c’est ici que ces constats — et les désaccords qu’ils suscitent — ont leur place.

Chaque contribution de ce numéro est publiée en libre accès sous licence CC BY 4.0, en PDF et en HTML, avec un DOI Crossref, et préservée de façon indépendante sur Zenodo. Nous appelons les contributions — articles, entrées *Foundations*, correspondance et désaccords — des experts immobiliers, des économistes et des ingénieurs.

William Bert Craytor, SRA  
*Rédacteur en chef*

## Lettres à la rédaction

RUBRIQUE : COURRIER DES LECTEURS

## Une invitation à correspondre

*Traduction française, préparée avec l'assistance d'une IA et revue par la  
rédaction ; la version anglaise fait foi.*

William Bert Craytor, SRA — Rédacteur en chef

Volume 1, Numéro 1 · Été 2026

Publié le 1<sup>er</sup> juillet 2026

DOI: 10.67330/r55mw896

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

Cet article en libre accès est diffusé selon les termes de la licence Creative Commons  
Attribution 4.0.

**Résumé.** Le rédacteur ouvre la rubrique du Courrier des lecteurs : ce que cette revue espère recevoir de ses lecteurs, et comment le lui adresser.

**Mots-clés :** courrier des lecteurs, correspondance, commentaire par les pairs

Une revue qui ne fait que parler n'est qu'une demi-revue. Cette rubrique existe pour l'autre moitié.

*The Valuation Engineer Journal* publie des méthodes destinées à être employées — et les méthodes employées se trouvent mises à l'épreuve de façons que leurs auteurs n'avaient jamais prévues. Si vous avez appliqué l'approche par contrainte résiduelle (*Residual Constraint Approach*) et trouvé une étape qui échoue sur votre marché, nous voulons cette lettre. Si une entrée de la rubrique *Foundations* définit un terme plus étroitement que votre pratique ne l'exige, nous voulons cette lettre. Si une décision de justice que nous avons résumée se lit autrement pour un praticien de ce ressort, nous voulons cette lettre également. Le désaccord, la réplique, l'affinement et la simple correction entrent tous dans notre champ ; il en va de même des brefs comptes rendus de pratique — une technique qui a fonctionné, un problème de données résolu, un échec dont il vaut la peine d'avertir les autres.

Les règles sont peu nombreuses. Les lettres doivent être signées, concises (une page est l'idéal) et précises : citez l'article, la section ou l'entrée que vous visez, et apportez vos preuves comme vous le feriez dans un dossier de travail. Nous pouvons intervenir sur la longueur et la clarté, toujours avec l'accord de l'auteur. Les lettres sont publiées en libre accès sous licence CC BY 4.0, comme tout le reste de ces pages, et les auteurs dont les travaux sont discutés se verront offrir la possibilité de répondre dans le même numéro.

Adressez vos lettres via le site de la revue ([journal.valuation-engineer.com](http://journal.valuation-engineer.com)) ou directement au rédacteur. Les premières véritables lettres que recevra cette revue marqueront le moment où elle deviendra une conversation plutôt qu'une publication — nous les attendons avec impatience.

William Bert Craytor, SRA  
*Rédacteur en chef*

## Articles

# L'écart entre doctrine et pratique en évaluation immobilière : un exposé structuré des fonctions, des limites et des tensions

*Traduction française de l'article original en anglais. Traduction préparée avec l'assistance d'une*

*IA et revue par la rédaction ; la version anglaise fait foi.*

William Bert Craytor

Volume 1, Numéro 1 · 21 mai 2026

Publié le 1<sup>er</sup> juillet 2026

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

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

Cet article en libre accès est diffusé selon les termes de la [licence Creative Commons Attribution 4.0](#).

**Résumé.** L'évaluation immobilière occupe une position d'intérêt public significatif ; pourtant, sa doctrine officielle — centrée sur des opinions de valeur de marché dans des conditions passées et actuelles — entre en tension croissante avec l'usage réel des produits d'évaluation dans les chaînes financières contemporaines. Cet article construit cette démonstration en cinq étapes, chacune accompagnée d'une base de connaissances Prolog explicite et interrogeable : les deux premières établissent les fonctions sociales et économiques que remplit l'évaluation ; la troisième identifie son principal produit final, une opinion de valeur de marché ; et la quatrième examine l'orientation temporelle et méthodologique de l'évaluation par rapport à la prévision quantitative et financière, ainsi que la frontière réglementaire qui les sépare. Ces étapes convergent vers la thèse centrale de la cinquième : dans le crédit hypothécaire, la valeur de marché pondérée par un ratio prêt-valeur (LTV) fonctionne comme un substitut implicite de la valeur de saisie future — un usage prospectif que la doctrine de la valeur présente ne reconnaît pas. De nombreux experts, ignorant ce rôle d'évaluation de la garantie, ne traitent pas la valeur de marché avec la prudence à l'égard des tendances futures de la valeur qu'il exige — un problème que le présent exposé entend mettre au jour et soumettre à des travaux ultérieurs.

**Mots-clés :** évaluation immobilière, valeur de marché, doctrine de l'évaluation, représentation des connaissances, normes d'évaluation, valeur de saisie

## Table des matières

|                                                                                                        |                    |
|--------------------------------------------------------------------------------------------------------|--------------------|
| <a href="#">1 Importance sociale</a>                                                                   | <a href="#">2</a>  |
| <a href="#">2 Importance économique</a>                                                                | <a href="#">6</a>  |
| <a href="#">3 L'importance des estimations de valeur de marché comme produit final des évaluations</a> | <a href="#">10</a> |
| <a href="#">4 L'orientation temporelle de l'activité d'évaluation</a>                                  | <a href="#">14</a> |
| <a href="#">5 La valeur de marché × LTV comme substitut de la valeur de saisie</a>                     | <a href="#">21</a> |
| <a href="#">6 Conclusion</a>                                                                           | <a href="#">28</a> |

# 1 Importance sociale

Les évaluations de biens immobiliers et d'autres actifs remplissent une fonction sociale et économique essentielle en aidant les acheteurs, les vendeurs, les prêteurs, les assureurs, les investisseurs, les tribunaux, les autorités fiscales et le grand public à prendre des décisions éclairées et équitables concernant le transfert, le financement, la taxation et la bonne gestion des biens et d'autres formes de patrimoine. Des évaluations de haute qualité contribuent à la transparence des marchés, à la confiance du public, à l'allocation efficiente du capital et à la stabilité des institutions financières en fournissant des opinions de valeur indépendantes et solidement étayées, fondées sur des éléments probants, l'analyse et le jugement professionnel. (LACOUR-LITTLE et MALPEZZI 2003; K'AKUMU et LARSEN 2024; HENDERSHOTT et KANE 1995; LEE, PENG et LIAO 2020)

Des évaluations de mauvaise qualité peuvent au contraire produire des conséquences graves à plusieurs horizons temporels. À court terme, des évaluations inexactes peuvent causer aux acheteurs ou aux vendeurs individuels un préjudice financier immédiat sous la forme d'un surpaiement, d'une vente à perte, d'une taxation excessive ou de décisions de prêt inappropriées. À moyen terme, des insuffisances systématiques de l'évaluation peuvent fausser les signaux du marché, compromettre la souscription de crédit, encourager les comportements spéculatifs, contribuer à une taxation ou à une couverture d'assurance inéquitables, et accroître les contentieux ainsi que la défiance transactionnelle. À long terme, des défaillances persistantes de la qualité des évaluations peuvent saper la confiance dans les marchés et les institutions, contribuer à l'instabilité financière, à la mauvaise allocation des ressources, au déclin ou au surinvestissement des quartiers, et à des inégalités sociales plus larges touchant les ménages, les entreprises et les collectivités. (LACOUR-LITTLE et MALPEZZI 2003; MAYER et NOTHAFT 2022)

Parce que l'immobilier et les autres actifs représentent souvent une part substantielle du patrimoine des particuliers, des entreprises et des personnes publiques, la qualité des évaluations n'est pas une simple question technique d'ordre privé, mais un enjeu d'intérêt public significatif, lié à la résilience économique, à l'équité et au bien public à long terme. (MAYER et NOTHAFT 2022)

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

:- module(appraisal_social,
    [ social_horizon/1
    , social_party/1
    , social_decision_type/1
    , social_involves/2
    , % DecisionType, Party
```

```

, social_effect/4          % Quality, Horizon, Domain, Effect
, social_contribution/1
, social_hallmark/1
, social_equity_principle/1
, social_public_interest_claim/1
, social_positive_effect/3
, social_negative_effect/3
, social_horizon_summary/2
, social_consequences_of_quality/2
, social_parties_affected_by/2
, social_public_harm/2
, social_private_harm/2
, social_appraisal_is_public_interest/0
]).

:- 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 Importance économique

Les évaluations de biens immobiliers et d'autres actifs jouent un rôle économique fondamental en soutenant le fonctionnement efficient des marchés du crédit, les décisions d'investissement, les systèmes de taxation, l'information financière, la souscription d'assurance, l'administration des successions, les opérations entre entreprises et la planification des politiques publiques. Des évaluations fiables contribuent à établir des mesures crédibles de la valeur des garanties et de la valeur de marché, qui permettent au capital de circuler plus efficacement entre emprunteurs, prêteurs, investisseurs, administrations publiques et entreprises. (LACOUR-LITTLE et MALPEZZI 2003 ; EBRAHIM et HUSSAIN 2010)

À court terme, des évaluations exactes réduisent l'incertitude transactionnelle, améliorent les décisions de prêt, facilitent des négociations équitables et aident à prévenir les pertes financières immédiates résultant d'une surévaluation ou d'une sous-évaluation. À moyen terme, la qualité des évaluations influe sur la stabilité et l'efficacité des marchés dans leur ensemble en affectant le risque hypothécaire, l'allocation des investissements, l'équité fiscale, l'adéquation de la couverture d'assurance et le comportement de formation des prix des participants au marché. Des évaluations faibles ou systématiquement biaisées peuvent, sur cette période, contribuer à des bulles d'actifs, à la dégradation des portefeuilles de prêts, à la distorsion des incitations au développement immobilier et à l'accroissement du risque de défaut ou de contentieux. (LACOUR-LITTLE et MALPEZZI 2003 ; HENDERSHOTT et KANE 1992 ; GRIFFIN et MATURANA 2016)

À long terme, les conséquences économiques cumulées de la qualité des évaluations deviennent encore plus significatives : elles influencent les dynamiques de formation du capital, l'investissement en infrastructures, la transmission intergénérationnelle du patrimoine, les recettes publiques, l'accessibilité du logement et la résilience globale du système financier. Des insuffisances durables des pratiques d'évaluation peuvent contribuer à une mauvaise allocation chronique du capital, à l'instabilité bancaire, à l'érosion de la confiance des investisseurs et à un usage inefficace du foncier et des actifs productifs, tandis que des normes d'évaluation solides

soutiennent des marchés stables, une planification économique saine et la croissance économique à long terme.(GHENT, TOROUS et VALKANOV 2019; HENDERSHOTT et KANE 1992; EBRAHIM et 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
    ]).

:- disjoint economic_effect/4.
:- disjoint 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 L'importance des estimations de valeur de marché comme produit final des évaluations

Les évaluations de valeur de marché jouent un rôle important dans l'analyse du risque de garantie parce qu'elles fournissent une estimation, élaborée par un professionnel, du prix probable qu'un bien ou un actif pourrait atteindre dans des conditions normales de marché, servant ainsi de référence fondamentale aux prêteurs, aux investisseurs et aux analystes de garanties qui évaluent l'exposition au risque de crédit et les issues potentielles d'une saisie. (AGARWAL, BEN-DAVID et YAO 2015) Bien que la valeur de saisie ou la valeur de liquidation puisse s'écarter substantiellement de la valeur de marché en raison de conditions de commercialisation dégradées, de délais de vente accélérés, de frais juridiques, de la détérioration physique du bien ou d'un sentiment de marché défavorable, les analystes de garanties s'appuient fréquemment sur les évaluations de valeur de marché comme point de départ pour estimer le potentiel de recouvrement dans des scénarios de tension. (CAMPBELL, GIGLIO et PATHAK 2011 ; PENNINGTON-CROSS 2006 ;

CLAURETIE et DANESHVARY 2009)

En analysant la relation entre la valeur de marché, les encours de prêts, les ratios prêt-valeur, les tendances du marché, l'état du bien, la liquidité et les conditions locales d'offre et de demande, les analystes de garanties apprécient la probabilité et la sévérité des pertes en cas de défaut de l'emprunteur.(QI et YANG 2009)

À court terme, des évaluations exactes de la valeur de marché aident les prêteurs à prendre des décisions de souscription prudentes et à établir des marges de garantie appropriées. À moyen terme, elles soutiennent la gestion du risque de portefeuille, la titrisation des prêts, l'allocation des réserves et la conformité réglementaire en aidant les institutions à évaluer l'évolution de l'adéquation des garanties au fil des cycles économiques. À long terme, des pratiques d'évaluation constamment fiables contribuent à la stabilité du système financier en réduisant le risque d'endettement excessif, de prêts mal garantis et de mauvaise valorisation systémique des actifs de garantie, susceptibles d'amplifier les pertes sur saisies et des perturbations économiques plus larges en période de tension sur les marchés.(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
    ]).

:- discontiguous collateral_effect/4.
:- discontiguous 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 L'orientation temporelle de l'activité d'évaluation

Les experts développent généralement des opinions de valeur de marché à un instant donné, à une date d'effet précise — rétrospective (passée), actuelle ou, dans des limites définies, prospective (future) — plutôt que des prévisions probabilistes portant sur des périodes étendues. Cette orientation temporelle et méthodologique constitue l'une des distinctions centrales entre la pratique réglementée de l'évaluation et le travail des analystes quantitatifs, des analystes financiers, des analystes en investissement et des économistes. (QUAN et QUIGLEY 1991)

Dans la plupart des missions d'évaluation, en particulier celles qui portent sur la valeur de marché au sens de normes telles que les Uniform Standards of Professional Appraisal Practice (USPAP, normes uniformes de pratique professionnelle de l'évaluation) ou les International Valuation Standards (IVS, normes internationales d'évaluation), l'expert produit une estimation ponctuelle réconciliée (ou, dans les cas autorisés, une fourchette étroite), ancrée dans des éléments de marché observables, les conditions prévalentes et des méthodologies reconnues à une date d'effet unique. Les USPAP autorisent explicitement des conclusions de valeur exprimées comme un montant précis, une fourchette de valeurs ou une relation à une référence ; l'opinion réconciliée demeure néanmoins une indication ponctuelle crédible rattachée à cette date. (THE APPRAISAL FOUNDATION 2024 ; INTERNATIONAL VALUATION STANDARDS COUNCIL 2025)

Bien que les experts intègrent nécessairement les attentes raisonnables des participants au marché concernant les tendances, l'offre et la demande, la dépréciation, les taux de capitalisation ou d'actualisation et l'absorption, ils le font pour étayer une valeur soutenue par le marché à la date d'effet, non pour prévoir des mouvements spéculatifs. Les opinions de valeur prospective (par exemple, la valeur stabilisée à l'achèvement d'un projet) exigent des hypothèses extraordinaires

ou des conditions hypothétiques explicites et restent circonscrites à une date d'effet future définie, avec un étayage dérivé du marché.

En revanche, les analystes quantitatifs, financiers et en investissement traitent fréquemment l'incertitude au moyen d'estimations couvrant un éventail de résultats possibles sur plusieurs périodes, en recourant à des distributions de probabilité, des simulations de Monte-Carlo, des analyses de scénarios, des tests de sensibilité ou des mesures de valeur en risque (*value-at-risk*). Leur attention se porte souvent sur la performance future attendue, les rendements ajustés du risque ou la valeur d'investissement pour une partie déterminée sur des horizons variables, plutôt que sur un point unique de valeur de marché à une date donnée. (HULL 2018)

Un tel travail présuppose en outre des compétences techniques qu'une licence d'expert n'exige ni ne certifie. L'ajustement d'un modèle de type Multivariate Adaptive Regression Splines (MARS), par exemple, requiert le jugement nécessaire pour maîtriser la complexité du modèle afin qu'un bon ajustement sur l'échantillon (un  $R^2$  élevé) ne soit pas obtenu au prix d'un surajustement, mais se transpose aux données de validation (un  $R^2$  élevé en validation croisée), ce qui suppose à son tour une compréhension opérationnelle des algorithmes sous-jacents. (FRIEDMAN 1991 ; STEURER, HILL et PFEIFER 2021) Outre cette aisance quantitative, ce travail récompense la capacité conceptuelle à manier des concepts fortement structurés, à discerner de fines distinctions de sens et à assimiler des méthodes nouvelles. Il s'agit de compétences et non de titres : un expert agréé peut les posséder et un analyste non agréé peut en être dépourvu, ce qui est précisément la raison pour laquelle elles ne déterminent pas, à elles seules, de quel côté de la ligne réglementaire se situe une tâche donnée.

Les experts peuvent également intégrer des éléments limités tournés vers l'avenir, tels que la projection de flux de revenus, de la vacance, des charges et des valeurs de réversion dans des analyses de flux de trésorerie actualisés, ou l'estimation de valeurs prospectives dans des contextes de développement ou de contentieux. Ces composantes demeurent dans les limites de l'évaluation lorsqu'elles sont fondées sur des éléments de marché actuels et sur des normes professionnelles qui privilégient le caractère étayé des conclusions et l'exclusion de toute spéculation non fondée. (DIAZ et WOLVERTON 1998 ; GALLIMORE 1996 ; CROSBY, LIZIERI et MCALLISTER 2010) La pratique moderne de l'évaluation intègre de plus en plus les outils statistiques, les considérations de scénarios et l'analytique du risque — en particulier dans l'évaluation de masse, les modèles d'évaluation automatisés ou le travail de portefeuille — mais le livrable central demeure typiquement une opinion de valeur de marché à un instant donné. (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 et NEAL 2023)

On pourrait caractériser la pratique de l'évaluation comme essentiellement probatoire et orientée vers des estimations ponctuelles allant du passé au présent (ou prospectives dans des limites étroites), tandis que l'analyse financière et quantitative est plus probabiliste, distributionnelle et explicitement tournée vers l'avenir sur divers horizons temporels. (GELTNER 1991) Les experts demandent : « Quelle est l'indication de valeur de marché la mieux étayée à la date d'effet spécifiée, compte tenu des éléments actuels et des hypothèses définies ? » Les analystes financiers et quantitatifs demandent plus souvent : « Quel éventail de résultats est probable sur une période future, avec quelles probabilités, et comment les ressources devraient-elles être allouées ? »

La frontière n'est pas absolue et continue d'évoluer avec la complexité des marchés. Des différences fondamentales persistent néanmoins quant à la finalité (valeur de marché contre valeur d'investissement), aux exigences probatoires et aux cadres réglementaires.

La ligne opérante, toutefois, n'est pas tracée par la compétence mais par le droit : la question déterminante est simplement de savoir si une tâche donnée est de celles pour lesquelles la

licence d'expert est légalement requise. Les analystes quantitatifs, les analystes financiers, les économistes, les analystes en investissement, les gestionnaires de portefeuille et les professionnels apparentés peuvent légalement estimer, prévoir, modéliser ou analyser des valeurs d'actifs futures — y compris des projections distributionnelles ou par fourchettes — sans licence d'expert immobilier, pourvu que le travail ne constitue pas une évaluation réglementée. (MURPHY 2012) Les distinctions essentielles reposent sur :

1. le type de valeur (par exemple, une estimation ponctuelle de valeur de marché contre une valeur d'investissement ou une prévision probabiliste) ;
2. le point focal temporel (une date d'effet précise contre une période ou un horizon) ;
3. la forme du résultat (un point ou une fourchette étroite contre une distribution de probabilité complète) ;
4. la finalité visée et le fait que le travail soit ou non présenté comme une évaluation (*appraisal*) ; et
5. l'implication de contextes de prêt réglementés ou de contextes juridiques exigeant la conformité à des normes telles que les USPAP.

Ce cadre est largement issu de réglementations bancaires telles que FIRREA (1989) aux États-Unis, qui visait à garantir des opinions de valeur de marché crédibles et indépendantes pour les transactions relevant du niveau fédéral (*federally related transactions*). (UNITED STATES CONGRESS 1989 ; K'AKUMU et LARSEN 2024) De nombreuses activités d'évaluation sophistiquées — objectifs de cours en recherche actions, études de faisabilité, modélisation du risque de portefeuille, valorisation d'options ou analytique prédictive — échappent à l'obligation de licence parce qu'elles servent des finalités d'investissement, de prévision ou de conseil plutôt que des fonctions d'évaluation réglementée.

```
%% appraisal_scope_regulation.pl
%%
%% Knowledge base for the boundary of regulated appraisal practice.
%%
%% TWO TIERS, deliberately kept separate:
%%
%% TIER 1 -- THE DEFINING TEST (regulation cluster). A task is a
%% regulated appraisal IFF it is one for which an appraiser's
%% license is legally required. requires_license/1 is that test;
%% it -- and only it -- decides regulated_appraisal/1 vs.
%% unregulated_valuation/1.
%%
%% TIER 2 -- DESCRIPTIVE TENDENCIES (scope cluster). How the two
%% populations -- `appraisal' and `quant_financial' -- typically
%% differ in approach, temporal scope, output form, and skill set.
%% These are CORRELATES, not the boundary: the populations overlap
%% (a licensed appraiser may do quant work; a quant may be
%% unlicensed), and nothing here determines licensure. No claim is
%% made about intelligence -- only about training and methodology.
%%
%% `temporal scope' and `approach' are DISTINCT axes: an approach has a
%% typical (defeasible) temporal mode, but the two are not the same
%% viewpoint -- which is why an appraiser using an AVM (appraisal
%% approach, distributional methods) is possible.

:- module(appraisal_scope_regulation,
    [ % --- TIER 1: the defining test (regulation) ---
      regulation_framework/1
    , regulation_framework_origin/2
    , regulation_distinguishing_factor/1
    , regulation_triggers_licensure/1
```

```

, regulation_definition_of_appraisal/1
, regulation_unlicensed_activity/2
, regulation_licensed_activity/2
, regulation_legal_exposure_condition/1
, requires_license/1
, exempt_from_appraisal_license/1
, regulated_appraisal/1
, unregulated_valuation/1
, regulatory_asymmetry/2

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

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

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

regulation_framework(firrea_1989).
regulation_framework(uspap).
regulation_framework(ivs).
regulation_framework(state_licensing).
regulation_framework(state_certification).

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

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

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

```

```

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

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

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

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

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

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

:- dynamic activity_satisfies_definition/1.

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

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

unregulated_valuation(Task) :-

```

```

    regulation_unlicensed_activity(_, Task),
    \+ regulated_appraisal(Task).

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

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

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

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

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

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

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

temporal_mode(point_in_time).
temporal_mode(distribution_over_horizon).

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

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

```

```

scope_output_form(quant_financial, probability_distribution).
scope_output_form(quant_financial, range_of_outcomes).

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

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

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

%% --- Bounded future-oriented contexts that remain within appraisal ---
scope_future_oriented_context(income_property, future_income_streams).
scope_future_oriented_context(income_property, vacancy_rates).
scope_future_oriented_context(income_property, expenses).
scope_future_oriented_context(income_property, reversion_values).
scope_future_oriented_context(income_property, dcf_analysis).
scope_future_oriented_context(development, prospective_stabilized_value).
scope_future_oriented_context(conservation, future_damages).
scope_future_oriented_context(environmental, future_damages).
scope_future_oriented_context(insurance, future_damages).
scope_future_oriented_context(litigation, future_damages).
scope_future_oriented_context(eminent_domain, future_damages).
scope_future_oriented_context(general, remaining_economic_life).
scope_future_oriented_context(general, anticipated_market_reactions).
scope_future_oriented_context(market_analysis, price_trend_direction).

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

%% --- Contrasts across the descriptive axes:
↪ scope_contrast(A,B,Dimension,Distinction) ---
scope_contrast(appraisal, quant_financial, orientation,

```

```

        evidentiary_vs_probabilistic).
scope_contrast(appraisal, quant_financial, temporal_mode,
               point_in_time_vs_distribution_over_horizon).
scope_contrast(appraisal, quant_financial, output_form,
               point_or_narrow_range_vs_probability_distribution).
scope_contrast(appraisal, quant_financial, purpose,
               market_value_vs_investment_worth).
scope_contrast(appraisal, quant_financial, skill_set,
               appraisal_licensing_toolkit_vs_advanced_statistical_toolkit).
scope_contrast(appraisal, quant_financial, typical_methods,
               sales_comparison_cost_income_vs_simulation_option_pricing_var).

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

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

```

## 5 La valeur de marché × LTV comme substitut de la valeur de saisie

Les sections précédentes ont exposé à quoi sert l'évaluation, ce qu'elle produit et où se situent ses limites doctrinales et réglementaires; la présente section atteint la tension vers laquelle elles convergeaient.

L'une des réalités pratiques profondes qui sous-tendent le crédit hypothécaire et l'analyse des garanties est que, dans sa fonction économique réelle, une évaluation de valeur de marché utilisée dans la souscription hypothécaire sert souvent de substitut implicite de la recouvrabilité attendue de la garantie sur une période de détention future, alors même que, formellement, l'évaluation est présentée comme une opinion de valeur de marché présente à une date d'effet précise. (FOSTER et VAN ORDER 1984; KAU et al. 1995; DENG, QUIGLEY et VAN ORDER 2000)

Un ratio prêt-valeur de 80 % n'a rien d'arbitraire. Historiquement, il s'est en partie constitué comme un coussin de risque destiné à protéger les prêteurs contre :

- des baisses modérées du marché,
- les frais de saisie,
- les décotes de liquidation,
- les coûts de portage,
- les frais juridiques,
- les intérêts courus,
- et l'incertitude pendant le temps nécessaire pour céder la garantie après un défaut. (FOOTE, GERARDI et WILLEN 2008; QI et YANG 2009)

Ce coussin explique aussi pourquoi le montant du prêt lui-même — le produit de la valeur de marché par le ratio prêt-valeur autorisé — peut être lu comme l'estimation de travail, par le prêteur, de la valeur de saisie recouvrable : plafonner le prêt à une fraction de la valeur de marché revient, en pratique, à parier que le produit d'une vente forcée ne tombera pas en dessous

de l'encours du prêt. Quelle *devrait* être l'ampleur de la décote est toutefois bien plus difficile à cerner que ne le suggère le chiffre rond de 80 %. Les estimations de la décote de saisie varient fortement selon les marchés, les méthodes et les périodes ; une étude fréquemment citée sur les ventes forcées la situe à près de 27 % en dessous des prix comparables de transactions librement négociées (CAMPBELL, GIGLIO et PATHAK 2011), tandis que d'autres analyses — corrigeant l'état du bien, la localisation et les effets de sélection — font état d'ordres de grandeur sensiblement différents. (PENNINGTON-CROSS 2006 ; CLAURETIE et DANESHVARY 2009)

Le seuil conventionnel de 80 %, en revanche, est pour l'essentiel un paramètre national unique, appliqué à des marchés locaux extrêmement hétérogènes et aux phases successives du cycle du crédit. Il ne peut suivre la décote d'aucun marché particulier et fonctionne donc comme une approximation très grossière et uniforme plutôt que comme une correspondance précise entre valeur de marché et recouvrement en cas de saisie. Et c'est plutôt une approximation mince : une décote de 20 % se situe *en dessous* de la décote de 27 % rapportée par cette étude, de sorte que le prêt conventionnel maximal peut excéder le produit brut d'une vente forcée avant même la soustraction des coûts de cession énumérés ci-dessus.

Lorsque l'apport de l'emprunteur est trop mince pour constituer le coussin, l'assurance hypothécaire s'y substitue. Cela se comprend moins comme un curseur de politique publique que comme un arbitrage d'allocation du risque : un segment d'emprunteurs est pauvre en actifs mais stable en revenus, et le risque incrémental de défaut et de perte lié aux prêts qui leur sont consentis au-delà de 80 % de LTV est prévu et supporté par un assureur — de manière privée, par l'assurance hypothécaire privée sur un prêt conventionnel à LTV élevé, ou de manière publique, par la Federal Housing Administration (administration fédérale du logement). Le marché sur lequel cet arbitrage est de routine est toutefois une création de la politique fédérale du logement : les entreprises paragonnementales (*government-sponsored enterprises*) qui dominent le crédit conventionnel ont l'interdiction statutaire d'acquiescer des prêts au-delà de 80 % de LTV sans un tel rehaussement de crédit (UNITED STATES CODE 1970). Les conditions auxquelles l'assurance prend fin relèvent, de même, de choix institutionnels plutôt que de calculs de risque, et elles diffèrent nettement d'un programme à l'autre.<sup>1</sup>

Ainsi, bien que l'expert estime officiellement la « valeur de marché actuelle », le prêteur interprète économiquement cette valeur à travers un prisme de risque prospectif. En pratique, le prêteur pose une question qui s'apparente plutôt à :

« Si cet emprunteur fait défaut à un moment quelconque dans un avenir prévisible, la garantie restera-t-elle vraisemblablement suffisante pour couvrir la dette restante et les pertes liées à la saisie ? »

C'est là, par nature, une question tournée vers l'avenir.

Cela repose sur une séparation conceptuelle délibérée entre :

- l'opinion de « valeur de marché actuelle » de l'expert, et
- l'usage pratique que le prêteur fait de cette opinion comme coussin prévisionnel contre une dépréciation future de la garantie.

Historiquement et juridiquement, les régulateurs se sont efforcés d'empêcher les experts de devenir des prévisionnistes spéculatifs parce que :

- la prévision spéculative avait contribué aux crises de crédit passées,
- les tribunaux voulaient des évaluations rattachées à des éléments observables,

---

1. L'assurance hypothécaire privée conventionnelle payée par l'emprunteur prend fin automatiquement à 78 % de LTV et peut être résiliée sur demande à 80 % en vertu du Homeowners Protection Act de 1998. (UNITED STATES CONGRESS 1998) Les primes de la Federal Housing Administration sont plus strictes : pour les prêts émis à compter du 3 juin 2013 avec un ratio prêt-valeur initial supérieur à 90 % (par exemple, 3,5 % d'apport), la prime annuelle court sur toute la durée du prêt, tandis que les prêts à 90 % ou moins la voient s'éteindre après onze ans. (U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT 2013) De nombreux emprunteurs FHA ne se défont donc de l'assurance hypothécaire qu'en se refinançant dans un prêt conventionnel proche de 80 % de LTV.

- et les régulateurs bancaires recherchaient des pratiques d'évaluation normalisées et défendables.

La doctrine de l'évaluation a donc évolué vers :

- la valeur présente,
- les éléments de marché actuels,
- des dates d'effet définies,
- et des hypothèses encadrées.

Pendant ce temps, prêteurs, investisseurs et analystes de garanties utilisent couramment ces mêmes évaluations dans des cadres probabilistes de risque futur.

En réalité, la souscription hypothécaire est un système intégré qui fait intervenir :

1. l'évaluation du marché actuel,
2. l'estimation du risque futur,
3. la probabilité de défaut,
4. le calendrier de la saisie,
5. la sévérité de la liquidation,
6. les anticipations macroéconomiques,
7. et l'analyse de tension du portefeuille.

L'expert contribue formellement pour l'essentiel au point (1), tandis que le prêteur combine en interne ce résultat avec les autres.

La distinction devient floue en pratique. Si un prêteur n'accordait pas un prêt sans être convaincu que la garantie conserverait vraisemblablement une valeur suffisante dans l'avenir, alors l'évaluation fonctionne indirectement comme un élément d'un modèle de protection de la valeur future, même si les normes d'évaluation évitent formellement de la décrire ainsi.

Cette tension s'est encore accentuée avec :

- les modèles d'évaluation automatisés (AVM),
- l'analytique institutionnelle de portefeuille,
- les modèles de titrisation,
- les systèmes d'évaluation par apprentissage automatique,
- la modélisation du capital selon Bâle,
- et les tests de résistance hypothécaires à grande échelle.

L'analytique moderne des garanties traite de plus en plus les valeurs immobilières de manière probabiliste dans le temps plutôt que comme des estimations statiques ponctuelles. (BASEL COMMITTEE ON BANKING SUPERVISION 2017 ; STEURER, HILL et PFEIFER 2021 ; BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM 2021)

On pourrait soutenir que la réglementation traditionnelle de l'évaluation reflète un compromis historique :

- les experts fournissent des ancrs de valeur présente disciplinés et fondés sur des éléments probants,
- tandis que les institutions financières assument la responsabilité de l'interprétation du risque futur.

Savoir si cette séparation demeure pleinement cohérente dans les systèmes financiers modernes est une question légitime et de plus en plus débattue.

Il en découle une implication pratique pour la pratique de l'évaluation elle-même. La plupart des experts n'ont pas conscience de ce rôle d'évaluation de la garantie — l'usage que fait le prêteur de la valeur de marché comme coussin prospectif est rarement visible de l'intérieur de la mission d'évaluation — et ne traitent donc pas leurs conclusions de valeur de marché avec la prudence à l'égard des tendances futures de la valeur que ce rôle exige. Reconnaître la manière dont l'opinion de valeur présente est réellement consommée est un premier pas vers la réduction de cet écart, et un champ que le présent exposé entend ouvrir à des travaux ultérieurs.

```

%% 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 Conclusion

Les cinq passages encodés dans les bases de connaissances qui précèdent décrivent une même profession depuis cinq points de vue complémentaires. Pris isolément, chaque passage se lit comme un exposé autonome : les enjeux sociaux et d'intérêt public qu'ils portent, les fonctions économiques que soutiennent les évaluations, leur rôle opérationnel dans l'analyse du risque de garantie, la frontière réglementaire qui les sépare du travail quantitatif adjacent, et la tension entre leur présentation formelle et leur usage pratique. Pris ensemble, et avec la structure rendue explicite en Prolog, ils décrivent quelque chose de plus intéressant — une profession dont la doctrine officielle et la fonction économique ont été délibérément tenues à distance l'une de l'autre, et dont la cohérence interne est de plus en plus mise à l'épreuve par les outils analytiques désormais appliqués à ses produits.

Les trois premières bases de connaissances partagent un schéma commun : `effect(Quality, Horizon, Domain, Effect)`. Chacune présente la qualité de l'évaluation comme un intrant causal dont les conséquences se propagent aux horizons de court, moyen et long terme, ne différant que par le domaine à travers lequel ces conséquences sont lues. Le prisme économique les lit à travers les marchés du crédit, l'allocation des investissements, la taxation et la formation du capital ; le prisme social les lit à travers l'équité, la confiance du public et le devenir des collectivités ; le prisme de la garantie les lit à travers la souscription, la gestion de portefeuille et la stabilité du système financier. Ce schéma partagé n'est pas une coïncidence. Il reflète une thèse substantielle qui traverse les trois passages : la qualité du travail d'évaluation a des conséquences, ces conséquences s'amplifient avec le temps, et les mêmes modes de défaillance sous-jacents se manifestent différemment selon les intérêts mis au premier plan. Les atomes partagés du côté des arguments des prédicats — les vocabulaires de qualité et d'horizon en particulier — rendent les trois vocabulaires joignables, de sorte qu'une défaillance unique telle qu'une surévaluation systématique peut être suivie simultanément à travers la mauvaise valorisation économique, l'iniquité sociale et la mesure erronée des garanties.

La quatrième base de connaissances change de registre. Elle ne décrit pas les conséquences de la qualité de l'évaluation ; elle décrit ce que l'évaluation *est*, méthodologiquement et juridiquement. Ses prédicats sont taxonomiques plutôt que causaux, et sa machinerie inférentielle répond à des questions de compétence plutôt que d'impact. Ce déplacement est nécessaire parce que le cadrage conséquentialiste des trois premiers passages présuppose une définition de l'activité évaluée, et que cette définition est elle-même contestée. Les prédicats de périmètre consignent ce que font les experts — des opinions de valeur présente fondées sur des éléments probants, avec des composantes prospectives étroitement bornées — tandis que les prédicats de réglementation consignent les conditions dans lesquelles cette activité relève de la pratique agréée. Les règles de chaînage avant (`requires_license/1`, `exempt_from_appraisal_license/1`, `regulated_appraisal/1`) fournissent un mécanisme pour demander si un travail d'évaluation donné se situe à l'intérieur ou à l'extérieur du périmètre réglementé, question que les trois premières bases de connaissances supposent déjà tranchée.

La cinquième base de connaissances avance une thèse. Là où les autres décrivent, classent ou tracent des conséquences, la base de connaissances du substitut (*proxy*) soutient que la séparation doctrinale soigneuse entre l'évaluation en valeur présente et la prévision de valeur future — la séparation que documente la base de connaissances réglementaire — est en tension avec la fonction économique réelle des évaluations dans le crédit hypothécaire. La convention du ratio prêt-valeur de 80 %, les sept risques spécifiques qu'elle amortit, la question sous-jacente du prêteur consignée verbatim (« Si cet emprunteur fait défaut à un moment quelconque dans un avenir prévisible, la garantie restera-t-elle vraisemblablement suffisante... »), et le système

intégré de souscription à sept composantes soutiennent tous la même thèse : l'évaluation est formellement une opinion de valeur présente et économiquement un coussin de valeur future. Le compromis historique qui a produit cet arrangement — les experts comme ancrés disciplinés de la valeur présente, les prêteurs comme siège de l'interprétation du risque futur — fut une réponse défendable aux conditions dans lesquelles l'agrément des experts a émergé, mais les intensificateurs modernes énumérés dans le passage (AVM, systèmes d'évaluation par apprentissage automatique, modélisation du capital selon Bâle, analytique de titrisation, tests de résistance à grande échelle) traitent les valeurs immobilières de manière probabiliste dans le temps. Ces outils ramènent des couches doctrinalement séparées dans une chaîne de calcul unique, et la séparation conceptuelle entre couches devient d'autant plus difficile à maintenir.

Savoir si cette réconciliation est réalisable, et à quelles conditions, est la question sur laquelle se clôt le passage du substitut. Les bases de connaissances n'y répondent pas. Elles fournissent en revanche un exposé explicite et interrogeable des considérations pertinentes pour y répondre : les conséquences de la qualité de l'évaluation selon les horizons et les domaines, la frontière de compétence qui détermine quelles conséquences relèvent de la pratique réglementée, et les raisons structurelles pour lesquelles cette frontière est aujourd'hui sous pression. Cet exposé est le socle sur lequel des arguments plus spécifiques — sur la méthodologie, sur la réforme des normes, sur la relation entre la pratique de l'évaluation et le système plus large de l'évaluation quantitative — peuvent être construits.

## Notes sur la représentation des connaissances

Trois régularités structurelles méritent d'être consignées explicitement, car elles ont émergé du processus d'encodage plutôt que d'un passage particulier.

*Premièrement, la famille de bases de connaissances n'est pas plate.* Trois des cinq partagent un schéma ; une quatrième fournit les qualifications de compétence qui déterminent quand ce schéma s'applique ; une cinquième avance une thèse sur l'écart entre le cadrage formel et la fonction économique. Une couche de liaison combinant les cinq ne les traite pas comme cinq colonnes d'une même matrice. Elle traite les trois premières comme décrivant des conséquences, la quatrième comme délimitant le champ où ces descriptions sont opérantes, et la cinquième comme identifiant le point où la contrainte elle-même est sous pression. Cette structure asymétrique est une qualité et non un défaut : elle reflète l'organisation réelle de la pratique professionnelle, dans laquelle les énoncés descriptifs, les énoncés de compétence et les énoncés critiques occupent des rôles épistémiques véritablement différents.

*Deuxièmement, la discipline de préfixage qui distingue les modules (*economic\_*, *social\_*, *collateral\_*, *scope\_/regulation\_*, *proxy\_/underwriting\_*) ne fait un travail réel que là où existe un risque de collision.* Les prédicats qui ont des homologues directs dans d'autres modules méritent leur préfixe ; les prédicats propres à un seul domaine conceptuel non. Le vocabulaire de valeurs partagé du côté des arguments — les atomes de qualité, d'horizon et de domaine — est précisément ce qui permet aux modules de communiquer, et préfixer ces valeurs irait à l'encontre de ce but. La régularité qui s'est dégagée est générale : préfixe sur les prédicats, atomes partagés sur les arguments, avec un préfixage sélectif plutôt qu'uniforme lorsque le prédicat est propre à un seul prisme.

*Troisièmement, la tension centrale identifiée dans la base de connaissances du substitut est étroitement liée à l'asymétrie réglementaire identifiée dans la base de connaissances périmètre/réglementation, tout en s'en distinguant.* Les deux sont consignées au moyen de prédicats binaires parallèles — *regulatory\_asymmetry/2* et *proxy\_doctrine\_practice\_gap/2* — portant des noms délibérément différents parce qu'ils décrivent des axes différents d'un même arrangement sous-jacent. L'asymétrie réglementaire concerne qui est réglementé et avec quelle intensité ; l'écart entre doctrine et pratique concerne ce qui est formellement dit du produit de l'expert par opposition à ce qui en est économiquement fait. Les deux tensions se renforcent mutuellement. La réglementation pousse les experts vers la doctrine de la valeur présente ; la

fonction économique tire leur production vers un usage en valeur future. Le tableau d'ensemble est celui d'une profession dont l'autodescription officielle est de plus en plus difficile à concilier avec les chaînes analytiques qu'alimentent ses produits.

## Références

- AGARWAL, Sumit, Itzhak BEN-DAVID et Vincent YAO (2015). « Collateral Valuation and Borrower Financial Constraints: Evidence from the Residential Real Estate Market ». In : *Management Science* 61.9, p. 2220-2240. DOI : [10.1287/mnsc.2014.2002](https://doi.org/10.1287/mnsc.2014.2002).
- ALEXANDROV, Alexei, Laurie GOODMAN et Michael NEAL (jan. 2023). *Reengineering the Appraisal Process: Better Leveraging Both Automated Valuation Models and Manual Appraisals*. Rapp. tech. 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 (mars 2021). *Dodd-Frank Act Stress Test 2021: Supervisory Stress Test Methodology*. Rapp. tech. 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 et Parag PATHAK (2011). « Forced Sales and House Prices ». In : *American Economic Review* 101.5, p. 2108-2131. DOI : [10.1257/aer.101.5.2108](https://doi.org/10.1257/aer.101.5.2108).
- CLAURETIE, Terrence M. et Nasser DANESHVARY (2009). « Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time ». In : *Real Estate Economics* 37.1, p. 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 et Patrick M. MCALLISTER (2010). « Means, Motive and Opportunity? Disentangling Client Influence on Performance Measurement Appraisals ». In : *Journal of Property Research* 27.2, p. 181-201. DOI : [10.1080/09599916.2010.499014](https://doi.org/10.1080/09599916.2010.499014).
- DENG, Yongheng, John M. QUIGLEY et Robert VAN ORDER (2000). « Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options ». In : *Econometrica* 68.2, p. 275-307. DOI : [10.1111/1468-0262.00110](https://doi.org/10.1111/1468-0262.00110).
- DIAZ Julian, III et Marvin L. WOLVERTON (1998). « A Longitudinal Examination of the Appraisal Smoothing Hypothesis ». In : *Real Estate Economics* 26.2, p. 349-358. DOI : [10.1111/1540-6229.00749](https://doi.org/10.1111/1540-6229.00749).
- EBRAHIM, M. Shahid et Sikandar HUSSAIN (2010). « Financial Development and Asset Valuation: The Special Case of Real Estate ». In : *Journal of Banking & Finance* 34.1, p. 150-162. DOI : [10.1016/j.jbankfin.2009.07.011](https://doi.org/10.1016/j.jbankfin.2009.07.011).
- ERIKSEN, Michael D. et al. (2019). « The Influence of Contract Prices and Relationships on Appraisal Bias ». In : *Journal of Urban Economics* 111, p. 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 et Paul S. WILLEN (2008). « Negative Equity and Foreclosure: Theory and Evidence ». In : *Journal of Urban Economics* 64.2, p. 234-245. DOI : [10.1016/j.jue.2008.07.006](https://doi.org/10.1016/j.jue.2008.07.006).
- FOSTER, Chester et Robert VAN ORDER (1984). « An Option-Based Model of Mortgage Default ». In : *Housing Finance Review* 3.4, p. 351-372.
- FRIEDMAN, Jerome H. (1991). « Multivariate Adaptive Regression Splines ». In : *The Annals of Statistics* 19.1, p. 1-67. DOI : [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- GALLIMORE, Paul (1996). « Confirmation Bias in the Valuation Process: A Test for Corroborating Evidence ». In : *Journal of Property Research* 13.4, p. 261-273. DOI : [10.1080/095999196368781](https://doi.org/10.1080/095999196368781).
- GELTNER, David M. (1991). « Smoothing in Appraisal-Based Returns ». In : *The Journal of Real Estate Finance and Economics* 4.3, p. 327-345. DOI : [10.1007/BF00161933](https://doi.org/10.1007/BF00161933).

- GHENT, Andra C., Walter N. TOROUS et Rossen I. VALKANOV (2019). « Commercial Real Estate as an Asset Class ». In : *Annual Review of Financial Economics* 11, p. 153-171. DOI : [10.1146/annurev-financial-110118-123121](https://doi.org/10.1146/annurev-financial-110118-123121).
- GRIFFIN, John M. et Gonzalo MATURANA (2016). « Who Facilitated Misreporting in Securitized Loans? » In : *The Review of Financial Studies* 29.2. Evidence of appraisal inflation and its consequences, p. 384-419. DOI : [10.1093/rfs/hhv130](https://doi.org/10.1093/rfs/hhv130).
- HENDERSHOTT, Patric H. et Edward J. KANE (juill. 1992). *Office Market Values During the Past Decade: How Distorted Have Appraisals Been?* NBER Working Paper 4128. Available at SSRN: <https://ssrn.com/abstract=226786>. National Bureau of Economic Research. DOI : [10.3386/w4128](https://doi.org/10.3386/w4128). URL : <https://www.nber.org/papers/w4128>.
- (1995). « U.S. Office Market Values During the Past Decade: How Distorted Have Appraisals Been? » In : *Real Estate Economics* 23.2, p. 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<sup>e</sup> éd. 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](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. et James E. LARSEN (2024). « A Primer on Regulations and the Practice of Residential Property Appraisal ». In : *Journal of Real Estate Practice and Education*. Focus on regulatory and public interest aspects. DOI : [10.1080/15214842.2024.2377869](https://doi.org/10.1080/15214842.2024.2377869).
- KAU, James B. et al. (1995). « The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment ». In : *The Journal of Real Estate Finance and Economics* 11.1, p. 5-36. DOI : [10.1007/BF01097934](https://doi.org/10.1007/BF01097934).
- LACOUR-LITTLE, Michael et Stephen MALPEZZI (2003). « Appraisal Quality and Residential Mortgage Default: Evidence from Alaska ». In : *The Journal of Real Estate Finance and Economics* 27.2, p. 211-233. DOI : [10.1023/A:1024728420837](https://doi.org/10.1023/A:1024728420837).
- LEE, Ming-Hsiang, Chien-Wen PENG et Hui-Fen LIAO (2020). « An Analysis of Objectivity in the Real Estate Appraisal Process ». In : *International Real Estate Review* 23.4, p. 483-504.
- MAYER, Yanling G. et Frank E. NOTHAFT (2022). « Appraisal Overvaluation: Evidence of Price Adjustment Bias in Sales Comparisons ». In : *Real Estate Economics* 50.3, p. 862-881. DOI : [10.1111/1540-6229.12351](https://doi.org/10.1111/1540-6229.12351).
- MURPHY, Edward V. (2012). *Regulation of Real Estate Appraisers*. CRS Report RS22953. Congressional Research Service. URL : <https://sgp.fas.org/crs/misc/RS22953.pdf>.
- OFFICE OF THE COMPTROLLER OF THE CURRENCY AND BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM AND FEDERAL DEPOSIT INSURANCE CORPORATION AND NATIONAL CREDIT UNION ADMINISTRATION AND CONSUMER FINANCIAL PROTECTION BUREAU AND FEDERAL HOUSING FINANCE AGENCY (2024). *Quality Control Standards for Automated Valuation Models*. Final Rule, 89 Fed. Reg. 64538 (Aug. 7, 2024). URL : <https://www.federalregister.gov/documents/2024/08/07/2024-16197/quality-control-standards-for-automated-valuation-models>.
- PENNINGTON-CROSS, Anthony (2006). « The Value of Foreclosed Property ». In : *Journal of Real Estate Research* 28.2, p. 193-214. DOI : [10.1080/10835547.2006.12091177](https://doi.org/10.1080/10835547.2006.12091177).
- QI, Min et Xiaolong YANG (2009). « Loss Given Default of High Loan-to-Value Residential Mortgages ». In : *Journal of Banking & Finance* 33.5, p. 788-799. DOI : [10.1016/j.jbankfin.2008.09.010](https://doi.org/10.1016/j.jbankfin.2008.09.010).
- QUAN, Daniel C. et John M. QUIGLEY (1991). « Price Formation and the Appraisal Function in Real Estate Markets ». In : *The Journal of Real Estate Finance and Economics* 4.2, p. 127-146. DOI : [10.1007/BF00173120](https://doi.org/10.1007/BF00173120).

- STEURER, Miriam, Robert J. HILL et Norbert PFEIFER (2021). « Metrics for Evaluating the Performance of Machine Learning Based Automated Valuation Models ». In : *Journal of Property Research* 38.2, p. 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) : cadre méthodologique et protocole

*Traduction française de la version actualisée de l'article initialement publié sur Zenodo le 5 mars 2025. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

William Bert Craytor, SRA

Volume 1, Numéro 1 · Été 2026

Publié le 1<sup>er</sup> juillet 2026

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

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

Cet article en libre accès est diffusé selon les termes de la [licence Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/).

**Résumé.** Une limite importante des méthodes d'évaluation conventionnelles réside dans leur dépendance à des techniques manuelles dépassées, telles que l'analyse par paires appariées (*matched pairs*), qui ignorent les contributions à la valeur des caractéristiques du bien comme si elles n'existaient pas : la grille de comparaison traditionnelle (*Sales Grid*) ne comporte aucune colonne pour les contributions à la valeur, seulement des ajustements. Or, les contributions à la valeur fournissent la base nécessaire aux contraintes mathématiques qui préviennent la surévaluation et la sous-évaluation par les évaluateurs traditionnels.

Le Residual Constraint Approach (RCA, fr. : approche par contraintes résiduelles) fait progresser le Sales Comparison Approach (SCA, comparable à la méthode par comparaison française) traditionnel en intégrant un processus d'évaluation en plusieurs phases qui associe les Multivariate Adaptive Regression Splines (MARS) à des contraintes mathématiques rigoureuses sur les contributions à la valeur des caractéristiques — en particulier les caractéristiques non mesurées : des variables latentes telles que l'esthétique, la conception, l'état et la qualité, dont les valeurs sont généralement soumises au jugement subjectif et aux biais de l'évaluateur traditionnel.

Les contraintes mathématiques du RCA reposent sur l'application experte de la régression MARS pour estimer les contributions à la valeur des caractéristiques mesurables du bien au prix de vente. Cette estimation ne représente toutefois, en règle générale, qu'environ 80% du prix de vente réel, à l'exclusion des caractéristiques appréciées subjectivement comme la conception, l'utilité fonctionnelle, l'état et la qualité. Dans la San Francisco Bay Area, le résidu — qui représente ces composantes subjectives — s'élève en moyenne à environ 20% de la valeur totale du bien. Le ratio 80/20 reflète l'expérience de l'auteur dans la San Francisco Bay Area dans des conditions de données typiques et ne doit pas être interprété comme une constante invariante d'un marché à l'autre. Dans les marchés où le  $R^2$  atteignable est plus faible, la réponse du cadre méthodologique consiste à identifier et à mesurer les déterminants de la valeur non mesurés — la distance à l'océan dans les communes littorales, l'altitude dans les quartiers à flanc de colline et d'autres variables similaires collectées auprès de sources externes lorsque les données MLS ne les fournissent pas — plutôt que d'accepter cette limite et de publier des bandes d'erreur plus larges. Alors que les résidus MARS sont souvent traités comme des erreurs d'estimation dans d'autres contextes, ils servent, en évaluation immobilière, de mesure indirecte de la valeur des variables latentes. Bien que cela puisse d'abord sembler peu praticable, une démonstration établit que le résidu peut être décomposé de manière significative en composantes descriptives, pourvu que leurs contributions se somment au résidu — sans affecter le résultat final de l'évaluation.

Cette méthodologie est le fruit de deux décennies d'application empirique par l'auteur et a été développée au fil d'implémentations itératives de MARS dans les environnements R et Python.

**Mots-clés :** évaluation immobilière, évaluation résidentielle, Residual Constraint Approach, MARS, Multivariate Adaptive Regression Splines, Sales Comparison Approach, analyse des résidus, variables latentes, évaluation automatisée

## Table des matières

|          |                                                          |           |
|----------|----------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                      | <b>4</b>  |
| <b>2</b> | <b>Contexte</b>                                          | <b>6</b>  |
| <b>3</b> | <b>Valuation Engineer</b>                                | <b>9</b>  |
| <b>4</b> | <b>Hypothèses</b>                                        | <b>10</b> |
| <b>5</b> | <b>Workflow RCA</b>                                      | <b>12</b> |
| 5.1      | Types d'analyse MARS . . . . .                           | 12        |
| 5.2      | Workflow de traitement . . . . .                         | 12        |
| 5.3      | Workflow des données . . . . .                           | 13        |
| 5.3.1    | Exigences de données pour MARS . . . . .                 | 13        |
| 5.3.2    | Protocoles : listings de ventes conclues . . . . .       | 14        |
| 5.4      | Structure des dossiers de projet . . . . .               | 16        |
| 5.4.1    | Protocoles : exécutions MARS (earth) multiples . . . . . | 17        |
| 5.5      | Le classeur Excel central . . . . .                      | 19        |
| 5.6      | Sauvegardes . . . . .                                    | 19        |
| <b>6</b> | <b>Notation du classeur</b>                              | <b>20</b> |
| 6.1      | Dimensions descriptives . . . . .                        | 20        |
| 6.1.1    | La dimension projet (« a ») . . . . .                    | 21        |
| 6.1.2    | La dimension des versions d'exécution (« r ») . . . . .  | 22        |
| 6.1.3    | Étapes du flux de travail (« w ») . . . . .              | 22        |
| 6.1.4    | Feuilles du classeur Excel (« s ») . . . . .             | 23        |
| 6.1.5    | Listings de biens (« p ») - lignes de MLSData . . . . .  | 23        |
| 6.1.6    | Variables MLS (« v ») colonnes de MLSData . . . . .      | 24        |
| 6.1.7    | Exemple final de notation . . . . .                      | 24        |
| 6.1.8    | Justification . . . . .                                  | 24        |
| 6.1.9    | Matrices ? . . . . .                                     | 25        |
| 6.1.10   | Première simplification . . . . .                        | 25        |
| 6.1.11   | Seconde simplification . . . . .                         | 25        |
| <b>7</b> | <b>La Sales Grid RCA</b>                                 | <b>27</b> |

|                                                                                   |           |
|-----------------------------------------------------------------------------------|-----------|
| <b>8 Variables</b>                                                                | <b>31</b> |
| 8.1 Ontologie de l'origine des variables . . . . .                                | 31        |
| 8.2 Ontologie de l'application des variables . . . . .                            | 32        |
| 8.3 Termes du deuxième et du troisième degré . . . . .                            | 34        |
| 8.3.1 Variables d'ajustement . . . . .                                            | 34        |
| 8.3.2 Variables d'agrégation . . . . .                                            | 34        |
| 8.3.3 Variables spéciales & notation . . . . .                                    | 35        |
| 8.3.4 Décomposition des résidus . . . . .                                         | 35        |
| <b>9 Démonstrations fondamentales de la RCA</b>                                   | <b>37</b> |
| <b>10 Modèle MARS</b>                                                             | <b>39</b> |
| 10.1 Fonctions « $g$ » de variables du modèle et d'interaction . . . . .          | 39        |
| 10.2 Indexation des fonctions $g$ . . . . .                                       | 39        |
| 10.3 Dimensions de représentation graphique . . . . .                             | 40        |
| 10.4 L'équation finale du modèle . . . . .                                        | 40        |
| 10.5 Nombre de graphiques nécessaires pour représenter un modèle . . . . .        | 41        |
| <b>11 Caractéristiques des données</b>                                            | <b>44</b> |
| 11.1 Rôle du résidu . . . . .                                                     | 44        |
| 11.2 Les notations d'état et de qualité comme régresseurs approximatifs . . . . . | 45        |
| 11.3 Erreurs de données . . . . .                                                 | 45        |
| 11.4 Qu'est-ce que l'exactitude ? . . . . .                                       | 47        |
| 11.5 Le bien à évaluer comme anomalie . . . . .                                   | 48        |
| 11.6 L'importance du CVR2 face aux données inattendues . . . . .                  | 49        |
| <b>12 Résidus</b>                                                                 | <b>50</b> |
| 12.1 Ranking & Scoring (classement et notation) . . . . .                         | 50        |
| 12.2 Courbe CQA-résidu . . . . .                                                  | 53        |
| 12.3 Caractéristiques de la courbe CQA . . . . .                                  | 53        |
| 12.3.1 Analyse des composantes de la courbe CQA . . . . .                         | 53        |
| 12.3.2 Dynamique du marché et implications sociales . . . . .                     | 54        |
| 12.4 Résidu du bien à évaluer . . . . .                                           | 54        |
| 12.5 Conclusion de valeur RCA . . . . .                                           | 56        |
| 12.5.1 Considérations générales . . . . .                                         | 56        |
| 12.5.2 Révision, audit et réclamations . . . . .                                  | 56        |
| <b>13 Conclusion</b>                                                              | <b>58</b> |

## 1 Introduction

Cet article présente les fondements logiques et mathématiques d'une approche moderne visant à estimer avec exactitude la valeur de marché (valeur vénale) des biens immobiliers résidentiels et d'autres types de biens. La méthode, élaborée au cours de 20 années d'expérience dans l'application du logiciel Multivariate Adaptive Regression Splines (MARS) à des maisons du nord de la Californie — en particulier dans les communes de la San Francisco Bay Area — constitue un progrès significatif en matière de précision des évaluations. MARS améliore la qualité des évaluations en captant habilement les relations non linéaires entre les caractéristiques du bien et la valeur, tout en identifiant et en hiérarchisant efficacement les variables les plus influentes. Cette capacité permet aux évaluateurs de mettre au jour la dynamique de la valeur dans de nouvelles zones de marché bien plus rapidement que ne le permettent les méthodes traditionnelles. MARS constitue l'ossature de la Residual Constraint Approach (RCA) (fr. : approche par contraintes résiduelles) proposée ici. La méthode RCA doit son nom à la rigueur de son cadre :

- Elle impose des contraintes strictes sur les contributions à la valeur des caractéristiques du bien, tant mesurées que non mesurées :
  1. Toutes les contributions à la valeur dérivées de MARS, y compris le résidu, doivent avoir pour somme le prix de vente net de chaque bien comparable.
  2. Le résidu peut être décomposé en sous-composantes dotées de valeurs assignées, à condition que leur total soit égal au résidu du bien comparable.
- La RCA minimise le biais dans l'estimation des caractéristiques subjectives — telles que l'attrait esthétique, l'état et la qualité — en liant leur valeur totale au résidu, lequel est quantifié avec exactitude lorsqu'un Valuation Engineer (VE) compétent construit le modèle MARS.

Bien que principalement conçue pour l'immobilier résidentiel, où les caractéristiques subjectives influencent sensiblement la valeur de marché, la RCA présente un potentiel d'application à l'évaluation d'autres catégories d'actifs. Des rapports d'évaluation exacts revêtent une importance critique au regard de la valeur estimée du marché immobilier résidentiel américain, soit environ 45 billions de dollars (\$45 trillion). Fannie Mae relève que les évaluations typiques s'écartent d'environ 5% du prix de vente effectif, encore que ce chiffre masque un biais, les évaluateurs subissant souvent des pressions pour s'aligner sur les prix de vente. En outre, dès lors que 60%–70% des rapports d'évaluation servent à des refinancements — où l'on ne dispose pas de prix de vente comparables — la nécessité d'une précision telle que celle offerte par la RCA devient manifeste. L'analyse conduite par la Réserve fédérale en 2012 estimait que 15%–20% des évaluations antérieures à la crise de 2008 s'écartaient de plus de  $\pm 10\%$  des prix de vente, une fourchette vraisemblablement encore pertinente aujourd'hui pour les évaluations de refinancement, variable selon le type de bien et le marché. Pourquoi l'exactitude importe-t-elle ? Pour les transactions à faible ratio prêt-valeur (loan-to-value, LTV) — par exemple, un acheteur au dossier de crédit solide versant un apport de \$500,000 sur une maison de \$1 million — les prêteurs peuvent renoncer au rapport d'évaluation, en supposant un risque de saisie minimal. Toutefois, les acheteurs qui se passent d'évaluation risquent de payer trop cher et de ne découvrir les défauts qu'après l'achat. Cela souligne l'intérêt de méthodes d'évaluation robustes telles que la RCA. Quantifier l'exactitude des évaluations demeure difficile. Les chercheurs ne disposent pas de mesures directes des écarts par rapport à la « valeur de marché » et s'appuient sur des indicateurs indirects. Les méthodes d'évaluation traditionnelles, fondées sur le jugement subjectif et sur une compétence variable des évaluateurs, compliquent l'analyse. Parmi les stratégies employées :

1. TZIOUMIS (2016) a étudié 4,782 évaluateurs et 259,436 rapports d'évaluation dans 11 États, constatant qu'un tiers se situait à moins de 1% du prix contractuel, que la moitié se situait

entre 1% et 3% au-dessus, et qu'une minorité se situait en moyenne 5% au-dessus — ce qui suggère un biais à la hausse. L'étude a identifié deux types d'évaluateurs : ceux « à la valeur de marché » (proches du prix contractuel) et ceux « au-dessus de la valeur de marché » (systématiquement plus élevés).

2. KRIVOROTOV et LACOUR-LITTLE (2020) a constaté que 90% des évaluations d'achat atteignaient ou dépassaient le prix contractuel, mettant en évidence le « biais d'ancrage » — la tendance à s'aligner sur une valeur souhaitée — en particulier dans les contextes de refinancement.

Ces études révèlent un biais dans les pratiques d'évaluation, amplifié par les limites des outils traditionnels. Si les évaluateurs étaient uniformément compétents et exempts de biais, leurs évaluations convergeraient ; la surévaluation persistante de la part de certains groupes signale des problèmes systémiques.

Les études sur le biais citées ci-dessus, et le schéma systémique qu'elles mettent en évidence, reflètent un trait plus profond de la pratique traditionnelle de comparaison des ventes que la RCA est conçue pour corriger. La méthodologie traditionnelle opère sur deux canaux au statut épistémique très différent. Sur le canal des caractéristiques mesurables — GLA (surface habitable brute), nombre de pièces, superficie du terrain, âge, localisation — l'évaluateur est censé dériver les valeurs d'ajustement de manière objective, par analyse de paires appariées ou par régression linéaire locale sur les ventes comparables. Les évaluateurs compétents effectuent ce travail, et les ajustements qui en résultent sont au moins approximativement défendables au regard des données. Même là, cependant, la grille traditionnelle ne consigne que des ajustements ; elle ne comporte même pas de colonne pour les contributions à la valeur. Les ajustements traditionnels ne sont pas non plus calculés comme des différences entre contributions à la valeur — ils sont estimés directement, à partir de paires appariées ou de procédures analogues, sans que les contributions sous-jacentes ne soient jamais calculées. Dans la RCA, au contraire, chaque ajustement est la différence entre les contributions à la valeur du bien à évaluer et du bien comparable pour la caractéristique considérée — la structure dont la Démonstration II précise les conséquences. Sur le canal des caractéristiques subjectives — état, qualité de construction, conception, utilité fonctionnelle, esthétique — les procédures de justification disponibles vont du ténu à l'inconsistant. Les ajustements relatifs aux caractéristiques subjectives peuvent être étayés par une analyse de paires appariées portant sur des biens extérieurs à l'ensemble immédiat des comparables, par un renvoi à la pratique reconnue des pairs, ou par référence à l'expérience documentée de l'évaluateur sur le marché considéré. Ces trois procédés sont formellement conformes aux normes professionnelles. Aucun n'atteint la rigueur statistique que le canal mesurable atteint de manière routinière.

Il en résulte une asymétrie procédurale entre ce que les normes exigent et ce que la méthodologie peut fournir. Les normes exigent une justification pour chaque ajustement, qu'il soit mesurable ou subjectif. La méthodologie fournit une justification robuste sur un canal et un éventail de procédures plus faibles sur l'autre, allant d'une paire appariée anecdotique de taille d'échantillon un à la citation de ce que d'autres évaluateurs utilisent habituellement. L'évaluateur honnête et compétent qui effectue un travail soigné de paires appariées pour un ajustement de vue produit une justification sensiblement meilleure que celui qui invoque la pratique des pairs sans fondement, mais un réviseur comparant les deux rapports d'évaluation sur le formulaire standard ne peut aisément les distinguer, car le formulaire consigne la valeur de l'ajustement, et non la rigueur de sa dérivation. La méthodologie tolère cette asymétrie parce qu'aucune procédure plus solide pour les caractéristiques subjectives n'a été, de manière générale, à la disposition des évaluateurs en exercice.

La RCA améliore les deux canaux, mais de manières catégoriquement différentes. Sur le canal mesurable, la régression MARS capte des relations non linéaires, des interactions entre variables et une structure de fonctions de base que l'analyse de paires appariées et la régression linéaire

ordinaire ne peuvent représenter. Ajustées sur l'ensemble complet des ventes qualifiées de la zone de marché — typiquement plusieurs centaines — plutôt que sur les trois à neuf comparables de la grille traditionnelle, les contributions à la valeur dérivées de MARS sont à la fois objectives et sensiblement plus exactes que la méthodologie qu'elles remplacent. Sur le canal subjectif, la RCA fait quelque chose de différent en nature de toutes les procédures de justification traditionnelles. Elle ne tente aucunement d'estimer des valeurs d'ajustement pour les caractéristiques subjectives individuelles. Les contributions de toutes les caractéristiques subjectives — variables latentes — apparaissent collectivement sous la forme du résidu : la différence entre le prix de vente net effectif de chaque bien comparable et son prix prédit par MARS, calculable à partir des données sans intervention de l'évaluateur. La décomposition de ce résidu en composantes nommées — Conception, Utilité fonctionnelle, Vue, État, Qualité, etc. — est contrainte à avoir pour somme le résidu total, et, en vertu de la Démonstration II (Section 9), toute décomposition satisfaisant cette contrainte conduit à la même conclusion de valeur. Le jugement qui subsiste dans le flux de travail est le placement du bien à évaluer dans la distribution ordonnée des résidus des comparables, effectué au regard d'un classement inspectable et explicitement documenté dans le rapport. Un rapport d'évaluation RCA faible demeure possible — un VE inexpérimenté peut produire un modèle MARS surajusté, placer le bien à évaluer de manière peu plausible, ou manquer des anomalies dans les données — mais il est alors faible de manière qu'un réviseur peut identifier en inspectant le modèle et le classement, et non de manière qu'une revue sur formulaire standard ne saurait mettre au jour.

Au-delà du refinancement, les évaluateurs doivent aux propriétaires des évaluations exactes, en particulier pour les achats de grande valeur où les erreurs peuvent coûter des dizaines, voire des centaines de milliers de dollars. Pourtant, beaucoup rechignent à dépenser \$600–\$1,000 pour un rapport d'évaluation, doutant de sa valeur ajoutée. Les méthodes avancées telles que la RCA, quoique initialement exigeantes en ressources, promettent une précision accrue. Avec le perfectionnement des méthodes et l'expérience acquise, leur efficience — et leur adoption — devraient s'accroître.

## 2 Contexte

La formation universitaire de l'auteur relève des mathématiques, de l'informatique et de la statistique. De profession, il est ingénieur logiciel, avec une expérience secondaire en évaluation immobilière et en comptabilité. Il pratique l'évaluation immobilière, de manière intermittente, depuis 2002 et applique MARS depuis environ 2004 pour déterminer les ajustements de biens et les tendances de prix dans l'immobilier résidentiel.

Évaluer avec MARS constitue une « expérience approfondie » de l'évaluation immobilière. MARS génère des modèles qui reflètent, plus finement que toute autre méthode statistique, à la fois les variables qui contribuent à la valeur de marché (valeur vénale) et la manière dont les variations de ces valeurs influent sur le prix de vente. Ainsi, en évaluant avec MARS, l'évaluateur examine en permanence les relations causales entre les valeurs des variables et les prix de vente telles que générées par les modèles MARS.

Par souci de transparence, la méthodologie RCA (fr. : approche par contraintes résiduelles) s'applique de manière optimale aux zones résidentielles présentant les caractéristiques suivantes :

1. Des quartiers complexes, dont les biens diffèrent sensiblement les uns des autres
  - Maisons anciennes
  - Qualités de rénovation variées
  - Grande diversité d'architectures et de conceptions

- Variation des attributs des biens, tels que la GLA (surface habitable brute), la superficie du terrain, l'altitude et la vue.
- 2. Des conditions de marché complexes
  - Conditions de marché évoluant rapidement
  - Nombreux types d'acheteurs différents
- 3. De bonnes sources de données : le protocole RCA repose sur l'exploration de données (*data mining*) et requiert des données abondantes et de qualité. Bien qu'il puisse être utilisé dans des zones où les données sont rares, le temps et le coût nécessaires pourraient ne pas en valoir la peine.
- 4. Des estimations exactes de la valeur de marché sont recherchées. La RCA devrait, dans la plupart des cas, pouvoir fournir une exactitude de  $\pm 1-2\%$ . Le rapport du Valuation Engineer devrait commenter les conditions d'exactitude de la conclusion de valeur finale. Si les données ou les informations font défaut, ou si aucune régularité cohérente du marché n'a été mise au jour par l'exploration de données, l'exactitude en pâtira et le client devra être informé de la situation. Plus précisément, un  $R^2$  (coefficient de détermination) élevé et un CVR2 élevé ( $R^2$  validé de manière croisée, calculé en partitionnant de façon répétée le jeu de données MLS en partitions aléatoires d'entraînement et de test), un modèle MARS pertinent et une structure marquée des caractéristiques CQA (état, qualité, attrait) des biens dans les résidus classés constituent les principaux moyens d'apprécier l'exactitude et la fiabilité de la conclusion de valeur.

Le développement de ces protocoles RCA au cours des 20 dernières années et plus repose sur la disponibilité de MLSListings Inc. (Sunnyvale, CA) et d'autres MLS, actuellement accessibles via MLSListings, ce qui représente plus de 95% de l'expérience de l'auteur en matière d'analyse de données. L'expertise de l'auteur découle d'un travail d'évaluation approfondi mené dans 15 comtés principaux de Californie du Nord, complété par une expérience dans 10 comtés secondaires.

La méthodologie RCA est la plus efficace sur les marchés répondant à deux critères clés :

1. Une riche disponibilité de données d'évaluation de haute qualité
2. Des zones résidentielles complexes et hétérogènes, caractérisées par :
  - Des styles architecturaux variés
  - Des degrés divers de rénovation des biens
  - Des schémas de développement s'étalant sur plusieurs décennies

## Expérience des comtés californiens

Les affirmations formulées dans le présent article reposent sur l'expérience de l'évaluation de biens dans les comtés suivants de Californie du Nord :

|            |               |              |            |           |             |
|------------|---------------|--------------|------------|-----------|-------------|
| 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        |

Les zones ci-dessus recouvrent pour l'essentiel la Silicon Valley et les comtés environnants, en s'étendant vers des zones plus rurales, notamment la Central Valley et les contreforts de la Sierra. Il convient de mentionner que l'auteur a toujours eu accès à des données abondantes et

*relativement* exactes par rapport à celles dont disposent de nombreux évaluateurs dans d'autres régions moins développées des États-Unis.

Les données MLS utilisées sont certifiées RESO et conformes au RCF (RESO Common Format). Malgré la certification RCF, des données d'importance généralement secondaire sont parfois manquantes ou erronées pour certains biens, en raison d'erreurs de saisie initiales du Tax Assessor (administration d'évaluation fiscale) ou d'erreurs de saisie MLS ultérieures des agents immobiliers. Ces erreurs suivent souvent un schéma, tel que l'absence pure et simple de saisie lorsque la caractéristique n'existe pas. Par exemple, au lieu de saisir 0 pour « carport spaces » lorsqu'il n'y a pas de carport, rien n'est saisi. Dans les zones où les carports sont rares, l'analyste peut substituer 0 à l'absence de donnée afin d'obtenir un ajustement utile pour les carports. Bien que R/earth et d'autres implémentations de MARS puissent gérer les données manquantes, il appartient au Valuation Engineer (VE) de décider de la manière de les traiter.

Les données de MLSListings sont également disponibles auprès des MLS de certains autres États, de la Floride à Hawaï. Cela peut être utile pour tester le bon fonctionnement des protocoles et du code dans différents États.

### 3 Valuation Engineer

#### Définition 1 : Valuation Engineer

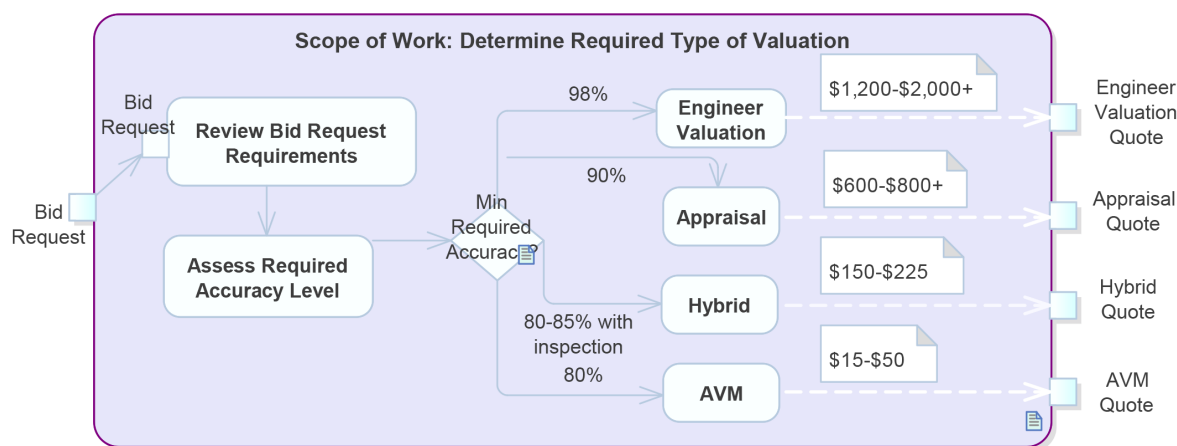
Afin d'éviter toute ambiguïté, le terme « Valuation Engineer » (ou « VE » en abrégé) désigne une personne possédant l'expertise d'un évaluateur agréé, associée à des connaissances, des compétences et une expérience avancées dans l'utilisation de l'exploration de données, de la statistique, de la programmation, de protocoles détaillés et de l'intelligence artificielle en vue de déterminer la valeur de marché de biens complexes. Il importe également de reconnaître que l'évaluation est une tâche qui nécessite une forme d'agrément en raison du potentiel avéré de fraude.

Les prêteurs proposent généralement plusieurs options d'évaluation :

1. Les modèles d'évaluation automatisée (*Automated Valuation Models*, AVM) - l'option la moins coûteuse
2. Les évaluations hybrides - recourant à des inspecteurs non agréés sous la supervision d'un évaluateur agréé
3. Les évaluations standard - réalisées par des évaluateurs agréés

Les évaluations de niveau ingénieur sont absentes des considérations habituelles en matière de travaux d'évaluation. Il existe tout simplement peu d'évaluateurs disposant des compétences nécessaires pour les réaliser. Le besoin devrait pourtant exister. Elles peuvent offrir une exactitude nettement supérieure et garantir presque une objectivité assez bonne grâce à l'emploi de la méthodologie RCA (MARS et contraintes).

Pour les biens conformes situés dans des lotissements récents, les évaluations standard suffisent généralement. En revanche, les évaluations d'ingénieur sont recommandées pour les maisons anciennes situées dans des zones présentant un historique de rénovation important s'étendant sur plus de 30 ans. Le choix entre ces méthodes devrait suivre un processus décisionnel structuré fondé sur les caractéristiques du bien et les conditions de marché.



## 4 Hypothèses

### Hypothèse 1 : Zone géographique & type de bien

Le champ du présent article porte principalement sur les maisons individuelles et les copropriétés (condominiums) des zones métropolitaines et urbaines de Californie. Bien que la Sales Comparison Approach (comparable à la méthode par comparaison française) s'applique à d'autres types de biens et à d'autres régions géographiques, les approches par le revenu et par le coût jouent généralement un rôle plus dominant lors de l'évaluation de terres agricoles, d'immeubles collectifs et d'immobilier commercial. Dans de tels cas, l'approche RCA peut nécessiter des modifications ou n'avoir qu'une applicabilité limitée.

### Hypothèse 2 : Stockage des données

Les formats CSV et Excel sont massivement utilisés par les systèmes MLS et sont supposés constituer le principal moyen d'alimenter en données les programmes RCA, bien que des flux de courtiers puissent être utilisés pour alimenter en continu des bases de données SQL. La sauvegarde et le stockage intermédiaire recourent souvent à des bases de données SQL, dont deux jouent un rôle majeur : SQLite et PostgreSQL, libres et open source. Des fichiers texte ou RTF sont également utilisés pour des fragments de rapport, et des fichiers PNG, JPG et TIFF pour les graphiques. Notons que SQL Server de Microsoft est également utilisé par beaucoup, mais l'édition gratuite « SQL Server Express » est très limitée (p. ex. un seul cœur de CPU peut être utilisé) et les autres versions comportent des frais de licence élevés pour un usage commercial. Les bases de données immobilières en Californie peuvent être très volumineuses et les jointures très chronophages ; un ordinateur de 8 cœurs ou plus doté de 128G de RAM peut donc être utile pour les grandes jointures. Les bases de données SQL sont également importantes lorsqu'il faut retrouver tous les biens situés dans une zone délimitée de manière arbitraire. Ainsi, si l'on a par exemple utilisé la classification automatique (« Cluster Analysis ») pour définir des quartiers, on sera enclin à utiliser une base de données SQL. Toutefois, les différentes bases de données possèdent des fonctions intégrées différentes pour le traitement de l'information géographique et ne sont donc pas compatibles entre elles. Pour la présente série d'articles sur la RCA, PostgreSQL sera supposé pour les tâches SQL complexes et SQLite pour les tâches simples.

### Hypothèse 3 : « R/earth » est utilisé pour le traitement MARS.

Jusqu'en 2017, le logiciel MARS de Salford Systems constituait une solution abordable à \$270 par an. Cependant, après l'acquisition de Salford Systems par Minitab, MARS a été intégré à leur suite SPM au prix de \$16,000 par an — un tarif prohibitif pour la plupart des petites entreprises. Heureusement, le package « earth » du langage de programmation R (MILBORROW 2024a), maintenu par Stephen Milborrow, a évolué au point d'égaliser, voire de dépasser, les capacités du MARS de SPM de Minitab. Le package earth offre des avantages supplémentaires : il est disponible gratuitement, s'intègre parfaitement à l'écosystème statistique complet de R et peut être utilisé depuis Python au moyen de Pandas, offrant une excellente flexibilité pour différents flux de travail. Les discussions relatives à MARS supposent l'utilisation spécifique de R/earth.

### Hypothèse 4 : Notation

Les données brutes issues de sources externes telles que les fichiers CSV et les feuilles de calcul Excel sont représentées à l'aide de caractères latins standard. Cela inclut les données de biens du Multiple Listing Service (MLS), les paramètres de configuration de MARS (Multivariate Adaptive Regression Splines) utilisant le package R/earth, ainsi que les données propres au projet. Cette notation s'étend à leurs représentations correspondantes sous forme de data frames en R ou en Python.

Lorsqu'il est question des *caractéristiques des biens* et de leurs divers types et sous-ensembles, la notation passe aux lettres grecques pour une précision et une clarté analytique accrues. Si cette notation mathématique peut paraître plus rigoureuse, elle sert un objectif crucial : distinguer les caractéristiques abstraites des biens de leurs représentations sous forme de données brutes. Par convention, les caractères grecs désignent systématiquement les caractéristiques des biens, tandis que les caractères latins indiquent les structures de données de type feuille de calcul. Par exemple, le symbole  $M$  représente la matrice de données MLS, organisée avec les biens en lignes et leurs caractéristiques correspondantes en colonnes.

### Hypothèse 5 : Langages de programmation

Dans ce domaine, R et Python avec Pandas sont les langages dominants, C ou C++ étant utilisés lorsque la vitesse est importante. R/earth est lui-même écrit en R et en C.

### Hypothèse 6 : Indexation

L'indexation peut varier selon les langages de programmation et les mathématiques, ce qui peut prêter à confusion :

- Les mathématiques et R commencent l'indexation à 1.
- C, C#, C++ et Python commencent à 0.

Dans le présent article, par souci de cohérence et de simplicité, toute indexation commence à 0 (comme en Python ou en C). Voici comment cela fonctionne :

- Pour les listes, matrices ou data frames :
  - L'indice 0 est réservé au « bien à évaluer » (le bien principal étudié).
  - Exemple : dans un tableau dont les lignes sont des biens, la ligne 0 est le bien à évaluer et la colonne 0 contient les identifiants uniques de tous les biens.
  - Conseil pratique : disposer les biens en plaçant le bien à évaluer en premier (ligne 0), puis lister les autres de la date de vente la plus récente à la plus ancienne, en les numérotant séquentiellement (0, 1, 2, etc.).
- Lorsqu'il n'y a pas de bien à évaluer :
  - Si une structure de données ne contient pas de biens (p. ex. seulement des variables ou des fonctions), l'indice 0 est omis et l'indexation commence à 1.
  - Note particulière pour les utilisateurs de R : R utilisant une indexation à base 1, lors du codage en R des mathématiques du présent article, il convient d'ajouter 1 à tous les indices pour passer de la base 0 à la base 1.

## 5 Workflow RCA

Comprendre le workflow RCA est essentiel pour une analyse efficace. Le processus suit une séquence structurée de tâches analytiques et de points de décision

### 5.1 Types d'analyse MARS

Le workflow d'une analyse MARS dépend du point de savoir si seule une analyse de marché est requise, ou si la valorisation ou l'évaluation immobilière d'un bien donné est demandée. On notera qu'un rapport d'évaluation portant sur un bien particulier exige en général une analyse de sa zone de marché.

- Analyse de marché : elle porte sur des zones étendues telles que des comtés, des villes ou des régions définies. Cette approche ne fait intervenir aucun bien à évaluer particulier et n'exige donc aucune conclusion de valeur. Le workflow est centré sur le développement du modèle, sa validation et la production d'analyses à l'appui. Le processus s'achève une fois le modèle optimal identifié et documenté.
- L'analyse d'un bien individuel englobe tous les éléments de l'analyse de marché, auxquels s'ajoute la détermination de la valeur d'un bien particulier. Elle diffère fondamentalement des méthodes d'évaluation traditionnelles. Dans les approches traditionnelles, la valeur est habituellement obtenue en ajustant 3-9 ventes comparables et en calculant une moyenne pondérée. En revanche, la RCA détermine la valeur principalement à partir de deux composantes :
  1. La valeur prédite par le modèle MARS
  2. Le résidu propre au bien estimé (avec le score CQA associé)

Dans la RCA, le Sales Grid sert avant tout à expliquer les différences de valeur entre le bien à évaluer et les comparables en illustrant les contributions des variables. Un avantage clé de la RCA est sa stabilité : si les clients peuvent débattre de la sélection des comparables, la conclusion de valeur demeure constante quels que soient les comparables retenus pour la présentation, puisque toutes les ventes (qu'il y en ait 15 ou 600) sont ajustées au moyen du même modèle. Bien que les ventes plus anciennes soient automatiquement ajustées aux conditions actuelles du marché, les ventes récentes sont généralement préférables. L'efficacité du modèle à capturer les facteurs de prix diminue avec l'éloignement temporel, les caractéristiques non mesurées du marché prenant davantage d'importance sur de longues périodes.

### 5.2 Workflow de traitement

Le workflow de traitement global peut être décrit par les étapes suivantes :

1. Périmètre de mission.
2. Mise en place des dossiers et fichiers initiaux du projet.
3. Téléchargement des données MLS. Se reporter aux Protocoles 1.1–1.3 concernant le nombre d'enregistrements de biens nécessaires et la fixation de limites sur les plages de dates.
4. Renseignement des feuilles de configuration : Regression Variables, Interactions, Aggregations, Calculations.
5. Exécution de la régression MARS (R/earth) pour créer le modèle de prix.
6. Génération des contributions à la valeur des biens & des résidus (résidu pur et résidu/SF) à partir du modèle MARS.
7. Classement des biens selon le résidu/SF.
8. Création des scores CQA pour chaque vente comparable.

9. En supposant qu'un classement des biens par qualité a été produit et qu'il représente fidèlement l'attrait global des biens, trouver la meilleure position du bien à évaluer dans ce classement et lui attribuer un score CQA, en étudiant les tendances d'ensemble du classement.
10. Attribuer au bien à évaluer le résidu/SF correspondant, soit  $\text{résidu/SF} \cdot \text{GLA}$  (surface habitable brute) du bien à évaluer.
11. **Additionner l'estimation de prix MARS et le résidu attribué pour aboutir à la conclusion de valeur du bien à évaluer.**
12. Décomposer le résidu du bien à évaluer en composantes descriptives et en contributions à la valeur.
13. Calculer les contributions à la valeur produites par le modèle pour les comparables & les ajustements, ainsi que tous les sous-totaux et les prix de vente ajustés.
14. Les prix de vente ajustés seront égaux à la conclusion de valeur finale du bien à évaluer.
15. Sélectionner les 6-12 biens les plus similaires pour le Sales Grid.
16. Décomposer les montants de résidu total des comparables en contributions à la valeur des composantes résiduelles, en fixant la valeur totale du résidu à 0 une fois la décomposition achevée.
17. Recalculer tous les sous-totaux et les prix de vente ajustés des comparables du Sales Grid. Vérifier les valeurs.
18. Charger le Sales Grid dans le rapport.

On notera qu'avec la RCA, la conclusion de valeur finale ne découle PAS directement des valeurs des comparables figurant dans le Sales Grid. Avec la méthode RCA, le Sales Grid devient un outil servant uniquement à expliquer pourquoi des biens qui paraissent similaires au bien à évaluer se vendent plus ou moins cher que la conclusion de valeur finale ou que la valeur d'expertise du bien à évaluer. Il faut envisager les choses autrement : la conclusion de valeur du bien à évaluer dépend de l'estimation de prix du modèle MARS et du résidu estimé par le Valuation Engineer. Mais ces deux quantités dépendent très fortement de l'ensemble des biens entrés dans la régression  $R/\text{earth}$ , laquelle a produit le classement des résidus des biens ayant déterminé le score CQA du bien à évaluer et, partant, son résidu, par le placement dans le classement effectué par le Valuation Engineer. Les ajustements du Sales Grid sont en réalité déterminés par les contraintes mathématiques imposées par la RCA ; bien que la décomposition du résidu dépende dans une certaine mesure du jugement du VE, les totaux des contributions à la valeur issues de la décomposition doivent toujours correspondre au résidu associé. Nous savons par la Démonstration 2 que, si les valeurs effectives des composantes résiduelles peuvent varier selon le jugement du VE, leurs totaux pour un bien donné sont entièrement contraints et ne modifieront donc pas le prix de vente ajusté du bien associé, quelle que soit la manière dont le VE modifie la décomposition du résidu.

### 5.3 Workflow des données

On notera que le présent article ne présente qu'un workflow de haut niveau pour la méthode RCA. D'autres articles approfondiront les différentes étapes et sous-étapes de l'analyse RCA.

#### 5.3.1 Exigences de données pour MARS

Il conviendra de spécifier une date de début pour les transactions de vente et une date de fin fondée sur la date d'effet du rapport d'évaluation ou de l'évaluation. Cela prend habituellement en considération le nombre estimé de variables de régression et d'interactions qui seront nécessaires pour chaque degré d'interaction supérieur à 1.

### 5.3.2 Protocoles : listings de ventes conclues

#### Protocole 1.1 : Détermination du nombre requis de listings MLS de ventes conclues

Ce protocole décrit comment calculer le nombre minimal recommandé de listings de ventes conclues à fournir en entrée de la régression MARS. Il faudra commencer par une estimation approximative, exécuter MARS en passes de premier, de deuxième et éventuellement de troisième degré afin de générer le modèle initial, puis examiner le nombre de lignes selon les règles suivantes pour déterminer si des lignes supplémentaires sont nécessaires. Ce protocole suppose que les variables indépendantes sont indépendantes à 100%, ce qui est rarement le cas en immobilier ; par exemple, la GLA et les nombres de pièces sont généralement fortement corrélés, et la GLA est également susceptible de présenter une corrélation avec la superficie du terrain. On touche ici au calcul des « degrés de liberté (DOF) ». Chaque variable véritablement indépendante reçoit un degré de liberté. Mais lorsqu'il existe des dépendances entre les variables dites indépendantes, un calcul exact des DOF exige une analyse qui dépasse le cadre du présent article. Il faut simplement garder à l'esprit qu'en raison d'une colinéarité probable entre les variables, les estimations ci-dessous sont sans doute plutôt élevées.

1. Pour les termes du premier ordre (sans interactions), la règle des  $\sim 15$  lignes par variable donne de bons résultats
2. Pour chaque interaction à deux variables autorisée, il convient d'ajouter environ 15-30 biens
3. Pour les interactions d'ordre supérieur (3 variables ou plus), chacune devrait disposer de 30 biens ou plus.

Il ne s'agit toutefois pas d'une règle absolue. Le nombre réellement nécessaire peut dépendre :

- Du rapport signal/bruit des données
- De la complexité des relations que l'on cherche à modéliser
- De la variabilité de la variable réponse
- De la qualité des données

Exemple :

Ainsi, si l'on dispose de :

- 10 variables principales
- 5 interactions à deux variables autorisées
- 2 interactions à trois variables autorisées

Le nombre minimal de ventes peut alors se calculer ainsi :

$$\begin{aligned} N &= (10 \text{ variables} * 15 \text{ lignes/variable}) + \\ &\quad (5 \text{ interactions-à-deux-variables} * 22 \text{ lignes/interaction-à-deux-variables}) + \\ &\quad (2 \text{ interactions-à-trois-variables} * 30 \text{ lignes/interaction-à-trois-variables}) \\ &= 320 \text{ lignes} \end{aligned}$$

**On peut également procéder comme suit :** après avoir exécuté R/earth sur un jeu de données, on peut utiliser le nombre de fonctions de base du modèle, plus un, comme estimation des degrés de liberté, puis multiplier ce nombre par 20-30 pour obtenir le nombre d'enregistrements d'entrée nécessaires.

**Protocole 1.2 : Les périodes de vente doivent couvrir une évolution minimale des goûts des acheteurs**

Lors de la collecte de données historiques sur les biens, il n'existe pas de limite temporelle stricte dès lors que les évolutions des préférences du marché peuvent être capturées par des variables appropriées. Les points clés suivants sont à considérer pour traiter l'évolution architecturale du marché étudié : les changements de styles architecturaux dominants (comme le passage du style California Ranch au style méditerranéen) peuvent être modélisés à l'aide :

- D'identifiants de lotissement
- De codes de zone MLS
- De plages de dates de construction

Toutefois, si ces évolutions de style ne peuvent être capturées de manière adéquate par des variables spécifiques, on se heurte à une contrainte importante : la date de vente étant habituellement utilisée pour l'ajustement aux conditions de marché, l'utiliser simultanément pour contrôler les préférences architecturales confondrait ces deux effets distincts. Dans de tels cas, l'approche à privilégier consiste :

1. Soit à limiter la collecte de données à la période postérieure à l'introduction du style architectural le plus récent
2. Soit à restreindre la sélection des comparables aux lotissements partageant le même style architectural

On garantit ainsi la cohérence des caractéristiques des biens dans l'analyse, tout en permettant un ajustement correct aux conditions de marché au fil du temps. Alternativement, du moins dans ce cas, à supposer que le style architectural puisse être extrait des données, on pourrait considérer l'interaction entre ces deux variables. Une telle extraction n'est cependant pas toujours possible.

**Protocole 1.3 : En zone de ventes rares, faire au mieux avec les données disponibles**

Lorsque l'on travaille avec la modélisation MARS sur des données de ventes, environ 15 transactions sont nécessaires pour une analyse significative. Si l'on dispose de moins de ventes (par exemple 5 biens), on peut dupliquer chaque enregistrement deux fois afin d'atteindre le seuil de 15 biens. Considérations importantes pour cette approche :

1. N'utiliser qu'une régression au premier degré afin de réduire le surajustement
2. Ignorer ou omettre la validation croisée, le jeu de données étant trop limité
3. Bien que MARS surpasse vraisemblablement les méthodes de paires appariées ou de régression linéaire standard dans ce scénario, il faut être conscient que la duplication des données affecte la validité des mesures d'incertitude et de la significativité statistique

La meilleure solution à long terme consiste à étendre le jeu de données avec des ventes réelles supplémentaires. Toutefois, s'il faut procéder avec des données limitées, la duplication des données constitue une solution temporaire praticable, dont il convient de reconnaître les limites statistiques.

## 5.4 Structure des dossiers de projet

Il n'est pas inutile d'aborder la structure des dossiers de projet. Il est judicieux de disposer d'un dossier distinct pour les analyses de marché et pour les rapports d'évaluation de biens individuels. Ces deux dossiers présentent une structure de sous-dossiers similaire :

```

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
│   │       │           └── ...

```

Le traitement de l'étape 1 crée le dossier de version avec ses sous-dossiers. Il copie le fichier `MLS_Original.xlsx`, bien que ce classeur ne *devrait* jamais changer, mais refléter exactement ce qui a été téléchargé depuis le MLS à une date et une heure données ; il reste toutefois possible, sur un projet de plus longue durée, qu'un nouveau téléchargement MLS soit nécessaire. Ensuite, cette étape copie dans le dossier de version le classeur `MLS.xlsx` étendu, avec ses paramètres de configuration. Il peut exister de nombreux dossiers de version, chacun créé pour une exécution complète de MARS.

Puis, les étapes suivantes génèrent des ajouts de colonnes, des tris et d'autres modifications

du classeur, et écrivent une version mise à jour sous le nom `MLS_Stage(N)` dans le dossier de données du projet lorsque le rapport d'évaluation est achevé. Répétons-le : le classeur racine `MLS.xlsx` subit des changements constants, de sorte que la copie de version est une mutation continue du classeur `MLS` d'origine, avec des suppressions et des ajouts de variables, ainsi que d'autres modifications. Le code de traitement R ou Python s'exécutera très probablement avec, en entrée, le classeur `MLS.xlsx` du dossier `Data`. Nous disposons ainsi toujours d'une copie de toutes les données sous-jacentes à chaque version de l'exécution RCA.

Autant de versions sont créées que nécessaire, jusqu'à ce qu'aucune amélioration ne soit plus possible ; la meilleure version est alors retenue comme version finale. On peut modifier le nom du dossier en lui ajoutant le suffixe « `_Final` », afin que cela apparaisse clairement lors du parcours du répertoire. Certaines versions plus anciennes, présentant des modèles intéressants ou concurrents, devraient généralement être conservées. Chaque dossier de version contient un certain nombre de fichiers de texte, de diagrammes et de graphiques pouvant être intégrés dans un rapport. Le classeur final `MLS_Stage5.xlsx` contient autant de feuilles que nécessaire pour le rapport. Il peut notamment comporter une feuille de calcul destinée au transfert vers un Sales Grid Fannie Mae. Il contiendra aussi nécessairement les calculs ou la décomposition sous-jacents aux agrégations. C'est potentiellement la partie la plus chronophage de l'approche RCA - la quête du modèle parfait dans une zone de marché difficile.

Tout ce qui est nécessaire pour recréer les Sales Grids doit être copié dans le dossier de version, y compris le code R ou Python utilisé - si ce code fait l'objet de modifications fréquentes.

#### 5.4.1 Protocoles : exécutions MARS (earth) multiples

##### Protocole 2.1 : Exécutions en surajustement pour étudier finement les facteurs de prix d'une zone de marché

Lors des premières exécutions de earth, le surajustement doit être employé de manière stratégique pour découvrir les facteurs de prix sous-jacents. Par exemple, une zone de marché représentée par 800 ventes peut ne contenir que 3 biens dotés d'un atelier, et dans ces cas isolés l'atelier peut avoir un impact majeur sur le prix de vente final. Une exécution limitée à la passe avant produira un modèle surajusté qui montre la contribution à la valeur approximative de ces ateliers. Cependant, comme il n'y a que 3 ventes de ce type sur 800, les termes relatifs à l'atelier seront éliminés lors de la passe arrière, et à plus forte raison en validation croisée. La valeur que la passe avant seule attribue aux ateliers peut être consignée et reportée dans l'analyse des variables latentes, pour être utilisée pour tout bien doté d'un atelier figurant dans le Sales Grid. De tels modèles surajustés fournissent également des estimations détaillées pour de nombreuses autres caractéristiques à données rares qui seraient éliminées par l'élagage de la passe arrière ou par les exécutions en validation croisée.

Chacune de ces exécutions exploratoires doit être sauvegardée dans son propre dossier de version : les contributions à la valeur consignées constituent des éléments que le VE voudra citer plus tard lors de la décomposition des résidus. Il faut bien entendu comprendre que le VE doit en dernier ressort s'appuyer sur l'élagage et la validation croisée pour rendre le modèle final robuste face à des biens que la régression n'a jamais vus — tel un bien à évaluer qui apparaît ultérieurement. Les exécutions surajustées sont des instruments d'étude ; elles ne constituent jamais le modèle final.

### Protocole 2.2 : Exécutions multiples en validation croisée sous différents paramètres de earth()

Les zones de marché diffèrent par leur structure et leur taille, et le VE doit généralement expérimenter une série d'exécutions : pour comprendre quelles variables influent le plus sur le prix ; pour juger si l'ensemble existant de variables mesurées estime la valeur de manière adéquate, ou si de nouvelles données doivent être ajoutées (distance à l'eau, altitude, distance aux centres commerciaux, etc.) ; et pour découvrir de nouveaux types d'interactions tout en écartant d'autres comme inefficaces ou trompeuses. Au final, le VE recherche un bon  $R^2$  sur les données d'entraînement et le meilleur CVR2 possible sur les données de test. Y parvenir peut exiger de collecter de nouvelles données, d'améliorer l'exactitude des données existantes, ou d'ajuster les paramètres de earth() à la taille et au caractère de la zone de marché (p. ex. au regard des degrés de liberté disponibles — voir le Protocole 1.1).

C'est précisément pourquoi la structure de projet présentée ci-dessus prend en charge de nombreux dossiers de versions d'exécution : chaque modèle candidat est conservé avec ses paramètres et son  $R^2$ /CVR2, de sorte que la trajectoire d'une exécution à l'autre peut être comparée. Une version dont le  $R^2$  grimpe tandis que son CVR2 stagne ou chute ajuste du bruit plutôt que le signal du marché, et c'est l'historique des versions qui rend ce phénomène visible.

### Protocole 2.3 : Expansion Prolog des remarques et des colonnes descriptives de caractéristiques

Bien que cette fonctionnalité soit encore dans une phase quelque peu expérimentale, earthUI est désormais livré avec un moteur Prolog intégré (vProlog) et utilise des grammaires à clauses définies (Definite Clause Grammars, DCG) pour analyser les colonnes de texte libre du MLS — Public Remarks avant tout, mais aussi les remarques d'agents et les listes de caractéristiques délimitées par des virgules, telles que les caractéristiques des pièces ou les structures existantes. Cette expansion transforme des expressions comme « owned solar », « 3 car garage » ou « 40x50 workshop » en colonnes de caractéristiques ordinaires pouvant être proposées à la régression, et les caractéristiques à données rares ainsi récupérées alimentent directement l'analyse des variables latentes (voir la sous-section ultérieure sur les notations d'état et de qualité comme régresseurs approchés ; la fonctionnalité elle-même est documentée dans l'article compagnon consacré à earthUI dans ce numéro). L'expansion textuelle multiplie l'ensemble des prédicteurs candidats, ce qui impose de multiples exécutions earth/RCA à des fins de test : les colonnes issues de l'expansion doivent être essayées dans le modèle et hors du modèle, consolidées en comptages ou en booléens dérivés lorsqu'elles sont colinéaires, et vérifiées au regard du CVR2 afin de confirmer qu'elles portent un signal de marché plutôt que du bruit. Chaque expérience de ce type doit être traitée comme son propre dossier de version d'exécution, exactement comme dans les Protocoles 2.1 et 2.2. Et puisque la grammaire d'analyse est elle-même en cours de développement, le VE devrait vérifier les colonnes extraites au regard du texte source avant de s'y fier dans un modèle.

## 5.5 Le classeur Excel central

Puisque la plupart des services MLS téléchargent les données des biens vers des feuilles de calcul Excel CVS, Excel est probablement le moyen le plus commode de stocker les données et les divers paramètres de configuration. Il existe un classeur Excel central pour chaque projet, nommé `MLS.xlsx`, qui comporte une feuille appelée « `MLSData` » pour stocker les listings de ventes conclues téléchargés depuis le MLS, ainsi que plusieurs feuilles de configuration et de paramètres :

```
Classeur : MLS.xlsx/  
├─ Feuille : Project Data  
├─ Feuille : MLSData  
├─ Feuille : Regression Variables  
├─ Feuille : Calculations  
├─ Feuille : Allowed Interactions  
├─ Feuille : Aggregations  
└─ Feuille : (paramètres d'exécution MARS)
```

Les données d'origine sont stockées dans un classeur nommé par exemple `MLS_Original.xlsx`. Une copie doit en être faite sous un autre nom, disons « `MLS.xlsx` », qui sera chargée dans le programme RCA en R ou en Python. Ces deux classeurs sont stockés dans le dossier `Data`. Or, le classeur `Data/MLS.xlsx` passera vraisemblablement par de nombreuses itérations d'exécutions, au fil d'essais de différentes combinaisons de variables, de paramètres MARS et de modifications d'autres réglages. Chaque itération crée un nouveau dossier d'entrée/sortie nommé avec la date et l'heure, au format `VER_YYYYMMDDHHMMSS`. Chaque dossier contiendra une copie du classeur `MLS.xlsx` d'entrée ainsi que les classeurs Excel générés pour chacune des quatre étapes de traitement (étapes 2–5), de même que les documents texte, diagrammes et graphiques produits en sortie. On notera que l'étape 5, l'étape finale, peut contenir un certain nombre d'autres feuilles pour le rapport, selon le type de client et le SOW (périmètre de mission).

## 5.6 Sauvegardes

Lors de l'écriture de feuilles de calcul Excel nouvelles ou modifiées dans les dossiers de version du projet, des sauvegardes peuvent également être effectuées vers des bases de données SQLite ou PostgreSQL dans le dossier du projet. Les fichiers de base de données risquent moins d'être corrompus involontairement au cours des périodes ultérieures, lors de consultations ou d'utilisations de référence.

## 6 Notation du classeur

### 6.1 Dimensions descriptives

Un classeur « M » comporte environ six dimensions descriptives :

1. Rapports d'évaluation « a »
2. Versions d'exécution « r »
3. Étapes du flux de travail « w »
4. Feuilles « s »
5. Listing de bien « p »
6. Ensemble de variables « v »

Ainsi, une cellule individuelle d'un classeur M donné peut être décrite comme suit :

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

(Remarque : « \* » désigne un indice valide quelconque)

ou :

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

où :

- $n_p$  = # de lignes (selon la feuille)
- $n_v$  = # de colonnes (selon la feuille)
- $n_r$  = # de versions d'exécution
- $n_m$  = # d'étapes du flux de travail
- $n_g$  = # de projets d'analyse
- $n_h$  = # de feuilles du classeur

Il convient de comprendre que « W » désigne l'intégralité du classeur Excel, composé de six feuilles de données principales possédant chacune leur propre table, lesquelles, lors de l'exécution du programme R ou Python associé, sont chargées dans des data frames internes distincts. R comme Python disposent de fonctions internes capables, à partir d'un nom, de déterminer à quelle ligne ou colonne du data frame celui-ci appartient. Par exemple, si dans l'indice  $v_j$ ,  $v =$  (« GLA », « LotSize », « BathRms », « BedRms », ...) et  $j=0$ , alors  $v_j =$  «GLA». R comme Python sauront résoudre en interne « GLA » vers la colonne dans laquelle cette variable est stockée, qui pourrait être, par exemple, la colonne 15 du data frame associé. Ainsi, dans le présent article, nous indexons nos tableaux de noms, lesquels transmettent un nom au programme en cours d'exécution, qui ré-indexe ce nom en l'indice réel de ligne ou de colonne du data frame interne. C'est un peu compliqué. Mais, pour être précis, cela est nécessaire et des plus commodes. Il s'agit d'une « séparation des responsabilités » entre le rôle d'analyste et le rôle de développeur, bien qu'en réalité le Valuation Engineer (VE) puisse assumer les deux rôles. Notons qu'un « data frame » de R s'appelle « dataframe » en Python.

Notons également que la feuille des paramètres d'exécution de MARS n'est pas une dimension de W, car W se rapporte à la version des données fournies en entrée à MARS, indépendamment des paramètres qui déterminent la manière dont MARS/earth traite ces données.

Par souci de clarté, une notation multidimensionnelle complexe du classeur est présentée, qui correspond plus directement aux data frames internes de R ou Python dans lesquels les feuilles de données du classeur sont chargées. Il s'agit d'une notation « lourde ». Mais lorsque nous aborderons les démonstrations, etc., nous nous délesterons d'une grande partie de ce bagage en formulant des hypothèses sur les objets manipulés. Néanmoins, cette notation aide à obtenir d'emblée une meilleure vue d'ensemble du flux de travail et des données.

Commençons par la notation de l'ensemble des classeurs que nous allons traiter.

### 6.1.1 La dimension projet (« a »)

Une fois qu'un Scope of Work (SOW, « périmètre de mission ») a été approuvé, un dossier de projet doit être créé pour stocker les données d'entrée et de sortie, y compris les paramètres de configuration, les graphiques de sortie, les diagrammes, les feuilles de calcul et les extraits de rapport. Le VE devrait identifier tous les projets par un simple ID séquentiel, tel qu'un entier, de sorte que la disparition éventuelle d'un rapport soit manifeste. Si des projets sont annulés, il devrait exister un moyen cohérent d'indiquer l'annulation du projet au sein du dossier de projet concerné, mais celui-ci ne devrait jamais être supprimé. Nous supposons que les ID de projet commencent à 0 et sont incrémentés de 1, bien que d'autres possibilités existent. Nous supposons en outre que tous les ID de projet valides sont stockés dans un ensemble ordonné « a » créé avec le programme de traitement.

Les données du projet d'évaluation figurent dans la feuille de calcul projet du classeur, mais pourraient également être stockées dans une base de données SQL centrale.

Voici les champs de projet suggérés :

1. ID local : un identifiant entier court généré séquentiellement pour l'entreprise. Il constituerait une excellente clé primaire pour le stockage - et pourrait être généré automatiquement par la base de données à chaque nouveau projet.
2. GUID global : si vous le souhaitez, vous pouvez créer un GUID (Globally Universal ID) à l'aide de

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

3. Localisation du bien (analyse de marché) ou adresse
4. Date d'effet de l'évaluation
5. Date de la commande
6. Date de l'inspection
7. Date du rapport original
8. Date du premier rapport mis à jour
9. Date du deuxième rapport mis à jour
10. Date du troisième rapport mis à jour
11. Nom du client
12. Adresse du client
13. Nom du prêteur
14. Adresse du prêteur
15. Nom du propriétaire
16. Adresse du propriétaire
17. ....

Exemple de liste d'ID de projet :

a =        (« 1 »,  
              « 2 »,  
              « ... »,  
              « 141 »,  
              « 142 »)

### 6.1.2 La dimension des versions d'exécution (« r »)

En deuxième lieu dans le flux de travail, nous exécutons la régression MARS sur les données d'entrée. Cela impliquera vraisemblablement de plusieurs à de nombreuses exécutions, à mesure que nous modifions divers paramètres pour améliorer le modèle. Chaque exécution créera un nouveau dossier de version pour la sortie, dont le nom repose sur la date et l'heure d'exécution. Par exemple,

r = ( « ver20241210\_083124 »,  
      « ver20241209\_151130 »,  
      « ver20241208\_160944 »)

latest\_version :  $r_1$  = "ver20241210\_083124"

Les versions prometteuses, dotées de modèles et de sorties compétitifs, devraient être conservées jusqu'à ce qu'une version finale soit arrêtée, voire plus longtemps si elles présentent une certaine utilité. Vous pouvez modifier le dossier de version pour indiquer qu'il s'agit de la version finale, en plaçant « Final » à la fin du nom.

### 6.1.3 Étapes du flux de travail (« w »)

Le flux de travail RCA est habituellement séparé en cinq étapes distinctes, énumérées ci-dessous. Chaque étape doit s'achever avec succès avant le démarrage de la suivante. Le VE devrait pouvoir arrêter le traitement à la fin d'une étape afin d'examiner les résultats avant de poursuivre. Cela est nécessaire car les zones de marché sont relativement complexes et souvent tout à fait singulières. Des problèmes inhabituels exigeant une nouvelle approche peuvent survenir. Le VE doit pouvoir intervenir dans le code R ou Python pour effectuer les modifications nécessaires, voire essayer de nouvelles approches. Et, dans de tels cas, il peut vouloir expérimenter et essayer différentes approches nécessitant plusieurs exécutions de la même étape sur laquelle il travaille. Cela étant, le VE devrait aussi être libre d'exécuter toutes les étapes en une seule commande, - sous certaines réserves (p. ex. l'agrégation du 3e degré doit déjà avoir été déterminée lors d'une exécution antérieure de l'étape 2). Cela signifie qu'à l'achèvement de chaque étape, celle-ci devrait stocker toutes ses données internes dans des fichiers ou des bases de données, de sorte que l'étape suivante puisse démarrer comme si l'étape précédente était encore en mémoire, en rechargeant en mémoire les données de l'étape précédente.

Voici les étapes :

w = ( « stage1 », -Configuration  
      « stage2 », -Traitement MARS  
      « stage3 », -Sales Grid finale  
      « stage4 », -Extraits de rapport  
      « stage5 », -Revue/Nettoyage

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

#### 6.1.4 Feuilles du classeur Excel (« s »)

Dans le classeur Excel RCA Core, nous disposons des feuilles suivantes, chacune étant chargée dans un data frame interne :

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

La raison pour laquelle les agrégations du 2e degré sont placées sur une feuille séparée est qu'elles se définissent le plus aisément au moyen d'une table symétrique présentant toutes les paires d'interactions possibles, permettant de spécifier, pour chaque paire, une variable d'agrégation à laquelle les contributions à la valeur et les ajustements associés seront agrégés (additionnés). Cela n'était pas possible pour les agrégations du 3e degré, car il faudrait une table symétrique tridimensionnelle, ce qui est impossible dans Excel. En outre, une table de 20x20x20 représenterait 8,000 cellules à gérer. Il est plus simple de lister les noms des éventuels termes du 3e degré dans la table des agrégations du 1er degré, par exemple :

| Variable            | Agréger vers |
|---------------------|--------------|
| 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 biens (« p ») - lignes de MLSDData

Les listings de biens forment les lignes de la feuille MLSDData. Ils n'apparaissent pas dans les autres feuilles avant l'étape 4, où les extraits finaux de rapport sont créés avec la sélection finale de biens comparables. Voici un exemple de liste d'ID de biens :

```
p = (« MLS823211 », « MLS823456 », « MLS823542 », « MLS823721 », « ... »)
```

### 6.1.6 Variables MLS (« v ») colonnes de MLSData

Nous reviendrons plus loin sur le traitement des variables. Qu'il suffise d'indiquer à ce stade que la feuille MLSData comporte typiquement 30-40 variables de types divers, mais que seules 10-20 d'entre elles seront généralement retenues comme variables indépendantes pour la régression MARS. Ces variables apparaissent également dans d'autres feuilles, généralement toutefois sous la forme d'une liste de toutes les variables susceptibles d'être soumises à MARS. Voici un exemple de liste de variables :

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

### 6.1.7 Exemple final de notation

Rassemblons la notation ci-dessus dans un exemple final. À l'aide des listes précédentes, posons

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

Notons que nous utilisons ici une indexation à partir de 0, de sorte que le premier élément d'une liste est supposé avoir l'indice 0. Python utilise également l'indexation à partir de 0, tandis que R utilise l'indexation à partir de 1.

Ainsi, ce qui précède signifie que nous parlons de la valeur de la cellule correspondant à :

- La variable  $v_0$ , soit la colonne « GLA »
- Pour le bien  $p_2$ , soit l'ID « MLS823542 »
- De la feuille  $s_0$ , soit « MLSData »
- Pour l'étape de flux de travail  $w_1$ , soit « stage2 »
- De la version  $r_1$ , soit  
« ver20241209\_151130 »
- Du projet  $a_{141}$ , soit « 142 »

La notation ci-dessus peut alors également s'écrire :

$$\begin{matrix} \text{«ver20241209_151130»} \\ \text{«stage2»} \end{matrix} \begin{matrix} \text{«142»} \\ W \\ \text{«MLSData»} \end{matrix} \begin{matrix} \text{«MLS823542»} \\ \text{«GLA»} \end{matrix} \quad (4)$$

### 6.1.8 Justification

« Pourquoi ne pas simplement utiliser des index ou des numéros dans la notation ? » Notre notation doit satisfaire deux exigences :

1. Les sous-ensembles : l'analyste sélectionne fréquemment des sous-ensembles d'une liste, par exemple des sous-ensembles de variables. Ainsi, la feuille MLSData peut contenir 35-40 variables, dont seulement 10-20 sont utilisées pour une exécution donnée. D'une exécution à l'autre, le VE change fréquemment les variables soumises à MARS/earth.
2. L'itération, p. ex.  $i = 1, 2, \dots, n_1$   
Nos algorithmes, protocoles et démonstrations doivent généralement pouvoir itérer sur tous les éléments d'une liste ordonnée, du premier au dernier – ou en extraire un élément particulier.

Il nous faut donc une lettre représentant le sous-ensemble ou la liste d'éléments, laquelle doit être munie d'un indice supplémentaire permettant d'itérer sur les éléments de la liste afin d'obtenir les chaînes de caractères servant à indexer les lignes ou les colonnes du data frame.

### 6.1.9 Matrices ?

Et qu'en est-il des matrices de l'algèbre linéaire ? Les matrices ne fonctionnent qu'avec des nombres, entiers ou réels. Les matrices en tant que structure de données sont prises en charge tant en R qu'en Python. Toutefois, nombre des variables que nous manipulons dans les données MLS sont des « chaînes de caractères » (strings), c'est-à-dire des données de type caractère. Même lorsque certaines variables ne contiennent que des nombres, il s'agit souvent non pas de nombres réels, mais de simples noms. Par exemple, les codes de zone MLS d'une ville peuvent être (« 660 », « 661 », « 662 », « 663 », ..., « 672 »). La régression earth permet de spécifier des variables comme variables « factor », auquel cas les nombres sont traités simplement comme des noms, sans aucun ordre. Dans ce cas, le modèle produit fournira simplement une contribution à la valeur pour chaque nombre, plutôt qu'une fonction définie sur la séquence. D'autres exemples incluent « C1 », ..., « C6 » pour l'état et « Q1 », ..., « Q6 » pour la qualité.

### 6.1.10 Première simplification

La notation ci-dessus pour le classeur Excel central du projet est utile pour relier les éléments entre eux, mais trop lourde à manipuler lorsqu'il s'agit d'expliquer des algorithmes ou des démonstrations mathématiques. Lorsque nous effectuons le travail de détail, nous pouvons supposer que nous nous trouvons dans le projet, la version, l'étape et la feuille considérés. Nous indiquons simplement l'objet sur lequel nous travaillons, s'il n'est pas déjà implicite. Tout ce dont nous avons réellement besoin dans la notation ci-dessus, une fois précisé l'objet du travail, est :

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

où :  $n_p$  = # biens  
 $n_v$  = # variables d'entrée  
 $n_s$  = # feuilles d'entrée

### 6.1.11 Seconde simplification

Il est plus commode de renommer les feuilles selon les types particuliers de data frames qu'elles prennent en charge plutôt que d'utiliser des index. Pour chaque feuille, les lignes et les colonnes ont des fonctionnalités différentes. Ainsi, l'ensemble W de data frames est décomposé en data frames M, V, A, I et C comme suit :

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

où :  $n_p$  = # biens  
 $n_v$  = # variables d'entrée

**Note :** le bien à évaluer figure par défaut toujours en tête des listes de biens

Dans la feuille de calcul Excel « MLSData » et le data frame associé, la première ligne est toujours réservée au bien à évaluer. Elle n'est jamais triée avec les autres biens MLS situés en dessous. Elle n'est jamais introduite dans la régression MARS comme donnée indépendante ; elle y entre plutôt comme donnée cible, c'est-à-dire comme variable dépendante. Il en va de même pour les listes de biens. Dans toute liste de biens, le premier élément, d'index 0, est toujours supposé être l'identifiant du bien à évaluer, sauf indication contraire.

Même lorsque la tâche consiste à créer un modèle pour une zone de marché sans aucun bien à évaluer, la première ligne est néanmoins remplie avec des données de bien fictives. Cela rend beaucoup plus commode l'écriture de programmes, la spécification d'algorithmes ou la rédaction de démonstrations. Ainsi, si nous utilisons  $c$  pour représenter la liste des biens,  $c_0$  est le bien à évaluer et les  $c_i$ ,  $i = 1, \dots, n_p$ , sont les listings comparables qui seront fournis en entrée à la régression MARS/earth.

◇ **Spécifications des variables de régression**

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

où :  $n_v = \#$  variables disponibles  
 $n_s = \#$  de spécifications

On notera que  $n_s \leq n_v$ , car nous sélectionnons habituellement un sous-ensemble des variables disponibles comme entrée de la régression.

◇ **Calculs**

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

où :  $n_v = \#$  de variables  
 $n_c = \#$  de spécifications de calcul

◇ **Interactions**

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

où :  $n_f = \#$  variables origine (from)  
 $n_t = \#$  variables destination (to)

On notera que le data frame I est un data frame symétrique, dont les variables sont énumérées par ordre alphabétique pour les lignes « from » (origine) et les colonnes « to » (destination). Les cellules sont marquées d'un « 1 » pour indiquer qu'une interaction entre la paire de variables correspondante est autorisée. À l'inverse, un « 0 » indique qu'aucune interaction n'est permise. Pour une interaction du second degré, la situation est relativement simple, une seule interaction étant possible. Avec une régression du troisième degré, il existe trois interactions par paires entre les trois variables, chacune de ces paires devant être autorisée pour engendrer le terme (ou la variable) d'interaction à trois variables.

◇ **Agrégations**

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

où :  $n_v = \#$  de variables

## 7 La Sales Grid RCA

Il convient de garder à l'esprit que le protocole RCA exige un programme R ou Python capable d'accéder à R/earth et d'exécuter l'ensemble des étapes détaillées. Le traitement est intensif en données et le travail ne pourrait en aucun cas être effectué manuellement. En même temps, à supposer que le programme soit exempt d'erreurs, un logiciel est nécessaire pour garantir que le travail requis soit accompli avec précision et sans erreur humaine.

Il est également indispensable que les VE qui effectuent ce travail soient en mesure de modifier de tels programmes selon les besoins, en particulier lorsqu'ils sont confrontés à de nouvelles zones de marché imprévues, à de nouveaux quartiers, à de nouveaux biens ou à d'autres exigences ou problèmes inédits.

Au sein du traitement RCA, on distingue trois types de Sales Grids :

1. La RCA Processing Sales Grid, désignée dans le présent article par « MLSDData », qui est un très grand data frame accompagné d'une feuille de calcul associée, pouvant contenir de nombreux détails intermédiaires qui passent finalement par un processus d'agrégation afin d'aboutir à la Sales Grid RCA mentionnée ci-dessous. Un exemple serait trop volumineux pour être présenté ici. Selon toute vraisemblance, les Valuation Engineers adapteront cette feuille de calcul à leurs préférences. J'ai l'intention de consacrer ultérieurement un article distinct à l'élaboration d'une telle grille.
2. La RCA Final Sales Grid, utilisée dans les rapports non destinés aux GSE, qui ressemble à [Table 2](#). Elle imite, par exemple, le formulaire de Fannie Mae, suffisamment pour offrir au lecteur un certain repère de familiarité. On remarquera toutefois qu'elle comporte davantage de colonnes par bien comparable que la grille de Fannie Mae : chaque comparable dispose d'une colonne « Contribution à la valeur » aux côtés de ses colonnes « Mesure » et « Ajustement », et le bien à évaluer en possède également une — des colonnes qui font défaut au formulaire de Fannie Mae. Une forme plus simple est également présentée dans la [Table 1](#).
3. La GSE Upload Sales Grid, qui sert à charger la Sales Grid dans des logiciels tels qu'Alamode. On notera toutefois que tous les logiciels de formulaires d'évaluation ne permettent pas de charger les données d'une Sales Grid à partir d'une feuille de calcul. Par nécessité, de telles grilles ne peuvent inclure de colonne distincte pour les contributions à la valeur, notion dont les GSE, telles que Fannie Mae, n'ont aucune compréhension. La stratégie consiste à leur fournir le formulaire qu'elles exigent, tout en l'étayant par une Sales Grid RCA ainsi que par d'autres pièces justificatives, telles que des précisions sur les éventuelles agrégations de variables.

Les RCA Processing Sales Grids ne ressemblent pas aux Sales Grids de Fannie Mae mais en constituent, grosso modo, une transposée, lignes et colonnes étant interverties : chaque ligne contient les données d'un bien, et chaque colonne les données d'une variable du bien. On peut facilement compter des centaines, voire des milliers de lignes de biens et 200 colonnes ou plus. La première ligne est réservée aux noms des variables, et la deuxième à un bien à évaluer, même lorsque l'on ne réalise qu'une analyse de marché. Lorsque la feuille de calcul est lue dans un data frame, la première ligne de la feuille devient l'attribut names, et la deuxième ligne, celle du bien à évaluer, devient la première ligne du data frame.

Les colonnes de la Sales Grid RCA comprennent des colonnes d'identification, puis les variables primaires, externes et synthétiques, combinées et fournies à MARS en tant que variables indépendantes. MARS sélectionne parmi ces variables celles qu'il estime avoir une incidence significative sur le prix de vente, également appelé variable dépendante ou variable cible. En régression de degré deux, on peut obtenir des paires de ces variables agissant comme de nouvelles variables. En régression de degré trois, on peut également obtenir des triplets de variables agissant

comme de prétendues variables indépendantes, bien qu'elles soient, évidemment, fortement corrélées avec les variables sous-jacentes.

TABLE 1 – Exemple de feuille de calcul RCA : disposition verticale

| Bien à évaluer                  |                          |                |                          | Comparable 1            |             |                          | Comparable 2            |                |  |
|---------------------------------|--------------------------|----------------|--------------------------|-------------------------|-------------|--------------------------|-------------------------|----------------|--|
| Mesure                          | Contribution à la valeur | Mesure         | Contribution à la valeur | Ajustement              | Mesure      | Contribution à la valeur | Ajustement              |                |  |
| Prix de vente                   |                          | \$1,150,000    |                          |                         | \$1,500,000 |                          |                         |                |  |
| Valeur de base                  | \$500,000                |                | \$500,000                | \$0.00                  |             | \$500,000                | \$500,000               | \$0            |  |
| Caractéristiques mesurées       |                          |                |                          |                         |             |                          |                         |                |  |
| Surface habitable               | 1,530                    | \$463,500      | 1,350                    | \$382,500               | \$81,000    | 1,700                    | \$540,000               | -\$76,500      |  |
| Superficie du terrain           | 6,000                    | \$200,000      | 6,500                    | \$220,000               | -\$20,000   | 7,000                    | \$240,000               | -\$40,000      |  |
| Âge                             | 53                       | \$63,600       | 57                       | \$68,400                | -\$4,800    | 51                       | \$61,200                | \$2,400        |  |
| Salles de bains                 | 2                        | \$15,000       | 3                        | \$30,000                | -\$15,000   | 3                        | \$30,000                | -\$15,000      |  |
| Chambres                        | 3                        | \$6,000        | 4                        | \$9,000                 | -\$3,000    | 4                        | \$9,000                 | -\$3,000       |  |
| Total mesuré                    |                          | \$1,248,100.00 |                          | \$1,209,900.00          | \$38,200    |                          | \$1,380,200.00          | -\$132,100     |  |
| Caractéristiques résiduelles    |                          |                |                          |                         |             |                          |                         |                |  |
| Résidu                          | CQA 5.8                  | \$30,000       | CQA 2.5                  | -\$59,900               | \$89,900    | CQA 8.2                  | \$119,800               | -\$89,800      |  |
| Conception                      | Décomposition du résidu  |                |                          | Décomposition du résidu |             |                          | Décomposition du résidu |                |  |
| État                            |                          | 7,000          |                          | -\$15,000               | \$22,000    |                          | \$35,000                | -\$28,000      |  |
| Qualité                         |                          | 5,000          |                          | -\$8,000                | \$13,000    |                          | \$20,000                | -\$15,000      |  |
| Utilité fonct.                  |                          | 8,000          |                          | -\$10,000               | \$18,000    |                          | \$30,000                | -\$22,000      |  |
| Équipements                     |                          | 5,000          |                          | -\$12,000               | \$17,000    |                          | \$15,000                | -\$10,000      |  |
|                                 |                          | 5,000          |                          | -\$14,900               | \$19,900    |                          | \$19,800                | -\$14,800      |  |
| Sous-total                      |                          | 30,000         |                          | -\$59,900               | \$89,900    |                          | 119,800                 | -\$89,800      |  |
| Total                           |                          |                | Total                    |                         |             | Total                    |                         |                |  |
| Contribution totale à la valeur |                          | \$1,278,100.00 |                          | \$1,150,000.00          |             |                          | \$1,500,000.00          |                |  |
| Prix de vente ajusté            |                          |                |                          | \$1,278,100.00          |             |                          |                         | \$1,278,100.00 |  |

[illegible]

## 8 Variables

Du côté des entrées de l'exécution de MARS, on ne dispose que de variables simples, à savoir :

1. Les caractéristiques du bien chargées depuis la feuille de calcul MLS.
2. Les variables dérivées créées par le VE afin d'affiner les variables MLS ou d'ajouter des transformations de variables existantes. Par exemple, le VE peut préférer transformer la date de vente en « DaysOffMarket », qui est égale à la date d'effet de l'évaluation moins la date de vente.
3. Le VE peut importer des données de sources tierces, telles que les taux d'intérêt hypothécaires. De même, si les données de latitude, de longitude ou d'altitude sont absentes des données MLS, il peut les importer d'autres sources. Ces variables pourraient être appelées « variables externes ».

Du côté des sorties de MARS, on travaille avec un modèle de prix qui est une combinaison de termes qui, soit :

1. Contiennent une seule des variables d'entrée et fournissent l'équation de la contribution à la valeur de cette variable. On les appelle termes du premier degré ou variables (de sortie). Le terme lui-même est traité comme une valeur de contribution à la valeur et reçoit un nom.
2. Contiennent deux des variables d'entrée et constituent l'équation de la contribution à la valeur de l'interaction des deux variables considérées. Ces termes peuvent être appelés termes du deuxième degré ou variables (de sortie).
3. Contiennent trois des variables d'entrée et constituent l'équation de la contribution à la valeur de l'interaction entre les trois variables. Ces termes peuvent être appelés, dans le présent article, termes du troisième degré ou variables (de sortie).

Les variables de sortie de MARS sont donc des contributions à la valeur, y compris la contribution à la valeur résiduelle. Elles sont à leur tour traitées comme des variables dans les calculs de la Sales Grid, tels que le calcul des ajustements ou la création de graphiques.

On peut avoir différents types de variables qui relèvent de deux ontologies fondamentales : (1) l'origine des variables et (2) l'application des variables, c'est-à-dire non résiduelles vs résiduelles.

### 8.1 Ontologie de l'origine des variables

1. Variables du premier degré
  - 1.1 Variables de régression MLS
  - 1.2 Variables de régression externes
  - 1.3 Variables de régression dérivées

2. Variables facteur :

Pour traiter les variables facteur, une fonction capable de les identifier est nécessaire. Puisqu'elles sont spécifiées dans le classeur Excel de configuration du projet, une fonction est créée qui renvoie vrai si la variable qui lui est passée a été classée comme étant de type facteur.

3. Variables d'interaction : elles sont créées par MARS lors de la génération du modèle, si le degré est fixé à 2 ou plus. Toutefois, trois variables en interaction constituent le maximum qui puisse être recommandé, en raison de la difficulté d'interprétation, de la probabilité de surajustement et de l'impact sur les degrés de liberté, ainsi que de l'augmentation résultante du nombre d'enregistrements de biens nécessaires pour garantir un modèle robuste.

- 3.1 Variables du deuxième degré : elles sont courantes, et de bons exemples seraient GLA :BathRms, GLA :LotSize, DateOfSale :BathRms, Latitude \_\_Longitude.
- 3.2 Variables du troisième degré : interactions moins courantes qui appellent à la prudence et suggèrent un possible surajustement.
- 4. Variables de décomposition du résidu

## 8.2 Ontologie de l'application des variables

- 1. Contributions à la valeur générées par MARS
  - 1.1 Variables du premier degré
    - 1.1.1. Variables de régression MLS
    - 1.1.2. Variables de régression externes
    - 1.1.3. Variables de régression dérivées
  - 1.2 Variables d'interaction
    - 1.2.1. Variables du deuxième degré
    - 1.2.2. Variables du troisième degré
  - 1.3 Variables de décomposition du résidu
  - 1.4 Variables d'agrégation
- 2. Variables & valeurs générées par l'ingénieur
  - 2.1 Variables de décomposition du résidu
    - 2.1.1. Conception/Esthétique
    - 2.1.2. Aménagement paysager
    - 2.1.3. Utilité fonctionnelle
    - 2.1.4. Vue
    - 2.1.5. État
    - 2.1.6. Qualité
    - 2.1.7. ...

Ces groupes de variables recevront leur notation et seront décrits plus en détail dans la section suivante.

### Protocole 3.1 : Variables facteur

Les variables facteur, ou simplement les facteurs, peuvent apparaître dans tous les termes, y compris les interactions. Elles sont importantes parce qu'elles doivent faire l'objet d'un traitement particulier pour la régression MARS et la représentation graphique.

Si les variables à valeurs de type caractère sont nécessairement des facteurs, les variables numériques doivent être signalées comme variables facteur pour que MARS les reconnaisse comme telles. Par exemple, les zones MLS sont souvent identifiées par un numéro. En particulier, les quartiers de la San Francisco Bay Area sont souvent identifiés par des entiers uniques. Si ces valeurs sont transmises à MARS en tant que facteurs, MARS tentera alors de découvrir si certaines zones présentent une valeur contributive moyenne pour les ventes de biens situés à l'intérieur de leurs limites, sans égard à une quelconque relation avec d'autres zones. En revanche, avec des variables ordinaires, il supposera qu'il existe une certaine forme de continuité dans les contributions à la valeur à mesure que la valeur de la variable change. C'est une distinction importante. En outre, ces variables facteur peuvent apparaître dans des termes du modèle avec des variables non-facteur, ce qui rend la représentation graphique difficile. Plus précisément, lors du tracé des termes à variable unique comportant des facteurs, on regroupe tous ces facteurs et on trace simplement un diagramme en barres montrant leurs contributions, positives ou négatives. Si les facteurs apparaissent avec d'autres variables dans des termes, ils apparaîtront toujours avec une seule valeur possible, car les variables facteur ne peuvent avoir qu'une seule valeur pour un bien donné : si une zone a la valeur 662, par définition elle ne peut avoir aucune autre valeur pour cette zone. Si cela vous semble déroutant, pour ne rien arranger, dans le modèle généré par MARS, il n'existe pas un nom de variable unique pour le facteur ; MARS prendra le nom racine de la variable facteur qui lui a été fournie en entrée, puis créera de nouvelles colonnes dont les noms de variables sont formés en y accolant les valeurs possibles. Ainsi, une seule colonne de la feuille de calcul Excel, telle que `MlsArea`, donnera lieu à la création de plusieurs colonnes de data frame pour chaque valeur possible, telles que `MlsArea660`, `MlsArea661`, etc. Une variable telle que `MlsArea661` aura la valeur 1 pour les biens situés dans la zone 661, et 0 sinon. MARS effectuera alors la régression sur ces colonnes comme si elles représentaient des variables booléennes. Ces noms modifiés seront utilisés dans le modèle fourni par MARS.

On voudra vraisemblablement reconvertir ces noms du modèle vers leur nom d'origine afin de rendre les graphiques plus compréhensibles pour des clients qui apprécient la simplicité (du moins autant que possible). Au lieu de voir `MlsArea661=1`, on verra `MlsArea=661`. On peut alors représenter graphiquement les termes en indiquant simplement dans l'en-tête du graphique une mention du type « Pour MLS Area = 662 ». Ainsi, les termes de degré N comportant M variables facteur sont tracés comme des termes de degré (N-M), et la dimension du facteur disparaît dans le titre.

Il convient de noter que deux variables facteur ou plus pourraient exister dans un même terme, de sorte qu'un graphique serait vraisemblablement nécessaire pour chaque paire de valeurs possible. Si la variable x, avec trois valeurs distinctes, et la variable y, avec deux valeurs distinctes, apparaissent dans le même terme avec la variable non-facteur z, il faudrait alors  $2 \cdot 3 = 6$  graphiques. Un exemple serait « `MlsArea` » + « `Architectural Style` » : par exemple, les maisons victoriennes de San Francisco peuvent valoir plus ou moins selon le quartier de San Francisco où elles se trouvent. Heureusement, dans la plupart des zones, il n'y a généralement que 1-3 styles architecturaux suffisamment importants comme contributeurs de valeur pour être utilisés dans le modèle (bien qu'il puisse exister 6+ styles architecturaux différents).

### 8.3 Termes du deuxième et du troisième degré

Avec une régression du deuxième ou du troisième degré, certains termes du modèle de sortie peuvent inclure respectivement deux ou trois variables. Par exemple :

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

On remarquera que l'on nomme ces termes en triant les noms de variables par ordre alphabétique puis en les concaténant, éventuellement avec un tiret bas. On procède de même pour les termes du troisième degré, sur les triplets de noms de variables produits dans le modèle. Par exemple, « BathRms\_GLA\_LotSz ».

Ces variables sont ajoutées aux variables d'entrée primaires à l'étape 3, lorsque le programme RCA calcule les contributions de toutes les variables pour chaque bien, y compris les variables d'interaction.

#### 8.3.1 Variables d'ajustement

Le programme RCA calculera également, à l'étape 3, les ajustements pour l'ensemble des variables et des biens, une fois les contributions à la valeur calculées, en soustrayant la contribution à la valeur du bien comparable correspondant, pour une caractéristique donnée, de la contribution du bien à évaluer pour cette même caractéristique.

#### 8.3.2 Variables d'agrégation

Les variables d'agrégation sont définies pour remplacer des sous-ensembles de variables, en additionnant toutes leurs contributions à la valeur et leurs ajustements dans une variable commune destinée à être utilisée dans le rapport. Par exemple, on peut avoir scindé la surface habitable, achevée comme inachevée, en variables de détail plus fines :

1. Surface achevée légale hors sol
2. Surface achevée légale en sous-sol
3. Surface achevée illégale hors sol
4. Surface achevée illégale en sous-sol
5. Surface inachevée hors sol
6. Surface inachevée en sous-sol

Oui, il existe des zones où cela a effectivement du sens.

La régression MARS portera sur toutes ces variables et générera un modèle qui fournira des contributions et des ajustements pour chacune d'elles. Mais pour un rapport sur formulaire, ces différentes contributions devront probablement être combinées ou agrégées en un nombre plus restreint de variables. Les détails de l'agrégation devront être fournis dans une annexe. Un exemple plus fréquent consisterait à agréger les ajustements d'interaction dans l'un des ajustements des variables primaires. Par exemple, on pourrait agréger l'ajustement de GLA\_LotSize dans GLA ou dans LotSize, ou même le répartir 50 :50 entre les deux, selon ce que détermine la feuille Aggregation.

### 8.3.3 Variables spéciales & notation

|                  |                                                                                                                                                    |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| $\pi_i$          | Prix de vente net effectif du bien $i = 1, 2, \dots, n_p$ ; pour le bien à évaluer, $\pi_0$ désigne la conclusion de valeur (voir Démonstration I) |
| $\alpha_i$       | Prix de vente net ajusté du bien $i = 1, 2, \dots, n_p$                                                                                            |
| $\rho_i$         | Prix de vente net estimé par MARS pour le bien $i = 0, 1, \dots, n_p$                                                                              |
| $\beta$          | Valeur de base du modèle                                                                                                                           |
| $\varphi_{i,j}$  | Contribution à la valeur de la variable de régression $j = 1, 2, \dots, n_r$ pour le bien $i = 0, 1, \dots, n_p$                                   |
| $\delta_{i,j}$   | Ajustement de la variable de régression $j = 1, 2, \dots, n_r$ pour le bien $i = 1, 2, \dots, n_p$                                                 |
| $\epsilon_i$     | Résidu total du bien $i = 0, 1, \dots, n_p$                                                                                                        |
| $\epsilon_{i,c}$ | Contribution à la valeur de la composante résiduelle $c$ , $c = 1, 2, \dots, n_c$ , pour le bien $i = 0, 1, \dots, n_p$                            |
| $\xi_{i,c}$      | Ajustement de la composante résiduelle $c$ , $c = 1, 2, \dots, n_c$ , pour le bien $i = 0, 1, \dots, n_p$                                          |

où

$n_p = \#$  of comparable properties

$n_r = \#$  of regression variables

$n_c = \#$  of residual breakdown variables

### 8.3.4 Décomposition des résidus

Une fois que la régression MARS a créé un modèle, elle génère les prix de vente estimés à partir des variables d'entrée sélectionnées par MARS. Elle soustrait cette valeur du prix de vente net correspondant afin d'obtenir une valeur résiduelle pour chaque bien. Le protocole RCA stipule alors qu'une fois que le VE a arrêté son choix définitif de biens comparables, généralement 6-12 listings, il doit passer en revue chaque listing et décomposer le résidu calculé selon les caractéristiques (variables latentes) les plus vraisemblables qui, selon lui, influencent la valeur, en ignorant la possibilité d'erreurs de données aléatoires. Il pourrait, par exemple, retenir les variables de décomposition suivantes pour un bien donné :

1. Conception
2. Utilité fonctionnelle
3. Vente au-dessus/au-dessous du marché
4. État
5. Qualité

Des biens différents auront vraisemblablement des variables de décomposition du résidu en partie différentes, qu'il convient de rassembler dans un seul ensemble ordonné.

**Estimation des contributions des caractéristiques rares.** Certaines caractéristiques porteuses de valeur sont trop rares pour le canal de la régression : un atelier présent dans 3 ventes sur 780 ne peut se voir attribuer un coefficient défendable, et la passe avant ignorera généralement une telle variable entièrement. Ce sont des caractéristiques résiduelles, et leurs contributions peuvent être estimées *dans l'espace des résidus* : on ajuste le modèle sans la caractéristique, puis on régresse les résidus du modèle sur la ou les colonnes exclues par les moindres carrés ordinaires. Pour un indicateur binaire, le coefficient est une contribution forfaitaire (de manière équivalente, le décalage résiduel moyen des ventes signalées) ; pour une colonne numérique, il s'agit d'une contribution unitaire. Lorsque plusieurs aménités corrélées sont estimées *conjointement* (p. ex. la surface des dépendances conjointement avec un indicateur d'atelier), la régression répartit le crédit entre elles plutôt que de compter deux fois les mêmes dollars dans deux lignes de décomposition. Les estimations fondées sur moins d'une dizaine d'observations sont seulement indicatives et devraient être corroborées par le coût déprécié ou par un échantillon élargi (p. ex. des indicateurs récupérés à partir des remarques publiques) ; la contrainte résiduelle elle-même plafonne quelle que soit la valeur finalement attribuée. earthUI (v0.11+) met en œuvre cette méthode directement sous la forme de l'onglet *Residual Contribution*, y compris l'estimation conjointe, les éléments de preuve résiduels par vente et une note documentaire pour le dossier de travail.

Notation : on utilise la notation du tableau ci-dessus pour arriver à :

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

où  $n_c$  est le nombre de composantes de la décomposition du résidu.

Le VE doit exercer son jugement pour décomposer le résidu selon les composantes données. On pourrait penser que cela permettrait à un biais de s'introduire dans la conclusion de valeur finale. Or, il n'en est rien, car on dispose d'une démonstration mathématique établissant que si les ajustements de la décomposition totalisent le résidu, ils n'auront aucune incidence sur la conclusion de valeur. L'objet de la décomposition n'est pas d'établir la valeur, mais d'expliquer pourquoi le résidu est ce qu'il est en comparaison des résidus des ventes des autres biens. Le seul point où le biais de l'ingénieur peut influencer sur la conclusion de valeur finale est l'estimation du résidu du bien à évaluer puisque, bien entendu, il n'existe pas de prix de vente effectif pour le bien à évaluer permettant de calculer un résidu pour celui-ci.

Chaque bien possède son propre ensemble de contributions à la valeur résiduelles, mais ceux-ci doivent être fusionnés en un seul ensemble C couvrant tous les biens :

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

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

où  $[\epsilon_{i,c}]$  est le nom de variable associé à la valeur  $\epsilon_{i,c}$ ,  $A_i$  est l'ensemble de ces variables et  $n_c$  est

le nombre de valeurs dans chaque  $\epsilon_i$ .

Ainsi, les données du bien à évaluer et des biens comparables figurant dans des colonnes séparées, on devrait voir, sous l'ensemble des contributions à la valeur des variables du modèle de régression pour chaque bien, l'ensemble C des  $n_c$  contributions à la valeur résiduelles correspondantes.

## 9 Démonstrations fondamentales de la RCA

Posons :

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

**Démonstration I : Chaque prix de vente ajusté de chaque bien comparable est égal au prix de vente estimé du bien à évaluer.**

Pour tout bien comparable  $i$ , le prix de vente ajusté  $\alpha_i$  est la somme du prix de vente net  $\pi_i$  et des ajustements : les différences entre les contributions à la valeur du bien à évaluer et celles du bien comparable pour la valeur de base, les variables de régression  $j = 1, 2, \dots, n_r$ , et le résidu. Pour le bien à évaluer, qui ne fait l'objet d'aucune vente effective, le prix de vente net  $\pi_0$  est défini comme la conclusion de valeur  $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$ . Alors :

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

**Démonstration II : Modifier la décomposition du résidu, sous la contrainte que toutes les contributions à la valeur des composantes résiduelles se somment au résidu du bien comparable, n'a aucune incidence sur la conclusion de valeur finale du protocole RCA.**

*C'est presque trivial. Pourtant, sans réfléchir aux mathématiques, il est facile de passer à côté, car les évaluateurs traditionnels se concentrent sur les ajustements plutôt que sur les contributions à la valeur. Ce sont les contributions à la valeur qui s'annulent ici — un oubli vieux de plusieurs décennies.* Pour un bien comparable  $k$ ,  $k > 0$ , l'ajustement du résidu total est égal à la somme des  $n_c$  composantes résiduelles de la décomposition, laquelle est à son tour égale à la somme des différences de contributions à la valeur, elle-même égale au résidu total du bien à évaluer moins le résidu total du bien comparable. Et cela est indépendant du nombre de composantes résiduelles créées ou de ce qu'elles représentent. Tant que la valeur des composantes résiduelles

est égale au résidu total, tant pour le bien à évaluer que pour le bien comparable choisi,  $\xi_k$  ne changera pas.

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

## 10 Modèle MARS

### 10.1 Fonctions « $g$ » de variables du modèle et d'interaction

Un principe fondamental de la méthode RCA est que le « raisonnement explicatif » revêt une importance capitale. Malheureusement, produire des graphiques parlants pour les diverses équations de contribution à la valeur, en particulier lorsque des interactions et des variables facteurs sont utilisées, peut s'avérer difficile. Il est donc jugé impératif de fournir un protocole clair et strict pour la création de ces graphiques. C'est ce qui est fait ici ; il convient toutefois de comprendre que ce protocole s'adresse au développeur qui écrit le code R ou Python générant les graphiques, et qu'il ne présente peut-être qu'un intérêt superficiel pour les personnes non impliquées dans la programmation.

Le modèle de prix généré par MARS est la somme d'une constante de base et de termes additionnels, chacun faisant intervenir 1 à 3 variables. Pour l'analyser, on décompose d'abord le modèle en termes individuels. Ensuite, les termes partageant le même ensemble de variables, représentant une seule interaction, sont regroupés en expressions d'interaction uniques. Supposons que le degré maximal de la régression soit 3. Considérons un modèle de régression de troisième degré composé d'une constante de base, de  $n_1$  expressions de premier degré, de  $n_2$  expressions d'interaction de deuxième degré et de  $n_3$  expressions d'interaction de troisième degré. On obtient ainsi quatre groupes distincts d'expressions, qui sont soit des expressions à variable unique, soit des interactions à plusieurs variables, chacune représentée graphiquement de manière différente selon deux facteurs : (1) la présence ou non d'une variable facteur (une variable catégorielle), et (2) le nombre de variables uniques qu'elles contiennent. Plus précisément :

- Les expressions de premier degré (1 variable) sont représentées en 2D.
- Les expressions de deuxième degré (2 variables) sont représentées en 3D.
- Les expressions de troisième degré (3 variables) sont représentées par un *ensemble* de graphiques 3D.
- Si une expression contient une variable facteur, chaque facteur unique réduit la dimensionnalité du graphique d'une unité (p. ex. un graphique 3D devient 2D si l'une des variables est un facteur).

L'indexation de chaque fonction  $g$  fournit les informations nécessaires pour déterminer comment elle doit être représentée graphiquement, en reflétant le nombre de variables uniques et la présence de variables facteurs.

### 10.2 Indexation des fonctions $g$

Chaque fonction  $g$  est indexée sous la forme  $f_{j,k} g_k^j$ , où :

- $f_{j,k}$  est le premier indice (un exposant placé avant  $g$ ) :
  - $j$  varie de 0 à 3 et identifie le groupe :
    - $j = 0$  : la constante de base (une exception à l'indexation habituelle).
    - $j = 1$  : expressions de premier degré.
    - $j = 2$  : expressions de deuxième degré.
    - $j = 3$  : expressions de troisième degré.
- $k$  est l'indice d'une fonction particulière au sein de son groupe (p. ex.  $k = 1$  pour la première fonction du groupe  $j$ ).
- La valeur de  $f_{j,k}$  indique le nombre de variables facteurs uniques dans la fonction :
  - $f_{j,k} = 0$  : aucune variable facteur.
  - $f_{j,k} > 0$  : le nombre de variables facteurs (p. ex. 1, 2 ou 3).

- Indices postposés de  $g$  :
  - L'exposant  $j$  après  $g$  (p. ex.  $g^j$ ) répète le numéro du groupe par souci de clarté.
  - L'indice inférieur  $k$  après  $g$  (p. ex.  $g_k$ ) correspond au  $k$  de  $f_{j,k}$  et identifie la fonction au sein de son groupe.

Bien que les indices de  $f$  et de  $g$  coïncident numériquement, leurs rôles diffèrent :  $f_{j,k}$  compte les variables facteurs, ce qui est déterminant pour les choix de représentation graphique, tandis que  $j$  et  $k$  sur  $g$  définissent son groupe et sa position.

### 10.3 Dimensions de représentation graphique

La dimension  $d$  du graphique d'une fonction  $f_{j,k} g_k^j$  se calcule comme suit :

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

où :

- $j$  est le nombre de variables uniques (ou le numéro du groupe, en ignorant le cas de la constante).
- $f_{j,k}$  est le nombre de variables facteurs.
- $d$  est le nombre de variables continues (non facteurs) à représenter, avec un maximum de 3 (puisque le degré maximal est 3). Une fonction à  $d$  variables continues se dessine en  $d + 1$  dimensions :  $d = 1$  donne un graphique 2D,  $d = 2$  un graphique 3D (ou une carte thermique 2D), et  $d = 3$  un *ensemble* de graphiques 3D.

Par exemple :

- Si  $j = 3$  (3 variables) et  $f_{j,k} = 0$  (aucune variable facteur), alors  $d = 3 - 0 = 3$ , ce qui exige un ensemble de graphiques 3D, chacun fixant l'une des variables.
- Si  $j = 3$  et  $f_{j,k} = 1$  (une variable facteur), alors  $d = 3 - 1 = 2$ , ce qui exige un graphique 3D (ou une carte thermique 2D) pour chaque valeur de la variable facteur apparaissant dans le modèle.
- Si  $j = 0$  (la constante), une représentation graphique n'est généralement pas nécessaire, puisqu'il s'agit d'une valeur unique.

Lorsque  $d = 3$  (p. ex. une expression de troisième degré sans variable facteur), la visualisation de la fonction complète peut nécessiter plusieurs graphiques 3D, chacun fixant une variable à différentes valeurs afin de montrer un sous-ensemble du comportement de la fonction.

### 10.4 L'équation finale du modèle

Le modèle MARS est décomposé en quatre groupes d'expressions (constante, premier, deuxième et troisième degré). L'indexation  $f_{j,k} g_k^j$  nous indique :

- $j$  : le groupe et le nombre de variables.
- $f_{j,k}$  : le nombre de variables facteurs.
- $d = j - f_{j,k}$  : la dimension du graphique.

Cette structure guide la visualisation de chaque fonction  $g$ , en garantissant que la méthode de représentation graphique corresponde à la complexité de l'expression et aux types de variables.

L'estimation du prix de vente  $\hat{P}$  fournie par le modèle MARS peut s'exprimer ainsi :

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

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

## 10.5 Nombre de graphiques nécessaires pour représenter un modèle

Combien de graphiques faut-il pour visualiser toutes les fonctions d'un modèle ? La réponse dépend des variables, de leurs interactions et de leur nature de variable facteur ou non.

Le VE peut restreindre les variables et les interactions dans le fichier de configuration Excel destiné à R/earth, ce qui influe sur les interactions autorisées. Une approche pratique consiste à laisser R/earth évaluer toutes les interactions possibles et sélectionner celles qui apportent une contribution significative à l'estimation du prix de vente, tout en vérifiant l'absence de problèmes tels que le surajustement ou la colinéarité. Certaines interactions peuvent être interdites, ce qui se répercute sur les interactions de degré supérieur (p. ex. si une paire est exclue dans une interaction de troisième degré, la combinaison entière est exclue).

Pour les variables facteurs (p. ex. MLS Area avec 12 valeurs en interaction avec GLA), la représentation graphique devient plus complexe. Cela pourrait nécessiter 12 graphiques—un par zone. Si un autre facteur comme Architectural Style (6 valeurs) interagit avec MLS Area et GLA, jusqu'à  $12 \cdot 6 = 60$  graphiques pourraient être nécessaires. Toutefois, R/earth ne juge que rarement toutes les combinaisons significatives ; en général, seules quelques interactions clés (p. ex. certaines zones et certains styles) méritent d'être représentées, ainsi que le modèle l'identifie.

Voici un guide simplifié du nombre de graphiques nécessaires :

1. **Fonctions à variable unique (non facteur) :** 1 graphique.
2. **Fonctions à variable unique (facteur) :** 1 graphique en barres regroupant toutes les valeurs du facteur.
3. **Fonction de base :** aucun graphique nécessaire.
4. **Interactions de 2<sup>e</sup> degré (sans facteur) :** 1 graphique (p. ex. carte thermique 2D avec des couleurs pour le prix de vente).
5. **Interactions de 2<sup>e</sup> degré (1 facteur) :** 1 graphique par valeur du facteur présente dans le modèle (p. ex. 2 valeurs = 2 graphiques).
6. **Interactions de 2<sup>e</sup> degré (2 facteurs) :** 1 graphique avec une matrice des combinaisons de facteurs (p. ex. 3 zones  $\times$  2 styles = 6 cellules).
7. **Interactions de 3<sup>e</sup> degré (sans facteur) :** plusieurs graphiques (p. ex. 4 graphiques pour un nombre de salles de bains = 1, 2, 3, 4).
8. **Interactions de 3<sup>e</sup> degré (1 facteur) :** 1 graphique par valeur du facteur présente dans le modèle.
9. **Interactions de 3<sup>e</sup> degré (2 facteurs) :** 1 graphique par combinaison de valeurs des facteurs présente dans le modèle.
10. **Interactions de 3<sup>e</sup> degré (3 facteurs) :** 1 graphique par valeur du facteur ayant le moins de valeurs distinctes, avec une matrice de couleurs pour les deux autres.

Des scénarios plus complexes peuvent exiger des techniques de représentation graphique plus avancées.

D'autres possibilités existent, avec des graphiques vraisemblablement plus sophistiqués.

Un exemple de modèle complet généré par MARS figure dans [Figure 1](#).



## 11 Caractéristiques des données

La section suivante traite de MARS. MARS dépend entièrement des données qu'il reçoit, et la plupart de ces données proviendront d'un MLS (Multiple Listing Service) ou des Tax Assessors (administration d'évaluation fiscale). Toutefois, elles peuvent transiter par d'autres fournisseurs de données qui entretiennent et améliorent les données, puis les revendent avec profit. Dans certaines zones, elles sont assez fiables ; dans d'autres, elles le sont moins. Dans les cas critiques, l'inspecteur du bien doit effectuer lui-même ses observations et ses mesures, dont l'exactitude peut être mise en question.

Le Valuation Engineer doit comprendre ce que les données représentent et dans quelle mesure elles remplissent bien leur rôle de représentation. Cela soulève toute une série de questions :

- Comment classer une variable comme étant une mesure ?
- Si une variable n'est pas une mesure, que peut-elle être d'autre ?
- Comment la régression peut-elle être utile si les données qu'elle reçoit ne sont pas exactes ?

La plupart des données que nous obtenons des services de listings MLS et des Tax Assessors aux fins de la régression sont des données numériques qui impliquent des mesures ou des dénombrements. Certaines sont logiques, Vrai ou Faux.

### 11.1 Rôle du résidu

Le résidu sert d'indicateur indirect de la valeur ajoutée par les variables non incluses dans le modèle de régression. Dans le langage de la littérature hédonique, ces caractéristiques non mesurées sont des *variables latentes* : des variables qui influencent le prix de vente mais ne sont pas mesurées directement, le plus souvent parce qu'il n'existe aucun moyen pratique de les mesurer objectivement. Leurs valeurs doivent au contraire être inférées et, en pratique, la valeur d'une variable latente est souvent estimée par un jugement subjectif — dans la RCA, un jugement ancré dans le classement des résidus. En ordonnant les biens selon leur résidu ou leur résidu par pied carré, on obtient un classement objectif allant des biens présentant le moins d'attrait à ceux qui en présentent le plus. Cette méthode de classement est efficace, à condition qu'un analyste qualifié doté d'une connaissance approfondie du marché local élabore le modèle MARS (Multivariate Adaptive Regression Splines). Voici quelques exigences supplémentaires :

- Qualité des données et du modèle : un modèle de régression bien construit et des données de haute qualité sont des préalables à des résidus fiables. Dans la San Francisco Bay Area, mon objectif est de développer un modèle présentant un  $R^2$  d'environ 75-80% et un  $R^2$  de validation croisée (CV  $R^2$ ) de 55-70%.
- Éviter le surajustement, limiter le  $R^2$  : le Valuation Engineer devrait viser un  $R^2$  maximal d'environ 80% dans la plupart des marchés matures des zones à prix élevés comme la SF Bay Area, car une grande partie de la valeur concurrentielle d'un bien résidentiel réside dans son attrait subjectif, qui n'est pas quantifiable et ne peut donc pas être intégré directement dans les modèles de régression. Un  $R^2$  nettement supérieur à 80% peut indiquer un modèle surajusté et appeler un examen plus approfondi. Une mise en garde s'impose toutefois : il existe des zones de marché présentant des valeurs de  $R^2$  supérieures à 90% sans surajustement, et il existe des zones difficiles à évaluer où le mieux que l'on puisse obtenir est de 50-60% ou moins.
- $R^2$  élevé dans certaines zones : il existe cependant des régions, comme les lotissements ou d'autres zones résidentielles uniformes, où des valeurs de  $R^2$  plus élevées sont étonnamment courantes en raison de l'homogénéité des biens.

## 11.2 Les notations d'état et de qualité comme régresseurs approximatifs

Les échelles de notation d'état et de qualité de Fannie Mae (C1–C6 et Q1–Q6) méritent un commentaire particulier. Ces échelles sont trop grossières pour être traitées comme de véritables caractéristiques mesurées et, parce qu'elles sont attribuées par l'évaluateur, elles constituent en pratique des notations subjectives — des variables latentes à part entière. Il est utile de distinguer explicitement deux niveaux : la notation grossière de l'évaluateur utilisée comme régresseur, *état à gros grain* (condition-1) et *qualité à gros grain* (quality-1), par opposition aux variables latentes à grain fin extraites du résidu, *état à grain fin* (condition-2) et *qualité à grain fin* (quality-2). Le VE peut accepter les notations à gros grain de l'évaluateur comme approximations et effectuer la régression sur celles-ci, obtenant ainsi des contributions à la valeur approximatives qui sont utiles comme valeurs initiales pour les variables latentes à grain fin. Lorsque l'on procède ainsi, le protocole consiste à soustraire de l'estimation du modèle les contributions à la valeur estimées par la régression pour ces deux variables à gros grain — en les retirant des attributs mesurés figurant dans la section régression de la Sales Grid — et à les rajouter au résidu. De là, elles sont ensuite retirées du résidu et affectées aux variables latentes à grain fin correspondantes, `res_condition1` et `res_quality1`. Des informations supplémentaires extraites des remarques publiques du MLS (`public_remarks`) et des listes de chaînes de caractères associées sont ensuite utilisées pour apporter des micro-ajustements aux valeurs de ces variables latentes. Le résultat final est que les contributions à la valeur à grain fin remplacent les contributions à la valeur à gros grain : les notations grossières ne servent que d'échafaudage pour le modèle approximatif, et les valeurs finalement rapportées pour l'état et la qualité sont celles à grain fin, dérivées du résidu.

## 11.3 Erreurs de données

D'après des commentaires sur les réseaux sociaux, les données du Multiple Listing Service (MLS) de Californie sont généralement de meilleure qualité que celles de nombreuses autres régions des États-Unis. En règle générale, on peut s'attendre à ce que les données MLS soient de bonne qualité dans la plupart des zones métropolitaines matures, mais de moindre qualité dans les zones rurales.

Bien que des erreurs occasionnelles surviennent dans les données pour diverses raisons, ces erreurs devraient généralement se compenser mutuellement et n'avoir qu'un impact minime sur les moyennes utilisées dans les modèles de prix, pour autant que les données soient suffisantes.

Cependant, des biais systématiques peuvent survenir dans certaines zones. Considérons par exemple un quartier où de nombreuses maisons ont été initialement construites avec un étage complet, mais où certains acheteurs se sont vu offrir la possibilité de supprimer des pièces de l'étage afin de créer des plafonds cathédrale dans une ou plusieurs pièces du rez-de-chaussée, réduisant considérablement la surface habitable brute (GLA). Ces modifications n'ont pas été reportées par la suite dans les registres du Tax Assessor, pour une raison ou une autre. Par conséquent, le MLS peut indiquer 2,400 pieds carrés alors que la valeur mesurée n'est que de 2,000 pieds carrés. Dans de tels scénarios, les données MLS devraient être corrigées avant d'effectuer la régression MARS (Multivariate Adaptive Regression Splines). Il est important de noter que l'obtention de cette information pour les biens comparables peut exiger un certain effort d'investigation. Les divergences devraient être consignées pour référence future.

Compte tenu de ce qui précède, il est raisonnable de supposer que, dans les zones où les données MLS sont de haute qualité, les erreurs qui se répercutent sur le résidu sont minimales par rapport à l'impact des caractéristiques subjectives et non mesurées telles que la conception, l'état et la qualité. C'est une considération importante lors de l'estimation de l'exactitude de la méthode RCA. Par exemple, si la valeur du  $R^2$  pour une zone de marché donnée est de 80%, alors environ 20% du prix est attribué au résidu CQA. Une erreur de  $\pm 10\%$  dans le positionnement du bien à

évaluer au sein du classement des résidus se propage donc en une erreur de  $\pm 10\% \cdot 20\% = \pm 2\%$  dans la conclusion de valeur finale. C'est ainsi que l'on améliore l'exactitude ! Dans de nombreux cas, des estimations de valeur exactes bien en deçà de  $\pm 1\%$  sont possibles.

Par conséquent, traiter les résidus comme des indicateurs indirects de la valeur collective des variables non modélisées dans la régression MARS est raisonnable, à condition que le Valuation Engineer dispose d'une expérience suffisante du marché local pour identifier où et pourquoi existent des biais de mesure systémiques. Cette approche garantit que l'analyse reflète fidèlement les valeurs réelles des biens, tout en reconnaissant les limites et les nuances des données immobilières.

### **Exemple détaillé : mesurer les déterminants de la valeur non mesurés à Sea Ranch**

Lorsque le  $R^2$  atteignable dans une zone de marché est inférieur à ce que l'homogénéité apparente du marché laisserait attendre, la réponse du cadre méthodologique consiste à identifier et à mesurer la variable manquante plutôt qu'à accepter cette limitation. La communauté de Sea Ranch, dans le comté de Sonoma, l'illustre directement.

Voici un cas concret de la réponse « mesurer plutôt qu'accepter ». La communauté de Sea Ranch, dans le comté de Sonoma, est un développement côtier où la distance à l'océan constitue un déterminant de la valeur de premier ordre : une maison située à quelques centaines de pieds de la falaise se vend à un prix sensiblement différent de celui d'une maison située à un quart de mile à l'intérieur des terres, toutes autres caractéristiques égales par ailleurs. Les données MLS de la communauté n'enregistrent pas la distance à l'océan comme variable. Un modèle MARS initial ajusté sur les variables MLS disponibles a produit un pouvoir explicatif inférieur à ce que l'homogénéité apparente du marché aurait dû permettre, avec une structure des résidus qui n'ordonnait pas suffisamment les biens selon une caractéristique visible dans les données — signe qu'un déterminant important manquait au modèle, plutôt que d'un marché véritablement bruité.

Le remède fut la mesure directe. L'auteur s'est procuré une carte de la communauté montrant la disposition des propriétés par rapport au littoral et a mesuré la distance à l'océan pour plusieurs centaines de maisons. Les distances mesurées ont été ajoutées comme variable au jeu de données MLS et la régression MARS a été relancée. La nouvelle variable a été retenue par le modèle comme significative, le pouvoir explicatif a augmenté substantiellement, et le classement des résidus a ensuite ordonné les biens selon des motifs cohérents avec les caractéristiques subjectives restantes — vue, état, qualité — plutôt que selon le signal latent de distance à l'océan qui l'avait auparavant contaminé. L'altitude a été traitée de manière similaire dans des quartiers à flanc de colline où la variable MLS est manquante ou peu fiable.

Le principe méthodologique est général. Lorsque le  $R^2$  atteignable tombe en dessous de ce que l'homogénéité apparente du marché devrait permettre, la première hypothèse est qu'une variable importante manque au modèle, et la réponse appropriée consiste à identifier de quoi il s'agit et à recueillir les mesures directement. Les sources comprennent les cartes communautaires, les données SIG, les registres du Tax Assessor, les registres du service des permis de construire, l'imagerie satellitaire et l'inspection sur le terrain. Le coût d'un tel travail de mesure est réel mais borné ; une fois la variable recueillie pour une zone de marché donnée, elle peut être reportée dans les missions futures dans cette zone. La RCA absorbe proprement les nouvelles variables parce que le cadre méthodologique est indifférent aux variables qui entrent dans la régression MARS — la structure des résidus indique au VE si la nouvelle variable capte un signal réel, et les contributions à la valeur s'ajustent en conséquence.

La réponse alternative — accepter un  $R^2$  faible et rapporter des bandes d'erreur plus larges — est défendable lorsque les déterminants manquants sont véritablement non mesurables (préférences propres à l'acheteur, sentiment de marché transitoire, circonstances de transaction

idiosyncratiques), mais pas lorsqu'ils sont simplement non collectés. Face à un modèle faible, le premier réflexe du VE devrait être de se demander si le signal manquant relève de ce second cas.

## 11.4 Qu'est-ce que l'exactitude ?

Une exactitude parfaite pour une évaluation exige :

1. Conformité à la valeur de marché : la transaction de vente hypothétique du bien à évaluer adhère à 100% à la définition donnée de la valeur de marché
2. Développement d'un modèle exact et robuste : un modèle MARS de haute qualité est développé, avec un  $R^2$  d'environ 75-80% et un CVR2 d'environ 55-70% selon la localisation.
3. Bonne analyse des erreurs de données : toutes les erreurs de données probables pour les variables mesurées sont comprises et contenues, avec peu de répercussions sur les résidus.
4. Prise en compte des biais systémiques : tout biais systémique dans les données doit être pris en compte et faire l'objet d'un ajustement.
5. Analyse du classement des résidus : s'assurer que les biens classés par résidu suivent un schéma raisonnable de caractéristiques, du rang le plus bas au rang le plus élevé. Les anomalies telles que les ventes successorales, les ventes en « short sale » et les ventes aux enchères doivent être examinées, expliquées et éventuellement retirées.
6. Positionnement du bien à évaluer : avec une compréhension approfondie du classement des résidus, le bien à évaluer doit être positionné avec exactitude. Il convient d'identifier un bien présentant un attrait nettement supérieur et un autre présentant un attrait nettement inférieur, puis de placer le bien à évaluer entre les deux et de le noter. On s'efforce ensuite de resserrer autant que possible le classement entre le bien supérieur et le bien inférieur.
7. Raisonnement explicatif : ne jamais sacrifier le raisonnement explicatif à l'exactitude. L'évaluateur doit être capable d'expliquer les différences de valeur entre le bien à évaluer et les biens comparables, du moins à une personne d'intelligence raisonnable. Par exemple, un bon modèle MARS présentant un  $R^2$  et un CVR2 optimaux peut être utilisé pour expliquer les contributions à la valeur des caractéristiques mesurées, tandis que la décomposition du résidu effectuée par le Valuation Engineer peut être utilisée pour expliquer les contributions à la valeur des autres caractéristiques.
8. Surajustement : éviter le surajustement des modèles en n'utilisant pas de caractéristiques du bien susceptibles d'identifier, individuellement ou en combinaison, un ou plusieurs biens. En particulier, éviter d'utiliser des variables facteurs dont les valeurs peuvent désigner un petit nombre de biens. Une variable facteur qui identifie des quartiers dans des jeux de données de taille conséquente est appropriée, mais une variable qui identifie les biens dotés d'une piscine dans une zone où il en existe très peu doit être utilisée avec prudence. Une caractéristique qui identifie des biens uniques, telle que des coordonnées SIG exactes, des numéros de parcelle ou des adresses, est absolument proscrite.

Si le  $R^2$  de la régression est de 80%, on s'attend à ce qu'environ 20% de la valeur du bien à évaluer provienne du résidu. Or, cela n'est qu'approximatif. Nous savons, d'après la courbe CQA-Résidu, que la majeure partie de la distribution des résidus se situe aux extrémités basses et hautes du CQA. Sur la base de la courbe CQA-Résidu et de la pente extrême pour les valeurs basses et hautes du CQA, on pourrait s'attendre à une estimation moins exacte du résidu dans ces zones. Toutefois, cela est contrebalancé par la rareté relative des biens dans ces zones et par la plus grande différence d'attrait observée entre biens adjacents. Ainsi, tandis que l'on pourrait parler de l'impact d'une fourchette de  $\pm 5\%$  dans la plage médiane, aux extrêmes on se préoccuperait plutôt d'une fourchette un peu plus étroite de  $\pm 2 - \pm 3\%$

Le calcul de l'exactitude de la RCA est, en général, assez difficile en raison de la non-linéarité de la courbe CQA-Résidu. Mais il est possible au cas par cas, à condition de disposer d'un classement des résidus raisonnable. La plage des scores CQA va de 0.0 à 10.0. Avec un bon modèle MARS, votre classement des résidus devrait présenter un schéma clair allant du moindre au plus grand attrait, et il ne devrait pas être difficile de noter à  $\pm 5\%$  près de la plage totale, soit  $\pm 0.5$  pour le score CQA. Si la régression comprend 100 biens, cela revient à classer le bien à évaluer avec exactitude au sein d'un groupe de 10 biens. Ainsi, en nous référant aux données de vente de Burlingame du tableau 3, calculons l'exactitude (ou l'erreur) approximative pour des CQA estimés par le VE de 7.5, 5.1 et 1.9, sa conclusion de valeur pour le bien à évaluer étant de \$3,000,000. Notons que, le tableau 3 ne contenant qu'un échantillon des plusieurs centaines de biens utilisés, il faudra interpoler le résidu si nécessaire (en supposant la linéarité). Pour être clair, nous supposons que le VE est, dans chaque cas, très sûr que son classement du bien à évaluer se situe dans une fourchette de  $\pm 5$  biens au sein du classement des résidus portant sur 100 biens seulement.

| CQA= 7.5 $\pm 0.5$ |           |            |                |                    |              |
|--------------------|-----------|------------|----------------|--------------------|--------------|
| CQA                | Résidu    | Diff. abs. | Erreur att.    | Prix de vente hyp. | Erreur en %  |
| 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 $\pm 0.5$ |           |            |                |                    |              |
|-------------------|-----------|------------|----------------|--------------------|--------------|
| CQA               | Résidu    | Diff. abs. | Erreur att.    | Prix de vente hyp. | Erreur en %  |
| 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 $\pm 0.5$ |            |            |                |                    |              |
|--------------------|------------|------------|----------------|--------------------|--------------|
| CQA                | Résidu     | Diff. abs. | Erreur att.    | Prix de vente hyp. | Erreur en %  |
| 2.40               | -\$228,411 |            |                |                    |              |
| 1.90               | -\$276,381 | \$47,970   |                |                    |              |
| 1.40               | -\$326,938 | \$50,557   | $\pm \$49,264$ | \$3,000,000        | $\pm 1.64\%$ |

Enfin, la seule manière objective de vérifier l'exactitude est d'établir un protocole strict (que le présent article ne fait qu'esquisser à grands traits) et d'en vérifier le respect. Cela est nécessaire parce que les prix de vente réels peuvent s'écarter de la valeur de marché en raison des imperfections du marché. La comparaison des rapports d'évaluation et des ventes ultérieures peut fournir une certaine indication de l'exactitude. Cependant, les complexités et l'instabilité du marché immobilier exigent l'expertise d'évaluateurs ou de Valuation Engineers hautement qualifiés pour fournir des estimations de valeur fondées sur des normes et des protocoles établis. Ceux-ci devraient être suffisamment rigoureux pour garantir que des Valuation Engineers compétents parviennent indépendamment à une conclusion de valeur presque identique. Cela devrait constituer un objectif atteignable dans la plupart des cas où de bonnes données sont disponibles.

## 11.5 Le bien à évaluer comme anomalie

Puisque l'estimation de la valeur de marché d'un bien exige l'analyse de transactions de vente récentes, il est légitime de se demander si les transactions récentes représentent bien les diverses

caractéristiques d'un bien à évaluer plus ancien, dépourvu de modernisation et de réparations. Cela est particulièrement vrai dans la San Francisco Bay Area, où les transactions récentes concernent généralement des maisons ayant fait l'objet d'une remise au goût du jour destinée à en améliorer le prix de vente. Ainsi, l'état et la qualité moyens des biens comparables vendus sont généralement supérieurs à ceux de maisons plus anciennes absentes du marché depuis longtemps. Cette différence de valeur devrait être captée par l'estimation du résidu du bien à évaluer, autrement dit par son score CQA. C'est une considération importante en matière de qualité des données et d'exactitude dans l'évaluation du bien à évaluer.

## 11.6 L'importance du CVR2 face aux données inattendues

La régression MARS repose sur de multiples itérations de R/earth appliquées à un sous-ensemble de biens sélectionné aléatoirement, qui comprend typiquement 80% à 90% du jeu de données MLS et que l'on appelle l'ensemble d'entraînement. Les 10% - 20% de biens restants constituent l'ensemble de test. Dans cette phase, le modèle MARS est créé à partir des données d'entraînement, puis utilisé pour prédire les prix de vente des biens de l'ensemble de test, lesquels sont ensuite comparés aux prix de vente réels. Cette méthodologie est reconnue par les spécialistes de la fouille de données comme la stratégie la plus efficace pour développer un modèle robuste capable de prédire les prix de vente de biens non encore observés.

Cependant, l'utilité première de la métrique CVR2 ne réside pas dans le reflet qu'elle donne de l'exactitude factuelle du modèle de régression au regard des ventes réelles de la zone de marché, mais plutôt dans l'optimisation des capacités prédictives du modèle concernant des ventes hypothétiques de biens à une date d'effet donnée. Ces biens ne figurent pas dans les données d'entraînement et peuvent présenter des combinaisons de caractéristiques nouvelles ou inattendues, absentes à la fois des données d'entraînement et des données de test.

Dans ce contexte, comment estimer le prix de vente d'un bien qui n'a pas été rénové au cours des 15 dernières années, en supposant qu'il ait été mis en vente, par exemple, un mois avant la date d'effet de l'évaluation ? Un modèle présentant un CVR2 solide est plus susceptible de fournir l'estimation la plus exacte qu'un modèle n'ayant pas subi cette validation croisée répétée. Il est également essentiel que le jeu de données comprenne un nombre suffisant de biens comparables présentant divers degrés de rénovation, de réparation et de modernisation.

En définitive, si nous disposons d'un modèle robuste qui prédit avec exactitude les contributions à la valeur des diverses caractéristiques des biens de notre jeu de données de biens comparables, nous pouvons placer efficacement le nouveau bien à évaluer dans le classement par résidu que le modèle établit pour ces biens. Ce placement nous permet d'attribuer au bien un score CQA et un résidu appropriés.

## 12 Résidus

Comme l'indique le nom « Residual Constraint Approach », les résidus occupent une place centrale dans cette méthodologie. Le terme résidu désigne ici la différence entre le prix de vente net d'un bien et le prix de vente prédit par l'analyse de régression.

Un résidu positif suggère que le bien s'est vendu plus cher que son prix estimé sur la base des attributs mesurables usuels, tels que la surface habitable brute (GLA) et la superficie du terrain, ce qui laisse penser qu'il présente un attrait supérieur à la moyenne. Un résidu négatif indique que le bien s'est vendu moins cher, ce qui peut suggérer un attrait moindre ou d'autres attributs négatifs non captés.

Étant donné que les maisons plus grandes présentent vraisemblablement des résidus plus importants, on peut utiliser le résidu par pied carré, calculé à partir de la GLA, afin d'obtenir une comparaison plus normalisée. Une discussion plus approfondie de cette question est réservée à un autre article.

### 12.1 Ranking & Scoring (classement et notation)

Après la création du modèle MARS (Multivariate Adaptive Regression Splines) pour les variables mesurées, les résidus sont calculés pour chaque bien. Ces résidus représentent la différence entre les prix de vente nets observés et les prix de vente prédits par le modèle. Une colonne supplémentaire doit être créée pour le Residual/SF, c'est-à-dire le résidu divisé par la surface habitable de la résidence, étant entendu qu'en général, le résidu d'un bien donné croît avec la taille de la surface habitable. D'autres statistiques peuvent donner de meilleurs résultats dans certains quartiers, mais elles feront l'objet d'un autre travail.

#### Hypothèse 7 : « Classement par résidu »

Dans cet article, les références au « classement par résidu » doivent s'entendre comme un classement par résidu, par résidu/SF, ou par toute autre fonction des résidus que le Valuation Engineer juge la plus appropriée pour la zone de marché. L'expérience de l'auteur dans la SF Bay Area indique que le résidu/SF est généralement le plus fiable, en particulier lorsque le bien à évaluer présente une GLA (Gross Living Area) inhabituellement petite ou grande.

Ensuite, les biens sont triés selon leurs résidus, du plus négatif au plus positif. Un score de Condition-Quality-Appeal (CQA — état, qualité, attrait) est alors généré pour chaque bien, sur une échelle de 0.0 à 10.0. Ce score est calculé comme suit :

Le score CQA correspond au rang centile du résidu d'un bien parmi l'ensemble des biens. Plus précisément, le score est obtenu en déterminant le pourcentage de biens dont les résidus sont inférieurs à celui du bien considéré, puis en divisant ce pourcentage par 10. Par exemple, si le résidu d'un bien est supérieur aux résidus de 50% des biens, son score sera de 5.0.

En raison des arrondis, un bien peut recevoir un score de 10.0 : celui qui présente le résidu le plus élevé. Toutefois, un score parfait de 10.0 est théoriquement impossible car, en toute logique, aucun bien ne peut avoir un résidu supérieur à celui de 100% des biens, lui-même compris. En revanche, le score minimal de 0.0 est possible, puisqu'au moins un bien n'a aucun autre bien présentant un résidu inférieur au sien. Notons que plusieurs biens peuvent recevoir un score de 0.0 en raison des arrondis lorsque de nombreux biens sont comparés.

Le score CQA fournit donc une mesure intuitive du degré d'adéquation entre les attributs d'un bien et les attentes du modèle, les scores les plus élevés indiquant une meilleure concordance

avec les prédictions du modèle.

Il est essentiel de reconnaître que le score CQA ne peut être considéré comme un indicateur indirect de l'attrait d'un bien comparable que lorsque les conditions de la vente sont en adéquation avec sa valeur de marché. Il incombe au Valuation Engineer de déterminer si la valeur résiduelle incorpore des éléments de sur- ou de sous-évaluation par rapport au marché, en fournissant une explication complète et une investigation approfondie de la transaction de vente. Bien que les ventes significativement au-dessus ou au-dessous du marché ne soient pas fréquentes, elles peuvent survenir dans des circonstances telles que les ventes forcées, les successions, l'incompétence, la collusion, etc.

Le tableau 3 présente un extrait semi-réaliste de données MLS, accompagné des résidus et des scores CQA, en valeur totale et par pied carré. Le nombre total de biens analysés dans cette étude s'élevait à environ 600. Il convient de noter que plusieurs variables ayant fait l'objet de la régression ne sont pas affichées. Autrement dit, les résidus présentés s'entendent après prise en compte, par la régression MARS, de l'impact d'autres variables, telles que la date de vente.

TABLE 3 – Biens comparables triés par résidu

| 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 Courbe CQA-résidu

Le tracé du CQA en fonction du Residual/SF pour un quartier produit invariablement la courbe caractéristique CQA-résidu présentée à la [Figure 2](#). Ce graphique particulier représente des données de Burlingame, CA, vers 2022. Des courbes similaires sont généralement observées lorsque l'on trace les scores CQA en fonction des valeurs de résidu ; je préfère toutefois utiliser le Residual/SF pour sa pertinence à contextualiser le CQA par rapport à la Gross Living Area (GLA) du bien à évaluer. En multipliant le Residual/SF par la GLA, on peut convertir le CQA d'un bien en une valeur résiduelle correspondante.

Le Valuation Engineer doit évaluer avec soin si la taille du bien, en particulier la Gross Living Area (GLA), influence de manière significative l'ampleur des résidus. Cette appréciation est critique, car elle peut déterminer la méthode la plus exacte pour calculer les valeurs des biens. Par exemple, des équipements de cuisine haut de gamme et des installations de salle de bains luxueuses peuvent accroître substantiellement la valeur résiduelle d'une maison. Cependant, dans les quartiers où la taille et l'agencement des cuisines et des salles de bains sont généralement uniformes, ces améliorations peuvent ne pas être fortement corrélées avec la GLA. Leur impact sur la valeur résiduelle peut tenir moins à l'espace qu'elles occupent qu'à la qualité qu'elles représentent.

À l'inverse, des éléments et matériaux extérieurs coûteux, qui améliorent notablement l'attrait extérieur d'un bien, peuvent entretenir une relation plus directe avec la taille de celui-ci. Les maisons plus grandes peuvent accueillir des extérieurs plus élaborés, tels que des aménagements paysagers étendus, des matériaux de toiture supérieurs ou de vastes patios, qui rehaussent naturellement l'attrait global du bien et, par conséquent, sa valeur résiduelle.

Il est donc essentiel que les Valuation Engineers prennent en compte ces dynamiques lors de l'analyse des résidus. Ils doivent non seulement apprécier la valeur intrinsèque ajoutée par des installations et des caractéristiques de haute qualité, mais aussi comprendre comment ces améliorations interagissent avec les dimensions physiques du bien. Cette approche globale garantit que les évaluations reflètent avec exactitude les attributs tant qualitatifs que quantitatifs d'un bien.

Par conséquent, il est recommandé de tracer à la fois les courbes CQA-résidu et CQA-résidu/SF et d'en étudier les différences, lesquelles peuvent révéler des tendances au sein d'une zone de marché. Le Valuation Engineer peut faire calculer les deux résidus par son programme et en prendre une moyenne pondérée afin de déterminer le résidu à attribuer au bien à évaluer. Cela peut faire une différence notable, ou non.

## 12.3 Caractéristiques de la courbe CQA

La courbe Condition-Quality-Appeal (CQA) offre un modèle sophistiqué pour comprendre comment les valeurs des biens fluctuent en fonction de l'attrait, de l'état et de la dynamique du marché au sein d'une zone donnée. Ce modèle capte l'interaction des facteurs économiques et physiques qui influencent le marché immobilier, reflétant les réalités nuancées de l'évaluation immobilière à travers différents segments de marché.

### 12.3.1 Analyse des composantes de la courbe CQA

1. Décroissance exponentielle sur la partie gauche de la courbe :

À mesure que le score CQA diminue d'environ 2 vers 0, les biens présentent des caractéristiques de détérioration rapide. Cette partie de la courbe est bien représentée par une

fonction de décroissance exponentielle, ce qui souligne la rapidité avec laquelle un bien peut perdre son attrait et sa valeur en l'absence d'entretien et de remises à niveau appropriés. La dégradation physique, telle que le bois qui pourrit, les couleurs qui passent et le métal qui rouille, accélère ce déclin, en particulier dans les maisons qui ne sont pas entretenues régulièrement.

2. Croissance linéaire dans la plage médiane :

L'essentiel du marché s'inscrit typiquement dans cette catégorie, où l'attrait augmente de façon quasi linéaire entre des scores d'environ 2 et 8. Cette section reflète des biens généralement entretenus selon les standards, mais qui ne présentent pas nécessairement de qualités ou d'emplacements exceptionnels. Elle représente la vaste classe moyenne du logement, englobant la majorité du marché immobilier, où les maisons sont proposées à des prix compétitifs sur la base de leurs caractéristiques standards et de leur état général.

3. Croissance exponentielle sur la partie droite de la courbe :

Ici, la courbe s'élève de façon exponentielle entre les scores 8 et 10, reflétant le segment du marché au plus fort attrait. Les biens de ce segment appartiennent souvent à des personnes plus fortunées, qui investissent non seulement dans un entretien supérieur, mais aussi dans des améliorations qui rehaussent considérablement l'attrait et la valeur du bien. La croissance exponentielle reflète la rareté de ces biens de grande qualité et la forte demande émanant d'acheteurs aisés. Et il faut bien comprendre qu'une maison de prix modeste peut obtenir une note très élevée en termes d'attrait, tandis qu'une maison grande et coûteuse peut obtenir une note faible. Ces dynamiques sont mises au jour par un usage expert de MARS et passent généralement inaperçues auprès de la plupart des évaluateurs.

### 12.3.2 Dynamique du marché et implications sociales

La courbe CQA raconte implicitement l'histoire de la richesse relative et des disparités économiques au sein de différentes régions. Dans des endroits comme le comté de San Mateo, la variation des valeurs immobilières entre des secteurs comme Burlingame et des quartiers plus fortunés tels que Hillsborough, Woodside ou Atherton illustre des clivages socio-économiques importants. Cette disparité affecte tout, de l'entretien des biens aux évaluations de marché, et influence en dernier ressort la démographie des acheteurs et les tendances du marché immobilier.

En outre, l'anecdote de la personne relativement fortunée qui investit dans une maison de grande qualité au sein d'un quartier de gamme intermédiaire où elle se sent à l'aise et où, pour diverses raisons, elle préfère vivre, met en lumière un autre aspect critique de l'économie immobilière — le concept de biens « surconstruits » (*overbuilt*). De tels biens, bien que très attrayants, peuvent ne pas trouver d'acheteurs locaux capables de payer le prix qu'ils auraient pu atteindre dans une zone de marché plus chère.

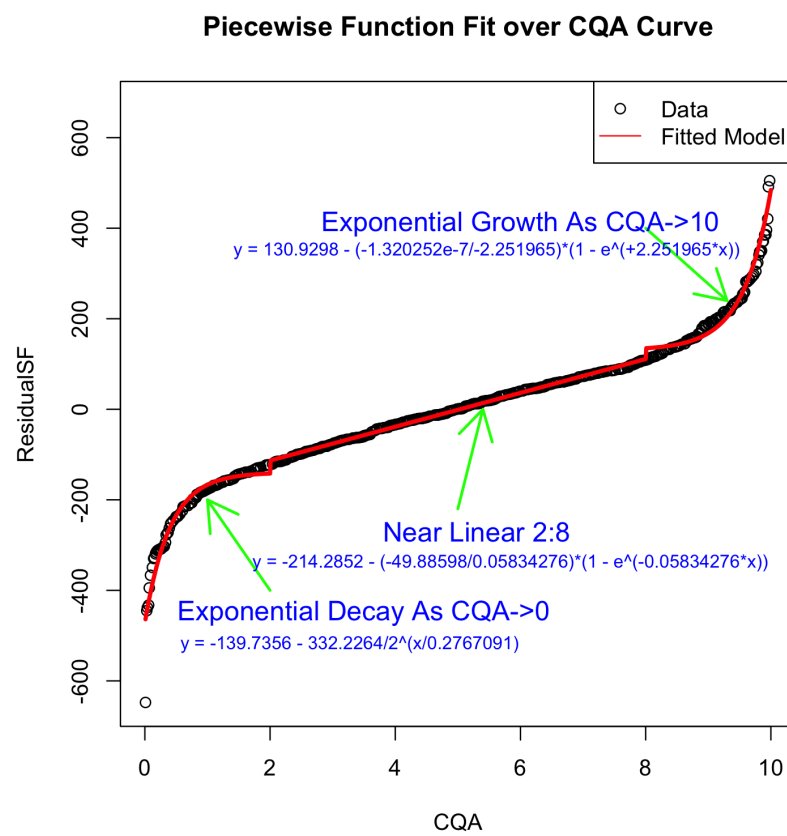
La courbe CQA-résidu est caractéristique de chaque zone de marché et ne doit pas être négligée lors de l'analyse.

### 12.4 Résidu du bien à évaluer

Contrairement aux biens comparables, le bien à évaluer ne possède pas de résidu déterminé de manière objective ; celui-ci doit être estimé par le Valuation Engineer (VE). Cette estimation introduit un degré de subjectivité dans le protocole du Residual Constraint Approach (RCA), marquant un point critique où le jugement personnel joue un rôle déterminant.

Pour estimer le score de Condition-Quality-Appeal (CQA), le VE analyse méticuleusement les biens classés, à la recherche de motifs discernables au sein du classement. Cette analyse implique

FIGURE 2 – Fonction caractéristique CQA-résidu



souvent un examen détaillé des photos MLS associées à chaque bien. Un examen exhaustif est impraticable pour un vaste jeu de données, tel que 600 biens. Le VE applique plutôt un raisonnement pratique pour approcher la position probable du bien à évaluer dans le classement global. Les biens nécessitant des réparations importantes, ou « fixers », se situent généralement près du bas du classement, tandis que ceux qui présentent des rénovations et un attrait supérieurs tendent à se classer près du sommet.

Le VE adopte ensuite une approche plus ciblée, en « sautant » au sein du classement pour identifier les biens comparables dont l'attrait correspond étroitement à celui du bien à évaluer. À mesure que la plage se resserre, le VE prête une attention croissante aux similitudes entre les biens, en particulier sur des aspects tels que la qualité, l'état, l'esthétique de la conception (y compris des éléments comme la peinture, les boiseries, la maçonnerie de pierre et les fenêtres), l'utilité fonctionnelle et d'autres caractéristiques pertinentes.

Il est crucial de comprendre que les jugements du VE sont fondamentalement guidés par les classements établis par le marché. Cela garantit que les éléments subjectifs du processus restent ancrés dans les réalités du marché, en visant une évaluation juste et exacte du bien à évaluer.

## **12.5 Conclusion de valeur RCA**

### **12.5.1 Considérations générales**

La détermination d'un score CQA est essentielle pour établir la valeur résiduelle estimée d'un bien, ces éléments étant intrinsèquement liés. S'il est possible d'évaluer un bien sans CQA, cette métrique fournit un contexte précieux en illustrant la position du bien par rapport aux autres dans la zone de marché. Par exemple, un score CQA de 3.2 indique que le bien présente un attrait supérieur à celui de 32% des biens comparables du voisinage. Si le score attribué fait l'objet d'objections, les contestataires doivent consulter les classements et justifier tout ajustement proposé sur la base de preuves photographiques ou d'autres données substantielles. Si le système de classement est robuste et que le Valuation Engineer (VE) a positionné le bien avec exactitude dans ce cadre, il est peu probable que des divergences significatives apparaissent.

Lors de la phase de rapport, on sélectionne typiquement 6 à 12 ventes comparables récentes et pertinentes pour étayer la conclusion de valeur. Les différences entre ces biens comparables et le bien à évaluer sont analysées, chaque variable faisant l'objet d'un ajustement correspondant. Ces ajustements sont ensuite appliqués aux prix de vente nets des comparables afin d'obtenir des prix de vente ajustés. Comme démontré mathématiquement, ces prix ajustés doivent converger vers le prix de vente estimé du bien à évaluer.

### **12.5.2 Révision, audit et réclamations**

Une fois le résidu estimé du bien à évaluer déterminé, des ajustements résiduels peuvent être calculés pour tous les biens comparables. Ces ajustements servent de contraintes pour les décompositions manuelles supplémentaires effectuées par le VE, adaptées aux variables CQA spécifiques jugées pertinentes. Il importe de souligner que ces décompositions détaillées sont explicatives et n'influencent pas l'évaluation finale, comme l'affirme la Démonstration II.

Concernant les réclamations soumises par des utilisateurs insatisfaits de la conclusion de valeur de l'évaluation, un évaluateur traditionnel doit typiquement perdre du temps à expliquer pourquoi tel ou tel bien, dont le prix est plus élevé ou plus bas, n'a pas l'impact sur la valeur attendu par l'utilisateur. Avec le RCA, toutes les ventes comparables possibles (habituellement 100-600) sont typiquement évaluées précisément à la même valeur que la conclusion de prix, de sorte qu'il faut peu d'efforts pour présenter les ajustements relatifs à un bien arbitraire. L'ensemble du processus

RCA est automatisé et objectif, à l'exception de la notation du résidu du bien à évaluer. Ainsi, un autre avantage du RCA est que la critique des imperfections du jugement subjectif du VE se limite pour l'essentiel à ce seul point — et elle peut être gérée de manière efficace et efficiente. De plus, en supposant un  $R^2$  d'environ 80%, l'impact d'une erreur de  $\pm 10\%$  dans l'estimation du résidu du bien à évaluer se traduit par une erreur de  $\pm 2\%$  dans l'estimation finale de la valeur.

Dans de nombreux cas, le bien à évaluer s'insère si parfaitement dans le classement des biens par résidu que l'on peut s'attendre à une exactitude de l'ordre de  $\pm 1\%$ .

La somme du résidu estimé et du prix de vente du bien à évaluer estimé par MARS donne la conclusion de valeur finale, appelée conclusion de valeur du Residual Constraint Approach (RCA). Ce processus garantit une approche complète et transparente de l'évaluation immobilière, renforçant la confiance dans l'exactitude et l'équité de l'appréciation.

## 13 Conclusion

La Residual Constraint Approach est un cadre porteur de trois contributions substantielles : une propriété structurelle d'anti-ancrage découlant de l'ajustement du modèle sur l'ensemble du marché plutôt que sur un petit ensemble de biens comparables (Démonstration I) ; un résultat d'invariance qui permet à la décomposition du résidu de servir de couche explicative sans perturber la conclusion de valeur (Démonstration II et la remarque qui la suit) ; et une régularité empirique — la courbe caractéristique CQA-Résidu — qui semble se reproduire d'une zone de marché à l'autre sous une forme caractéristique exponentielle-linéaire-exponentielle. Les limites du cadre sont tout aussi substantielles : la décomposition du résidu est invariante mais non identifiante, l'étape de positionnement CQA repose sur le jugement du VE, lequel n'a pas encore été validé par une étude formelle de concordance inter-évaluateurs, et les affirmations relatives à l'exactitude demeurent en attente d'une évaluation comparative face aux approches alternatives.

Ce que le présent article propose, c'est un cadre et un protocole. Ce qu'il ne propose pas encore, mais que le programme de recherche plus vaste actuellement en cours propose, c'est une chaîne de traitement exécutable et traçable à des fins d'audit, qui opérationnalise le protocole dans un logiciel doté de paramètres par défaut défendables. Cette chaîne de traitement — le package earthUI sur CRAN pour l'étape des fonctions de base MARS, le package glmnetUI pour le modèle structurel régularisé sur ces bases, et le package mgcvUI pour le lissage spatial — fait l'objet d'articles compagnons dans le présent numéro et constitue le pont entre la RCA en tant que méthode et la RCA en tant que livrable. L'écart entre doctrine et pratique dans la réglementation de l'évaluation immobilière (CRAYTOR, 2026, ce numéro) fournit la raison pour laquelle ce pont importe : les résultats d'évaluation sont économiquement consommés comme des intrants de risque prospectifs, et un cadre qui rend l'opinion de valeur actuelle auditable et reproductible est une condition préalable à tout engagement honnête avec cet usage prospectif. La RCA n'est qu'une étape de ce projet plus vaste. Elle n'en constitue pas la totalité, et le présent article ne prétend pas le contraire.

## Références

- AMBROSE, Brent W. et al. (2025). « Do Appraiser and Borrower Race Affect Mortgage Collateral Valuation? » In : *Review of Finance*. DOI : [10.1093/rof/rfaf046](https://doi.org/10.1093/rof/rfaf046).
- CRAYTOR, William Bert (2026). « The Doctrine–Practice Gap in Real Estate Appraisal: A Structured Account of Functions, Boundaries, and Tensions ». In : *Valuation Engineer Journal* 1.1.
- FRIEDMAN, Jerome H. (1991). « Multivariate Adaptive Regression Splines ». In : *The Annals of Statistics* 19.1, p. 1-141.
- HASTIE, Trevor, Robert TIBSHIRANI et 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 et Michael LACOUR-LITTLE (2020). « AVM versus Appraisal-Based Underwriting in Refinance Mortgages: The Trade-off Between Noise and Bias ». In : *SSRN Electronic Journal*. DOI : [10.2139/ssrn.3603327](https://doi.org/10.2139/ssrn.3603327).
- KUHN, Max et 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>.

- MILBORROW, Stephen (2024b). *Notes on the Earth Package*. Rapp. tech. URL : <http://www.milbo.org/doc/earth-notes.pdf>.
- (2024c). *Variance Models in Earth*. Rapp. tech. URL : <http://www.milbo.org/doc/earth-varmod.pdf>.
- PERRY, Andre M., Jonathan ROTHWELL et David HARSHBARGER (2018). *The Devaluation of Assets in Black Neighborhoods*. Rapp. tech. 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*. Rapp. tech. Brookings Institution.
- PROPERTY APPRAISAL AND VALUATION EQUITY TASK FORCE (2022). *Action Plan to Advance Property Appraisal and Valuation Equity*. Rapp. tech. U.S. Department of Housing et Urban Development.
- TZIOUMIS, Konstantinos (2016). « Appraisers and Valuation Bias: An Empirical Analysis ». In : *Real Estate Economics* 45.3, p. 679-712. DOI : [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- WILLIAMSON, Jake et Mark PALIM (2022). *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Rapp. tech. 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 : une interface interactive pour le package earth (MARS)

*Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

William Bert Craytor

Volume 1, Numéro 1 · Été 2026

Publié le 1<sup>er</sup> juillet 2026

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

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

Cet article en libre accès est diffusé selon les termes de la [licence Creative Commons Attribution 4.0](#).

**Résumé.** earthUI est une interface graphique pour le package R `earth`, qui implémente les Multivariate Adaptive Regression Splines (MARS). Elle propose trois modes d'utilisation — modélisation prédictive générale, évaluation immobilière et analyse de zone de marché — et guide l'utilisateur à travers l'importation des données, la configuration du modèle, l'ajustement, ainsi que l'interprétation des diagnostics et des sorties graphiques d'`earth`. Cet article documente les exigences de format de données d'`earthUI`, le flux de travail de modélisation, les affichages de sortie et la référence complète des fonctionnalités.

**Mots-clés :** MARS, multivariate adaptive regression splines, `earth`, R, interface graphique, évaluation immobilière

**Note de l'éditeur (juillet 2026) :** Cet article décrit earthUI 0.10.0. Peu avant la mise sous presse, earthUI 0.11.0 a absorbé les applications compagnes `glmnetUI` et `mgcvUI` (décrites dans les deux articles suivants), regroupant les flux de travail MARS, elastic-net et GAM dans une seule application. L'interface et le flux de travail décrits ici demeurent inchangés par ce regroupement.

## Table des matières

|          |                                                     |           |
|----------|-----------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                 | <b>4</b>  |
| 1.1      | Qu'est-ce qu'earthUI?                               | 4         |
| 1.2      | Un package, trois routines                          | 4         |
| 1.3      | Trois modes Purpose                                 | 4         |
| 1.4      | Fonctionnalités spécifiques à l'immobilier          | 5         |
| 1.5      | Installation d'earthUI et de vProlog                | 5         |
| 1.6      | Premiers pas                                        | 6         |
| <b>2</b> | <b>Exigences relatives aux données d'entrée MLS</b> | <b>7</b>  |
| 2.1      | Format et structure du fichier                      | 7         |
| 2.2      | Colonnes requises pour le mode Appraisal            | 7         |
| 2.3      | Conventions de nommage des colonnes spéciales       | 9         |
| 2.4      | Qualité et complétude des données                   | 9         |
| 2.5      | Placement de la ligne du bien à évaluer             | 9         |
| <b>3</b> | <b>Mode General Purpose</b>                         | <b>11</b> |
| 3.1      | Vue d'ensemble                                      | 11        |
| 3.2      | Le flux de travail de la barre latérale             | 11        |
| 3.3      | Les onglets du panneau principal                    | 12        |

|          |                                                                     |           |
|----------|---------------------------------------------------------------------|-----------|
| 3.4      | Persistence des paramètres                                          | 12        |
| 3.5      | Mode sombre                                                         | 12        |
| 3.6      | Locale et paramètres régionaux                                      | 13        |
| 3.6.1    | Pays pris en charge                                                 | 13        |
| 3.6.2    | Remplacement du format de papier                                    | 13        |
| 3.6.3    | Enregistrement des valeurs par défaut                               | 14        |
| <b>4</b> | <b>Mode d'évaluation (Appraisal Mode)</b>                           | <b>15</b> |
| 4.1      | Traitement de la ligne du bien à évaluer                            | 15        |
| 4.2      | Effective Date et ancienneté des ventes (Sale Age)                  | 15        |
| 4.3      | Désignations de colonnes spéciales (Special)                        | 15        |
| 4.4      | Residual Contribution (estimation des caractéristiques rares)       | 16        |
| 4.5      | Aperçu des ajustements RCA                                          | 17        |
| <b>5</b> | <b>Mode d'analyse de zone de marché (Market Area Analysis Mode)</b> | <b>18</b> |
| 5.1      | Différences par rapport au mode d'évaluation                        | 18        |
| 5.2      | Quand utiliser le mode marché                                       | 18        |
| <b>6</b> | <b>Sélection des variables</b>                                      | <b>19</b> |
| 6.1      | Variable(s) cible(s)                                                | 19        |
| 6.2      | La table des prédicteurs                                            | 19        |
| 6.3      | Détection des types de données et redéfinitions                     | 20        |
| 6.4      | Indicateurs Factor et Linear                                        | 20        |
| 6.5      | Référence des types de colonnes spéciaux                            | 20        |
| 6.6      | Cibles multiples                                                    | 21        |
| <b>7</b> | <b>Sélection des paramètres</b>                                     | <b>22</b> |
| 7.1      | Paramètres de base                                                  | 22        |
| 7.2      | Passe avant                                                         | 22        |
| 7.3      | Interactions autorisées                                             | 23        |
| 7.4      | Élagage                                                             | 23        |
| 7.5      | Validation croisée                                                  | 23        |
| 7.6      | Filtrage par sous-ensemble                                          | 24        |
| 7.7      | Valeurs de paramètres recommandées                                  | 24        |
| 7.7.1    | Paramètres de la passe avant                                        | 24        |
| 7.7.2    | Paramètres d'élagage                                                | 26        |
| 7.7.3    | Paramètres de validation croisée                                    | 26        |
| 7.7.4    | Modèle de variance                                                  | 27        |
| 7.7.5    | Fichier de sortie earth                                             | 27        |
| 7.8      | Réglages par défaut                                                 | 27        |
| <b>8</b> | <b>Ajustement du modèle Earth</b>                                   | <b>28</b> |
| 8.1      | Le bouton Fit                                                       | 28        |
| 8.2      | Fenêtre modale d'ajustement et sortie de trace                      | 28        |
| 8.3      | Journal d'ajustement                                                | 28        |
| <b>9</b> | <b>Téléchargement des données</b>                                   | <b>29</b> |
| 9.1      | Colonnes de sortie                                                  | 29        |
| 9.2      | Ordre des colonnes et mise en forme Excel                           | 29        |
| 9.3      | Scores CQA                                                          | 30        |
| 9.4      | Tri et ligne du bien à évaluer                                      | 30        |
| 9.5      | Export automatique vers mgcvUI                                      | 30        |

|           |                                                                                             |           |
|-----------|---------------------------------------------------------------------------------------------|-----------|
| 9.6       | Import dans glmnetUI . . . . .                                                              | 31        |
| 9.6.1     | Principaux avantages . . . . .                                                              | 31        |
| 9.6.2     | Important : la colonne de pondération doit correspondre . . . . .                           | 31        |
| 9.6.3     | Flux de travail . . . . .                                                                   | 31        |
| <b>10</b> | <b>Calculs RCA et téléchargement</b>                                                        | <b>32</b> |
| 10.1      | Ouverture de la boîte de dialogue RCA . . . . .                                             | 32        |
| 10.2      | Interpolation du score CQA . . . . .                                                        | 32        |
| 10.3      | Colonnes de sortie . . . . .                                                                | 32        |
| 10.4      | L'onglet RCA Adjustments . . . . .                                                          | 33        |
| <b>11</b> | <b>Sales Comparison Grid</b>                                                                | <b>34</b> |
| 11.1      | Vue d'ensemble et objet . . . . .                                                           | 34        |
| 11.2      | Fenêtre modale de sélection des comparables . . . . .                                       | 34        |
| 11.3      | Disposition de la grille . . . . .                                                          | 34        |
| 11.4      | Prix de vente, concessions et Net SP . . . . .                                              | 35        |
| 11.5      | Lignes groupées (localisation, site, âge) . . . . .                                         | 35        |
| 11.6      | Lignes de variables du modèle . . . . .                                                     | 35        |
| 11.7      | Ligne CQA Residual et lignes de caractéristiques résiduelles . . . . .                      | 36        |
| 11.8      | Formule du prix de vente ajusté (Adjusted Sale Price) . . . . .                             | 36        |
| 11.9      | Protection des feuilles et cellules modifiables . . . . .                                   | 36        |
| 11.10     | Utilisation de la grille dans Excel . . . . .                                               | 36        |
| <b>12</b> | <b>Téléchargement des rapports</b>                                                          | <b>38</b> |
| 12.1      | Generate Quarto Report . . . . .                                                            | 38        |
| 12.2      | Convert Quarto Report . . . . .                                                             | 38        |
| 12.3      | Contenu du rapport . . . . .                                                                | 38        |
| <b>13</b> | <b>Les Post Processing Modules : Elastic Net et GAM</b>                                     | <b>40</b> |
| 13.1      | Pourquoi des ajustements d'appui ? . . . . .                                                | 40        |
| 13.2      | Lancement et état partagé . . . . .                                                         | 40        |
| 13.3      | La routine elastic net . . . . .                                                            | 40        |
| 13.4      | La routine GAM . . . . .                                                                    | 40        |
| 13.5      | Un flux de travail pratique . . . . .                                                       | 41        |
| <b>14</b> | <b>Jeu de données de démonstration : Appraisal_1.csv</b>                                    | <b>42</b> |
| 14.1      | Description . . . . .                                                                       | 42        |
| 14.2      | Colonnes . . . . .                                                                          | 43        |
| 14.3      | Démarrage rapide suggéré . . . . .                                                          | 43        |
| <b>15</b> | <b>Fonctionnalité expérimentale : extraction de texte en Prolog et règles de dérivation</b> | <b>45</b> |
| 15.1      | Ce qu'elle fait . . . . .                                                                   | 45        |
| 15.2      | Prérequis : vProlog (Trealla), et non SWI-Prolog . . . . .                                  | 46        |
| 15.3      | Activation et exécution . . . . .                                                           | 46        |
| 15.4      | Règles de dérivation . . . . .                                                              | 46        |
| 15.5      | Fichier de données actif versionné . . . . .                                                | 47        |
| 15.6      | Interface programmatique et fichiers inclus . . . . .                                       | 47        |
| 15.7      | Expérimentation connexe : a priori par région climatique . . . . .                          | 47        |
| 15.7.1    | Au-delà de la Californie : packs juridiques et modules de zone de marché . . . . .          | 47        |

## 1 Introduction

### 1.1 Qu'est-ce qu'earthUI ?

earthUI est une interface graphique pour le package R **earth**, qui implémente les Multivariate Adaptive Regression Splines (MARS) (FRIEDMAN 1991). Le moteur de modélisation est le package **earth** de Stephen Milborrow (MILBORROW 2024), documenté en détail par MILBORROW (s. d.[a]) et MILBORROW (s. d.[b]). Il s'exécute sous la forme d'une application Shiny locale — sans identifiant, sans serveur et sans compte. On la lance depuis R, on importe un jeu de données (CSV ou Excel), on configure le modèle et on l'ajuste de manière interactive.

L'application offre un flux de travail complet : importation des données, configuration des variables, ajustement du modèle, graphiques de diagnostic, importance des variables, équations du modèle et rapports téléchargeables aux formats HTML, Word ou PDF.

### 1.2 Un package, trois routines

Depuis la version actuelle, le package earthUI *contient* également ses deux Post Processing Modules (modules de post-traitement), qui constituaient dans les versions antérieures les packages séparés glmnetUI et mgcvUI. L'installation d'earthUI installe les trois routines ; chacune s'exécute comme une application Shiny locale distincte :

- `earthUI::launch()` — l'application earth (MARS) décrite dans le présent article, sur le port 7878. `earth()` est le moteur central d'évaluation.
- `earthUI::launch_glmnet()` — le Post Processing Module elastic net (port 7879), qui importe les fonctions de base du modèle earth et les réestime sous régularisation lasso/elastic-net (voir « glmnetUI Import » ci-dessous).
- `earthUI::launch_mgcv()` — le Post Processing Module GAM (port 7880), qui utilise les emplacements de nœuds du modèle earth comme points de départ pour les termes de lissage de `mgcv` (voir « mgcvUI Auto-Export » ci-dessous).

Les trois routines partagent une même arborescence de projets, une même base de données de paramètres et une même copie de l'infrastructure de support, de sorte qu'un projet ouvert dans l'une des applications est immédiatement disponible dans les autres. Les routines glmnet et mgcv constituent des éléments de corroboration de la conclusion de valeur du modèle earth, et non des flux de travail concurrents : le schéma prévu consiste à ajuster et à affiner d'abord le modèle earth, puis à exécuter les modules sur le même projet lorsqu'une corroboration est souhaitée.

### 1.3 Trois modes Purpose

earthUI organise le travail en **projets** — des entités nommées qui regroupent les fichiers d'entrée, les sorties et les paramètres en un emplacement fixe. Chaque projet possède un **Purpose** (finalité), choisi dans l'assistant New Project lors de la création du projet. Il existe trois modes :

- **General** (général) — régression Earth pour tout type de population ou de jeu de données. Il s'agit du mode par défaut. Il offre le flux de travail complet de modélisation MARS, sans aucun ajout propre à un domaine particulier.
- **For Appraisal** (pour l'évaluation immobilière) — régression Earth adaptée à l'évaluation immobilière. Ce mode ajoute des fonctionnalités propres à l'évaluation d'un bien individuel, notamment la prise en charge du bien à évaluer, les désignations spéciales de colonnes, la Residual Constraint Approach (RCA) (fr. : approche par contraintes résiduelles) et la génération de la Sales Comparison Grid (grille de comparaison).

- **Market Area Analysis** (analyse de zone de marché) — régression Earth adaptée aux études de zones de marché. Ce mode ajoute des fonctionnalités destinées à l'analyse de groupes de biens au sein d'un marché défini, plutôt qu'à l'évaluation d'un bien unique.

Dans les trois modes, le moteur de modélisation central est identique — on ajuste toujours un modèle Earth (MARS). Le réglage du Purpose détermine quels outils et éléments d'interface supplémentaires sont disponibles.

**Le Purpose est fixé par projet.** La finalité d'un projet — de même que son pays, son État, son comté et sa ville — est verrouillée à la création du projet. Le passage à un autre projet réinitialise earthUI à un état par défaut vierge (les données importées, les résultats du modèle, les onglets, la configuration des variables et les paramètres earth sont tous effacés), puis restaure la configuration enregistrée de ce projet pour sa finalité.

## 1.4 Fonctionnalités spécifiques à l'immobilier

Lorsque le mode **For Appraisal** ou **Market Area Analysis** est sélectionné, earthUI active plusieurs fonctionnalités conçues pour l'analyse immobilière :

- **Désignations spéciales de colonnes** — Chaque prédicteur peut se voir attribuer un rôle spécial tel que `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Ces désignations déterminent la manière dont la colonne est traitée lors de l'ajustement, de la production des sorties et de la génération de la Sales Grid. Voir le chapitre 6 pour la référence complète.
- **Arrondi de la latitude et de la longitude** — Les colonnes désignées comme latitude ou longitude sont automatiquement arrondies à 3 décimales afin de prévenir le surajustement.
- **Colonne Sale Age** — Lorsqu'une colonne est désignée comme `contract_date` et qu'une Effective Date (date d'effet) est fournie, earthUI calcule une colonne `sale_age` (nombre de jours entre la date de vente et la date d'effet) et la substitue comme prédicteur.
- **Calculs RCA (évaluation uniquement)** — En mode évaluation, après l'ajustement du modèle, earthUI peut calculer les sorties de la Residual Constraint Approach (RCA), qui produisent des ajustements par bien comparable, des synthèses d'ajustements nets/bruts et un prix de vente ajusté pour le bien à évaluer.
- **Sales Comparison Grid (évaluation uniquement)** — Génère un classeur Excel à plusieurs feuilles comportant des formules pour les ajustements des biens comparables, les décompositions du résidu par caractéristique et les prix de vente ajustés. Voir le chapitre 11.

***Conseil :** le mode General convient à tout jeu de données — financier, scientifique, technique ou immobilier. Les modes évaluation et zone de marché ne font qu'ajouter des fonctionnalités de commodité destinées aux professionnels de l'immobilier.*

## 1.5 Installation d'earthUI et de vProlog

earthUI est publié sur le CRAN ; l'installation standard tient donc en une seule ligne de R :

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

Le présent article décrit la version actuelle. Si le CRAN propose encore une version plus ancienne au moment de la lecture, la version de développement s'installe depuis GitHub (noter le `subdir` — le package réside dans le dossier `pkg/` du dépôt) :

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

**vProlog (facultatif).** Le dispositif expérimental de traitement des remarques (voir « Prolog Text Extraction » vers la fin du présent article) utilise le package vProlog — un moteur Prolog embarqué autonome. Il ne s'agit délibérément *pas* d'une dépendance obligatoire : earthUI s'installe et s'exécute sans lui, et les fonctionnalités Prolog demeurent simplement désactivées tant qu'il n'est pas présent. vProlog est distribué via le dépôt r-universe de l'auteur, qui fournit des binaires précompilés pour Windows et macOS (aucune chaîne de compilation n'est nécessaire) :

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

Aucune installation système de Prolog (SWI-Prolog ou autre) n'est requise — le moteur Trealla est compilé dans le package lui-même. Tout le reste du présent article fonctionne sans vProlog.

## 1.6 Premiers pas

Pour utiliser earthUI :

1. **Installer le package** depuis le CRAN comme décrit ci-dessus.
2. **Lancer l'application** : exécuter `earthUI::launch()` dans R, ou, depuis la ligne de commande, exécuter `Rscript -e 'earthUI::launch()'`. L'application s'ouvre dans le navigateur web sur le port 7878. On peut également y accéder directement en naviguant vers `http://localhost:7878` dans le navigateur. (Les Post Processing Modules se lancent de la même manière : `launch_glmnet()` sur le port 7879, `launch_mgcv()` sur le port 7880.)
3. **Créer ou ouvrir un projet** dans la Section 1 de la barre latérale. L'assistant New Project recueille le nom du projet, le Purpose (General, For Appraisal ou Market Area Analysis), la localisation (pays, État, comté, ville) et, facultativement, un fichier de données initial, qui est copié dans le dossier `in/` du projet.
4. **Importer les données** dans la Section 2, qui répertorie les fichiers CSV et Excel présents dans le dossier `in/` du projet actif. Les fichiers ajoutés manuellement via le Finder ou l'Explorer apparaissent après un clic sur Refresh.
5. **Configurer les variables** — choisir la ou les variables cibles ainsi que les prédicteurs, définir les types de données et attribuer les éventuels rôles spéciaux de colonnes.
6. **Définir les paramètres Earth** — `degree`, `nprune`, filtres de sous-ensembles et autres arguments de `earth()`.
7. **Ajuster le modèle** — cliquer sur « Fit Earth Model » et examiner les résultats dans le panneau principal.
8. **Exporter** — télécharger les prédictions au format Excel, générer des lots de rapports Quarto ou (en mode évaluation) calculer les ajustements RCA et les Sales Comparison Grids.

La configuration est enregistrée côté serveur dans une base de données de projet, indexée par projet et par Purpose, via les boutons « Save current as default » des Sections 3 et 4 (voir le chapitre 3, « Settings Persistence »).

## 2 Exigences relatives aux données d'entrée MLS

Pour les flux de travail d'évaluation immobilière et d'analyse de marché, les données d'entrée proviennent généralement d'un export d'un Multiple Listing Service (MLS). Ce chapitre décrit la structure de fichier attendue ainsi que les colonnes que earthUI peut exploiter.

### 2.1 Format et structure du fichier

earthUI accepte les fichiers **CSV** et **Excel** (.xlsx, .xls). Lors de l'import, les noms de colonnes sont automatiquement convertis en **snake\_case** — par exemple, « Living SqFt » devient **living\_sqft**, « Contract Date » devient **contract\_date** et « Sale Price » devient **sale\_price**. Cette normalisation garantit des références de colonnes cohérentes tout au long du flux de travail. Le séparateur CSV et le signe décimal utilisés lors de l'import sont déterminés par les paramètres de locale (voir le chapitre 3, « Locale & Regional Settings » [paramètres régionaux]).

Le fichier de données doit se présenter sous la forme d'une table plate comportant une ligne par bien et une colonne par attribut. La première ligne du fichier doit contenir les en-têtes de colonnes.

### 2.2 Colonnes requises pour le mode Appraisal

Bien que earthUI fonctionne avec n'importe quel ensemble de colonnes, le flux de travail complet d'évaluation (ajustements RCA + Sales Comparison Grid, la grille de comparaison) tire parti de la présence des colonnes suivantes dans l'export MLS :

| Colonne          | Type spécial                 | Rôle                                                                                               |
|------------------|------------------------------|----------------------------------------------------------------------------------------------------|
| Sale Price       | (target)                     | Variable réponse du modèle                                                                         |
| Contract Date    | <code>contract_date</code>   | Sert au calcul de <code>sale_age</code> (nombre de jours par rapport à la date d'effet)            |
| Listing Date     | <code>listing_date</code>    | Utilisée avec la date de contrat pour calculer le DOM en l'absence de colonne DOM                  |
| Days on Market   | <code>dom</code>             | Nombre de jours sur le marché ; affiché dans la Sales Grid                                         |
| Concessions      | <code>concessions</code>     | Concessions de vente ; $\text{Net SP} = \text{Sale Price} - \text{Concessions}$ dans la Sales Grid |
| Living Area (SF) | <code>living_area</code>     | Permet le calcul des résidus par pied carré ( <code>residual_sf</code> , <code>cqa_sf</code> )     |
| Lot Size         | <code>lot_size</code>        | Regroupée dans la ligne « Site Size   Dimensions » de la Sales Grid                                |
| Site Dimensions  | <code>site_dimensions</code> | Regroupée avec la taille du terrain dans la Sales Grid                                             |
| Latitude         | <code>latitude</code>        | Arrondie à 3 décimales ; sert au calcul de proximité et au groupe Location                         |
| Longitude        | <code>longitude</code>       | Arrondie à 3 décimales ; sert au calcul de proximité et au groupe Location                         |
| Area ID          | <code>area</code>            | Regroupée dans la ligne « Loc : Long   Lat   Area » de la Sales Grid                               |
| Actual Age       | <code>actual_age</code>      | Regroupée dans la ligne « Actual Age   Effective Age » de la Sales Grid                            |
| Effective Age    | <code>effective_age</code>   | Regroupée avec l'âge réel dans la Sales Grid                                                       |
| Parcel Number    | (aucun)                      | Affiché dans la ligne APN de la Sales Grid                                                         |
| Listing ID       | (aucun)                      | Affiché dans la ligne APN de la Sales Grid                                                         |
| Address          | <code>display_only</code>    | Affichée dans la ligne Address de la Sales Grid ; exclue du modèle                                 |

Les noms de colonnes du tableur peuvent être rédigés dans une langue étrangère — les noms « spéciaux » sont en anglais afin que le programme R puisse leur réserver un traitement particulier. Pour le reste, les noms de colonnes fournis apparaissent tels quels dans les modèles de régression, les graphiques et (dans le cas d'un rapport d'évaluation) l'Intermediate Sales Grid (grille de comparaison intermédiaire).

Toutes les colonnes ne sont pas requises. earthUI s'adapte — si une colonne est absente, la caractéristique correspondante est simplement omise. Par exemple, si aucune colonne **concessions** n'est désignée, la ligne Net SP de la Sales Grid affiche le prix de vente sans déduction des concessions. Toutefois, pour les modèles de prix immobiliers, certaines colonnes sont fortement recommandées afin d'obtenir un ajustement acceptable :

1. **Sale Age** (ancienneté de la vente) — le nombre de jours entre la date de vente du contrat et la date d'effet de l'évaluation ou de l'analyse. Lorsque l'historique des ventes porte sur plusieurs années, en particulier sur des périodes supérieures à 5 ans, `sale_age` joue souvent un rôle central dans l'estimation du prix de vente. De fait, cette variable est souvent si importante que, sans elle, earthUI ne parvient à produire aucun modèle.
2. **Living Area** (surface habitable) — désignée également par des noms tels que « Living Sqft », « GLA » (*gross living area*, surface habitable brute), etc. Il s'agit d'un autre déterminant majeur du prix de vente.
3. **Total Bath Count** (nombre total de salles de bain) — le nombre total de salles de bain complètes, quart, demi et trois-quarts. Par exemple, deux salles de bain complètes et une demi-salle de bain correspondent à une valeur de 2.5.

4. **Garage Bays** ou **Garage Area** (places ou surface de garage) — le nombre de places de garage ou la surface du garage en pieds carrés.
5. **Lot Size** (taille du terrain) — la superficie du terrain, généralement exprimée en pieds carrés ou en acres.
6. **Longitude**, **Latitude** et, si disponible, **Area ID** (identifiant de zone). Les ajustements correspondants seront combinés en un ajustement unique Location dans la Sales Grid.

## 2.3 Conventions de nommage des colonnes spéciales

earthUI identifie les colonnes par leur **désignation de type spécial**, et non par leur nom de colonne. Les colonnes peuvent porter n'importe quel nom dans l'export MLS — ce qui importe est de leur attribuer le type spécial correct dans la table Variable Configuration (configuration des variables) présentée au chapitre 6.

Par exemple, le MLS peut exporter la surface habitable sous le nom « GLA », « Living SqFt », « liv\_area » ou « gross\_living\_area ». Après l'import (où le nom est converti en snake\_case), il suffit de la désigner comme `living_area` dans la liste déroulante Special. earthUI l'utilisera alors pour le calcul des résidus par pied carré et pour le regroupement dans la Sales Grid, quel que soit son nom d'origine.

## 2.4 Qualité et complétude des données

- **Valeurs manquantes (NA)** : les lignes comportant des valeurs NA dans une colonne prédicteur ou dans la colonne cible sont automatiquement supprimées avant l'ajustement. Le paramètre `na.action` est toujours fixé en interne à `na.fail`, la suppression des NA étant effectuée avant l'appel à `earth()`.
- **Colonnes de dates** : les dates doivent être dans un format que R peut analyser (p. ex. 2025-06-15, 06/15/2025, June 15, 2025). earthUI détecte automatiquement les colonnes de dates lorsque au moins 80% des valeurs sont analysées avec succès. Les années à deux chiffres (p. ex. 06/15/25) sont lues comme l'année correspondante la plus proche, de sorte que les exports MLS récents sont analysés correctement ; les années à quatre chiffres sont toujours prises littéralement.
- **Colonnes numériques** : le prix de vente, la surface habitable, la taille du terrain, les concessions et les champs similaires doivent être numériques. Si le MLS exporte les prix avec des symboles monétaires ou des séparateurs de milliers (p. ex. « \$350,000 » ou « 350.000,00 »), il peut être nécessaire de nettoyer ces valeurs avant l'import.
- **Colonnes facteurs** : les variables catégorielles telles que l'identifiant de zone, le style ou l'état doivent contenir un nombre raisonnable de valeurs uniques. earthUI détecte automatiquement les variables facteurs, mais cette détection peut être outrepassée.

***Conseil** : après l'import, examinez la colonne T/F/NA du tableau des prédicteurs — elle indique, pour chaque colonne, le nombre de valeurs true/false/NA. Les colonnes présentant plus de 30% de valeurs manquantes sont signalées en rouge. Il convient d'envisager d'exclure les colonnes à fort taux de NA ou de nettoyer les données avant l'import.*

## 2.5 Placement de la ligne du bien à évaluer

En mode **Appraisal** (évaluation), la **ligne 1 doit correspondre au bien à évaluer**. Toutes les lignes restantes sont des ventes comparables. Le prix de vente du bien à évaluer peut être laissé vide (NA) ou fixé à une valeur quelconque — earthUI le traite dans tous les cas comme NA lors de l'ajustement.

En mode **Market Area Analysis** (analyse de zone de marché), placer un bien à évaluer en ligne 1 est facultatif, et aucune ligne n'est exclue automatiquement. Pour tenir la ligne du bien

à évaluer (ou toute autre ligne) hors de l'ajustement, il faut lui attribuer une valeur nulle dans une colonne de pondérations désignée ou utiliser un filtre de sous-ensemble.

En mode **General** (général), aucune ligne ne fait l'objet d'un traitement particulier — toutes les lignes sont traitées de manière identique.

## 3 Mode General Purpose

### 3.1 Vue d'ensemble

Le mode General Purpose (mode à usage général) est le mode actif par défaut au lancement d'earthUI. Il offre le flux de travail complet de modélisation MARS pour tout jeu de données — et pas seulement pour l'immobilier. Vous pouvez utiliser earthUI pour des données scientifiques, des analyses financières, des études d'ingénierie ou tout problème de régression dans lequel vous souhaitez explorer des relations non linéaires et des interactions entre variables.

En mode General, l'interface omet les fonctionnalités propres à l'immobilier (colonnes spéciales, ancienneté de la vente, arrondi des coordonnées, RCA). La barre latérale est épurée afin de se concentrer sur la sélection des variables, la configuration des paramètres, l'ajustement du modèle et l'export.

### 3.2 Le flux de travail de la barre latérale

La barre latérale est organisée en sections numérotées et repliables qui vous guident de la sélection du projet jusqu'à l'export :

**1. Project** (projet) — Sélectionnez un projet existant ou créez-en un nouveau via l'assistant New Project, qui enregistre l'objet, le pays et la localisation administrative (tous verrouillés une fois le projet créé). Le projet actif détermine l'emplacement d'où sont lus les fichiers d'entrée et celui où toutes les sorties sont écrites — il n'existe pas de champ distinct pour le dossier de sortie.

**2. Import Data** (importation des données) — Choisissez un fichier CSV ou Excel dans le dossier in/ du projet actif; un lien Refresh réanalyse le dossier à la recherche de fichiers ajoutés manuellement. Pour les fichiers Excel comportant plusieurs feuilles, un sélecteur de feuille apparaît. Les noms de colonnes sont automatiquement convertis en snake\_case.

**3. Variable Configuration** (configuration des variables) — Sélecteur de variable cible (plusieurs cibles possibles), Effective Date (date d'effet) et tableau des prédicteurs avec cases à cocher Include, Factor, Linear et Latent. Voir le chapitre 6 pour tous les détails.

**4. Earth Call Parameters** (paramètres d'appel de earth) — Tous les arguments de la fonction `earth()` : degree, penalty, nk, méthode d'élagage, validation croisée, etc. Voir le chapitre 7 pour la référence complète des paramètres.

**Execute Prolog Processing** — Section optionnelle et non numérotée, située entre les sections 4 et 5, désactivée tant que la fonctionnalité expérimentale Prolog n'est pas activée dans Settings. Voir le chapitre 15.

**5. Fit Earth Model** (ajustement du modèle earth) — Un champ de graine aléatoire et le bouton qui exécute le modèle de manière asynchrone. Voir le chapitre 8 pour les détails de l'ajustement.

**6. Download Estimated Target Variable(s) & Residuals** — (intitulée « Download Estimated Sale Prices & Residuals » dans les modes appraisal et market) Exporte les prédictions, les résidus, les scores CQA et les contributions par fonction g sous forme de fichier Excel. Voir le chapitre 9.

**7–8. RCA Adjustments et Sales Grid** — Mode appraisal uniquement. Voir les chapitres 10 et 11.

**9. Generate Quarto Report** (génération du rapport Quarto) — Écrit un ensemble de rapport Quarto autonome (source .qmd et ressources associées) dans le dossier de sortie du projet. Numérotée 7 hors mode appraisal. Voir le chapitre 12.

**10. Convert Quarto Report** (conversion du rapport Quarto) — Convertit tout fichier .qmd en HTML, Word ou PDF. Voir le chapitre 12.

### 3.3 Les onglets du panneau principal

Après l'ajustement, le panneau principal propose dix onglets :

| Onglet                | Contenu                                                                                                                                                                                                                                                                                |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data                  | Aperçu des données importées. En mode appraisal, scindé en deux tables : Subject Property et Comparable Sales.                                                                                                                                                                         |
| Equation              | L'équation du modèle ajusté, affichée en LaTeX/MathJax, montrant chaque fonction de base et son coefficient.                                                                                                                                                                           |
| Summary               | Métriques clés ( $R^2$ , GRSq, GCV, RSS, CV $R^2$ ) et tableau des coefficients.                                                                                                                                                                                                       |
| Variable Importance   | Diagramme en barres et tableau classé des scores d'importance des prédicteurs.                                                                                                                                                                                                         |
| Contribution          | Graphiques de contribution des fonctions g. Sélectionnez une fonction g dans le menu déroulant ; les prédicteurs univariés sont représentés par des courbes, les interactions bivariées ( $\text{degree} \geq 2$ ) par des graphiques en perspective 3D et des graphiques de contours. |
| Correlation           | Carte thermique des corrélations entre prédicteurs.                                                                                                                                                                                                                                    |
| Diagnostics           | Trois graphiques : Residuals vs Fitted, Normal Q-Q et Actual vs Predicted.                                                                                                                                                                                                             |
| RCA Adjustments       | (mode appraisal uniquement) Histogrammes des valeurs Residual Adj %, Net Adj % et Gross Adj % sur l'ensemble des comparables, renseignés une fois l'étape 7 (Calculate RCA Adjustments) terminée.                                                                                      |
| Residual Contribution | Estime les contributions de marché des caractéristiques rares ou exclues en régressant conjointement les résidus du modèle sur des colonnes jamais utilisées par le modèle (voir <a href="#">subsection 4.4</a> ).                                                                     |
| ANOVA                 | Tableau de décomposition ANOVA montrant la contribution de chaque fonction de base à la RSS.                                                                                                                                                                                           |
| Earth Output          | Sortie texte brute de 'earth : :summary()', y compris les détails de la passe d'élagage.                                                                                                                                                                                               |

### 3.4 Persistance des paramètres

earthUI enregistre votre configuration côté serveur dans une base de données de projets (projects.sqlite à la racine de regProj), indexée par projet et par objet. L'enregistrement est **déclenché par bouton** : le bouton « Save current as default » de la section 3 enregistre la cible, la date d'effet et la configuration des prédicteurs, tandis que le bouton de la section 4 enregistre les paramètres d'appel de earth et la matrice Allowed Interactions. Chaque bouton ne met à jour que sa propre moitié, de sorte qu'ils ne s'écrasent jamais mutuellement, et **l'ajustement du modèle n'enregistre rien automatiquement** — vous pouvez expérimenter librement sans écraser vos valeurs par défaut enregistrées. Les différents fichiers d'entrée d'un projet (un petit extrait de test, le jeu de données complet) partagent une même configuration et, comme les paramètres résident dans l'arborescence du projet, ils l'accompagnent lorsque vous synchronisez, sauvegardez ou partagez le dossier regProj.

### 3.5 Mode sombre

Cliquez sur l'icône lune/soleil dans le coin supérieur droit pour basculer entre les thèmes clair et sombre. La préférence de thème est enregistrée dans le stockage local et persiste d'une session à l'autre.

### 3.6 Locale et paramètres régionaux

earthUI prend en charge les conventions internationales de formatage des nombres, des dates et des fichiers CSV grâce à un système de locale fondé sur le pays. Le menu déroulant **Country** (pays) du menu **Settings** (engrenage) de la barre de menus supérieure sélectionne un préréglage parmi 31 pays pris en charge. Chaque préréglage configure :

- **Séparateur CSV** — virgule (,) pour les États-Unis, le Royaume-Uni et le Japon, ou point-virgule (;) pour la majeure partie de l'Europe, où la virgule sert de marque décimale.
- **Marque décimale** — point (.) ou virgule (,).
- **Séparateur de milliers** — virgule (États-Unis/Royaume-Uni/Japon), point (Allemagne/Italie/Espagne), espace (Finlande/France/Pologne/pays baltes/Ukraine/Russie) ou apostrophe (Suisse).
- **Format de date** — MM/DD/YYYY (États-Unis), DD/MM/YYYY (majeure partie de l'Europe) ou YYYY-MM-DD (Suède/Lituanie/Japon/Canada).
- **Format de papier** — Letter (États-Unis/Canada/Mexique) ou A4 (partout ailleurs).

#### 3.6.1 Pays pris en charge

| Pays                                                                                             | Sép. CSV | Décimale | Milliers | Date    | Papier |  |
|--------------------------------------------------------------------------------------------------|----------|----------|----------|---------|--------|--|
| États-Unis,<br>Canada (csv)                                                                      | ,        | .        | ,        | MDY/YMD | Letter |  |
| Royaume-Uni,<br>Irlande, Australie,<br>Nouvelle-Zélande                                          | ,        | .        | ,        | DMY     | A4     |  |
| Allemagne,<br>Autriche, Belgique,<br>Pays-Bas, Italie,<br>Espagne, Portugal,<br>Danemark, Brésil | ;        | ,        | .        | DMY     | A4     |  |
| France, Pologne,<br>Rép. tchèque,<br>Norvège, Lettonie,<br>Estonie                               | ;        | ,        | (espace) | DMY     | A4     |  |
| Finlande, Suède                                                                                  | ;        | ,        | (espace) | DMY/YMD | A4     |  |
| Suisse                                                                                           | ;        | ,        | '        | DMY     | A4     |  |
| Ukraine, Russie                                                                                  | ;        | ,        | (espace) | DMY     | A4     |  |
| Turquie                                                                                          | ;        | ,        | .        | DMY     | A4     |  |
| Japon, Corée du<br>Sud                                                                           | ,        | .        | ,        | YMD     | A4     |  |
| Mexique                                                                                          | ,        | .        | ,        | DMY     | Letter |  |
| Lituanie                                                                                         | ;        | ,        | (espace) | YMD     | A4     |  |

#### 3.6.2 Remplacement du format de papier

Sous le sélecteur de pays, un menu déroulant **Paper** (Letter ou A4) vous permet de remplacer le format de page du rapport sans changer de pays. Le séparateur CSV, la marque décimale et l'ordre du format de date suivent le préréglage du pays sélectionné ; lorsque vous changez de pays, le remplacement du format de papier est réinitialisé à la valeur par défaut de ce pays.

### 3.6.3 Enregistrement des valeurs par défaut

Cliquez sur **Save** dans le menu Settings pour enregistrer globalement vos préférences de locale. Ces valeurs par défaut s'appliquent à toutes les sessions futures, quel que soit le projet ouvert. Les paramètres propres à chaque projet (cible, prédicteurs, paramètres) sont enregistrés séparément dans la base de données du projet ; les valeurs par défaut de locale persistent pour tous les projets via une base de données SQLite au niveau utilisateur. Cette approche à deux niveaux est conçue pour les organisations, telles que les cabinets d'audit, qui travaillent avec des données provenant de plusieurs pays — définissez votre pays le plus fréquent comme valeur par défaut, puis ajustez au besoin.

***Astuce :** Le formatage des nombres sur les axes des graphiques et sur les étiquettes de pente s'adapte automatiquement à votre locale. La locale allemande utilise des points pour les milliers (200.000), la locale finlandaise des espaces (200 000), la locale suisse des apostrophes (200'000). Aucun symbole monétaire n'est affiché — earthUI est indépendant de la devise.*

## 4 Mode d'évaluation (Appraisal Mode)

Lorsque vous sélectionnez **For Appraisal** (pour évaluation) comme Purpose (objectif), earthUI se configure pour l'évaluation d'un bien unique. Toutes les fonctionnalités décrites au chapitre 3 restent disponibles ; le présent chapitre ne couvre que les ajouts propres à l'évaluation immobilière.

### 4.1 Traitement de la ligne du bien à évaluer

En mode évaluation, **la ligne 1 de votre jeu de données correspond au bien à évaluer** et toutes les lignes restantes sont des ventes comparables. Votre fichier d'entrée doit être organisé en conséquence (voir chapitre 2). Le prix de vente du bien à évaluer peut être laissé vide ou fixé à une valeur quelconque — earthUI le traite automatiquement comme NA lors de l'ajustement du modèle.

Après l'importation, l'onglet Data se scinde en deux sections : **Subject Property** (bien à évaluer, ligne 1) et **Comparable Sales** (ventes comparables, lignes 2 et suivantes). La ligne 1 est toujours exclue de l'ajustement du modèle — la notification « Skipping row 1 (subject). Fitting on N rows. » (ligne 1 ignorée ; ajustement sur N lignes) le confirme. Après l'ajustement, le modèle génère néanmoins des prédictions pour la ligne du bien à évaluer, affichées sous la forme `est_<target>` dans les résultats.

### 4.2 Effective Date et ancienneté des ventes (Sale Age)

Un champ **Effective Date** (date d'effet) apparaît dans la section Variable Configuration dans chaque mode d'objectif (avec la date du jour par défaut) — les analyses générales et de marché disposent elles aussi d'une date d'effet, et non les seules évaluations — mais il n'est réellement *utilisé* que dans les modes évaluation et marché. Dans ces modes, si vous désignez une colonne comme `contract_date` dans le menu déroulant Special, earthUI calcule une colonne `sale_age` — le nombre de jours entiers entre la date de contrat de chaque vente et la date d'effet. Cette colonne remplace la colonne de date d'origine comme prédicteur.

La première fois que vous cliquez sur Fit après avoir désigné une date de contrat, earthUI crée la colonne `sale_age` et vous invite à cliquer de nouveau sur Fit pour l'inclure.

### 4.3 Désignations de colonnes spéciales (Special)

Dans les modes évaluation et marché, un menu déroulant **Special** apparaît pour chaque prédicteur dans le tableau Variable Configuration. Voici la liste complète des types spéciaux et de leurs effets :

| Type spécial    | Effet                                                                                                                                                                                                |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| no              | Valeur par défaut — aucun traitement particulier                                                                                                                                                     |
| contract_date   | Déclenche le calcul automatique de <code>sale_age</code> à partir de l'Effective Date                                                                                                                |
| listing_date    | Utilisée avec la date de contrat pour calculer le DOM lorsqu'aucune colonne <code>dom</code> n'existe                                                                                                |
| dom             | Identifie la colonne Days on Market (durée de mise en marché) ; affichée dans la Sales Grid (grille de comparaison)                                                                                  |
| concessions     | Identifie les concessions de vente ; utilisée dans la formule Net SP (Sale Price – Concessions) de la Sales Grid                                                                                     |
| latitude        | Valeurs arrondies à 3 décimales ; utilisées pour la proximité haversine et le groupe Location de la Sales Grid                                                                                       |
| longitude       | Valeurs arrondies à 3 décimales ; utilisées pour la proximité haversine et le groupe Location de la Sales Grid                                                                                       |
| area            | Regroupée dans la ligne « Loc : Long   Lat   Area » de la Sales Grid                                                                                                                                 |
| living_area     | Active les résidus au pied carré ( <code>residual_sf</code> , <code>cqa_sf</code> ) dans les fichiers téléchargés                                                                                    |
| lot_size        | Regroupée dans la ligne « Site Size   Dimensions » de la Sales Grid                                                                                                                                  |
| site_dimensions | Regroupée avec la taille du terrain dans la Sales Grid                                                                                                                                               |
| actual_age      | Regroupée dans la ligne « Actual Age   Effective Age » de la Sales Grid                                                                                                                              |
| effective_age   | Regroupée avec l'âge réel dans la Sales Grid                                                                                                                                                         |
| display_only    | Colonne incluse dans les exports mais exclue de l'ajustement du modèle (plusieurs colonnes autorisées)                                                                                               |
| remarks         | (Expérimental) Colonne de remarques en texte libre convertie en colonnes de caractéristiques par l'étape Prolog facultative (plusieurs colonnes autorisées) ; voir chapitre 15                       |
| list            | (Expérimental) Colonne à valeurs séparées par des virgules scindée en colonnes TRUE/FALSE par encodage one-hot lors de l'étape Prolog facultative (plusieurs colonnes autorisées) ; voir chapitre 15 |

Chaque type spécial ne peut être affecté qu'à une seule colonne (à l'exception de `display_only`, `remarks` et `list`). L'affectation d'un type spécial à une seconde colonne le retire automatiquement de la première. Un petit badge bleu apparaît à côté du nom de la variable pour indiquer le type spécial qui lui est affecté.

#### 4.4 Residual Contribution (estimation des caractéristiques rares)

Les caractéristiques trop rares pour le canal de régression — un atelier présent dans une poignée de ventes — relèvent de la couche résiduelle de la RCA. L'onglet **Residual Contribution** (contribution résiduelle) estime leur contribution de marché en régressant les résidus du modèle ajusté sur des colonnes que le modèle n'a jamais utilisées, par la méthode des moindres carrés ordinaires. La sélection de plusieurs colonnes les estime *conjointement*, de sorte que des aménités corrélées (p. ex. `outbuilding_sqft` et `pr_has_workshop`) se partagent le crédit au lieu d'être comptées deux fois. Les indicateurs binaires produisent des estimations forfaitaires et listent, à titre de preuve, le résidu de chaque vente concernée ; les colonnes numériques produisent des contributions unitaires ; l'erreur type rapportée correspond à un écart type d'estimation ( $t \geq 2$  servant de critère usuel de signification). Les termes appuyés par moins de dix observations reçoivent automatiquement la mention CAUTION dans la note de dossier de travail générée, laquelle documente en un seul paragraphe la méthode, les estimations et les réserves pour le rapport. Seuls les prédicteurs effectivement *utilisés* par le modèle sont exclus de la liste des candidats — un indicateur rare proposé à l'ajustement mais jamais retenu par la passe avant (forward pass) demeure éligible, puisque le modèle n'a absorbé aucune part de sa valeur.

#### 4.5 Aperçu des ajustements RCA

Le bouton **Calculate RCA Adjustments & Download** (calcul et téléchargement des ajustements RCA ; section 7 de la barre latérale, visible uniquement en mode évaluation après l'ajustement du modèle) calcule des ajustements dérivés du marché pour chaque bien comparable par rapport au bien à évaluer. Le flux de travail RCA complet est décrit au chapitre 10.

Après le calcul des ajustements RCA, le bouton **Generate Sales Grid & Download** (section 8 de la barre latérale) devient disponible. Le flux de travail de la Sales Grid est décrit au chapitre 11.

***Conseil :** le score CQA (état, qualité, attrait) que vous attribuez au bien à évaluer détermine la part de la distribution des résidus attribuée à celui-ci. Un score de 5.00 place le bien à évaluer à la médiane des biens comparables.*

## 5 Mode d'analyse de zone de marché (Market Area Analysis Mode)

Lorsque vous sélectionnez **Market Area Analysis** (analyse de zone de marché) comme Purpose, earthUI offre les mêmes fonctionnalités propres à l'immobilier que le mode évaluation (colonnes spéciales, ancienneté des ventes, arrondi des coordonnées), mais est orienté vers l'analyse d'un groupe de biens plutôt que vers l'évaluation d'un unique bien à évaluer.

### 5.1 Différences par rapport au mode d'évaluation

- **Pas d'exclusion automatique du bien à évaluer** — contrairement au mode évaluation, le mode marché n'exclut pas automatiquement la ligne 1 de l'ajustement. Toutes les lignes sont incluses ; pour tenir une ligne de bien à évaluer hors de l'ajustement, attribuez-lui une valeur nulle dans une colonne de pondérations désignée (comme dans le jeu de données de démonstration) ou utilisez un filtre de sous-ensemble.
- **Pas de sections RCA ni Sales Grid** — les étapes « Calculate RCA Adjustments & Download » et « Generate Sales Grid & Download » ne sont pas disponibles. Le mode marché est centré sur l'ajustement du modèle et ses sorties, et non sur les ajustements par bien comparable.
- **Numérotation de la barre latérale** — en l'absence des étapes RCA et Sales Grid, la section Generate Quarto Report porte le numéro 7 au lieu de 9 (la section Convert Quarto Report conserve le numéro 10).

### 5.2 Quand utiliser le mode marché

Le mode Market Area Analysis convient lorsque vous :

- construisez un modèle de régression pour un quartier ou une zone de marché afin d'en comprendre les déterminants de la valeur ;
- analysez la manière dont des variables telles que la surface, l'âge, la taille du terrain et la localisation influent sur les prix de vente d'un groupe de biens ;
- préparez des éléments à l'appui d'une analyse des conditions de marché ou d'une délimitation de quartier ;
- travaillez avec un jeu de données dont la ligne 1 contient un bien à évaluer, tout en souhaitant pouvoir l'inclure dans l'ajustement ou l'en exclure.

***Conseil :** le mode marché est également utile pour la régression immobilière générale lorsque vous souhaitez bénéficier des fonctionnalités de colonnes spéciales (ancienneté des ventes, arrondi des coordonnées) sans avoir besoin du flux de travail d'ajustements RCA.*

## 6 Sélection des variables

La section 3 de la barre latérale — **Variable Configuration** (configuration des variables) — est l'endroit où l'on choisit les colonnes qui participent au modèle et la manière dont elles sont traitées.

### 6.1 Variable(s) cible(s)

La liste déroulante **Target (response) variable(s)** (variable(s) cible(s), ou de réponse), en haut de la section 3, énumère toutes les colonnes du jeu de données. Sélectionnez une seule colonne pour un modèle standard à réponse unique, ou plusieurs colonnes pour un modèle à réponses multiples. Lorsque plusieurs cibles sont sélectionnées, le modèle ajuste toutes les réponses simultanément au moyen de `earth(cbind(y1, y2, ...) ~ .)`.

Les colonnes sélectionnées comme cibles sont automatiquement exclues de la liste des prédicteurs.

### 6.2 La table des prédicteurs

Sous le sélecteur de cible, une table énumère toutes les colonnes restantes avec les champs suivants :

| Colonne         | Description                                                                                                                                                                                                                                         |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Variable</b> | Nom de la colonne (le nom complet apparaît dans une infobulle s'il est tronqué), précédé d'un <b>bouton à bascule</b> $\times/+$ (voir ci-dessous). Dans les modes évaluation immobilière et marché, un badge bleu indique le type spécial assigné. |
| <b>Type</b>     | Liste déroulante du type de données : <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>logical</code> , <code>factor</code> , <code>Date</code> , <code>POSIXct</code>                                                  |
| <b>Include</b>  | Case à cocher — inclure cette colonne comme prédicteur dans le modèle                                                                                                                                                                               |
| <b>Factor</b>   | Case à cocher — traiter cette colonne comme variable catégorielle                                                                                                                                                                                   |
| <b>Linear</b>   | Case à cocher — imposer une entrée linéaire uniquement (aucune fonction charnière (hinge))                                                                                                                                                          |
| <b>Latent</b>   | Case à cocher — écarter entièrement cette colonne du modèle (exclue de <code>earth</code> ainsi que des données exportées vers <code>glmnetUI</code> / <code>mgcvUI</code> ), en la réservant au travail sur les variables latentes                 |
| <b>Special</b>  | Liste déroulante (modes évaluation/marché uniquement) — voir la référence des types de colonnes spéciaux ci-dessous                                                                                                                                 |
| <b>T/F/NA</b>   | Trois comptages par colonne : pour les colonnes logiques, TRUE/FALSE/NA ; pour les colonnes numériques, $\neq 0$ / $= 0$ /NA. Affichés en rouge lorsque plus de 30% des valeurs sont manquantes                                                     |

Une ligne d'aide au-dessus de la table explique les abréviations : « Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`). » (soit : Type = type de données de la colonne, Inc

= inclure comme prédicteur, Factor = traiter comme catégorielle, Linear = linéaire uniquement (sans charnières), Special = rôle de la colonne, p. ex. `contract_date`).

**Désactivation douce plutôt que suppression.** Le bouton à bascule  $\times$  placé à côté de chaque nom de variable grise la ligne et écarte la colonne du modèle sans la retirer des données ; il se transforme en un  $+$  vert qui réactive la colonne. Les colonnes ne sont jamais détruites.

### 6.3 Détection des types de données et redéfinitions

earthUI détecte automatiquement les types de données à l'importation. Les colonnes numériques, entières, logiques, facteurs et dates sont reconnues. Les colonnes de type caractère qui ressemblent à des dates (au moins 80% des valeurs s'analysent selon des formats de date courants) sont classées comme **Date**.

Toute détection peut être redéfinie en modifiant la liste déroulante **Type**. Lorsque l'on change une colonne en **character** ou **factor**, la case à cocher Factor est automatiquement cochée. Le changement de type influe sur la manière dont la colonne est transmise à la fonction **earth()**.

### 6.4 Indicateurs Factor et Linear

- **Factor** — Lorsque la case est cochée, la colonne est traitée comme une variable catégorielle. Cela convient aux colonnes telles que **style**, **area\_id** ou **grade**, qui représentent des groupes discrets plutôt que des mesures continues. Les colonnes facteurs entrent dans le modèle sous forme de variables indicatrices (dummy).
- **Linear** — Lorsque la case est cochée, la colonne est contrainte d'entrer linéairement dans le modèle — aucune fonction charnière (linéaire par morceaux) n'est créée pour elle. Cela est utile lorsque l'on sait qu'une variable entretient une relation strictement linéaire avec la cible.

### 6.5 Référence des types de colonnes spéciaux

Dans les modes évaluation immobilière et marché, la liste déroulante **Special** propose les options suivantes. Chaque type ne peut être assigné qu'à une seule colonne au plus (à l'exception de **display\_only**, **remarks** et **list**, qui en admettent plusieurs) :

#### Types de date et d'heure :

- **contract\_date** — Déclenche le calcul automatique de **sale\_age**. La colonne de date d'origine est remplacée par une colonne d'entiers mesurant le nombre de jours entre la date de vente et l'Effective Date (date d'effet de l'évaluation).
- **listing\_date** — Utilisée comme solution de repli pour calculer les Days on Market (DOM = date de contrat – date de mise en vente ; nombre de jours de mise en marché) lorsqu'aucune colonne **dom** explicite n'est désignée.
- **dom** — Identifie la colonne des Days on Market. Affichée dans la ligne APN et la ligne Date of Sale du Sales Grid (grille de comparaison).

#### Types monétaires :

- **concessions** — Identifie les concessions de vente (crédits du vendeur, incitations à l'acheteur, etc.). Utilisée dans le Sales Grid pour calculer le prix de vente net : Net SP = Sale Price – Concessions.

#### Types de taille et de localisation :

- **latitude** — Les valeurs sont automatiquement arrondies à 3 décimales afin de prévenir le surajustement. Utilisée pour les calculs de proximité par la formule de Haversine (distance entre le bien à évaluer et chaque bien comparable) et regroupée dans la ligne Location du Sales Grid.

- **longitude** — Même traitement d'arrondi que la latitude. Utilisée pour la proximité et le groupe Location.
- **area** — Typiquement un identifiant de quartier ou de zone. Regroupée avec la latitude et la longitude dans la ligne « Loc : Long | Lat | Area » du Sales Grid.
- **living\_area** — Active les calculs de résidus par pied carré (**residual\_sf** et **cqa\_sf**) dans le fichier téléchargé en sortie.
- **lot\_size** — Regroupée dans la ligne « Site Size | Dimensions » du Sales Grid.
- **site\_dimensions** — Regroupée avec la taille du terrain dans le Sales Grid (p. ex. « 75x120 »).

#### Types d'âge :

- **actual\_age** — Regroupée dans la ligne « Actual Age | Effective Age » du Sales Grid.
- **effective\_age** — Regroupée avec l'âge réel dans le Sales Grid.

#### Types d'affichage :

- **display\_only** — La colonne est incluse dans les exportations Excel mais entièrement exclue de l'ajustement du modèle. À utiliser pour les champs d'adresse, les numéros MLS, les identifiants de parcelle ou d'autres données de référence qui ne doivent pas servir de prédicteur. Plusieurs colonnes peuvent recevoir cette désignation.

#### Types d'extraction de texte (expérimental) :

- **remarks** — Une colonne de texte libre (p. ex. **public\_remarks**) destinée à être analysée en colonnes de caractéristiques par l'étape Prolog optionnelle. Sans effet tant que cette fonctionnalité n'est pas activée dans Settings. Plusieurs colonnes admises. Voir le chapitre 15.
- **list** — Une colonne délimitée par des virgules (p. ex. **existing\_structures**) destinée à être scindée en colonnes TRUE/FALSE <col>\_<value> en encodage one-hot (encodage disjonctif complet) par la même étape. Plusieurs colonnes admises. Voir le chapitre 15.

## 6.6 Cibles multiples

La sélection de plus d'une variable cible ajuste un modèle Earth à réponses multiples. Lorsque plusieurs cibles sont sélectionnées :

- Le modèle prédit toutes les réponses simultanément, en partageant les fonctions de base entre les cibles
- Le bouton **wp (response weights)** (pondérations des réponses) devient actif et permet d'assigner un poids numérique à chaque cible
- Les modèles de variance (**varmod.method**) sont désactivés (non pris en charge en réponses multiples)
- Les onglets de résultats affichent, pour chaque réponse, les métriques, les équations et les graphiques de diagnostic

**Conseil :** les colonnes désignées comme **display\_only** demeurent dans le jeu de données et figurent dans l'exportation Excel, mais sont entièrement exclues de la table des prédicteurs. À utiliser pour les colonnes d'identifiants, les adresses ou d'autres champs de référence.

## 7 Sélection des paramètres

La section 4 de la barre latérale — **Earth Call Parameters** (paramètres d'appel de earth) — donne accès à l'ensemble des arguments acceptés par la fonction `earth()`. Chaque paramètre est accompagné d'une icône d'aide bleue (?) dotée d'une infobulle explicative. Les paramètres sont organisés en sous-sections repliables.

### 7.1 Paramètres de base

| Paramètre      | Valeur par défaut | Description                                                                                                                                                                          |
|----------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>subset</b>  | (vide)            | Expression de filtrage des lignes. Voir « Filtrage par sous-ensemble » ci-dessous.                                                                                                   |
| <b>weights</b> | NULL              | Sélecteur de colonne pour les poids d'observation (par ligne). Seules les colonnes numériques sont proposées.                                                                        |
| <b>wp</b>      | NULL              | Poids des réponses pour les modèles à cibles multiples. Un bouton ouvre une boîte de dialogue comportant un champ numérique par cible (1.0 par défaut pour chacune).                 |
| <b>keepxy</b>  | off               | Conserve x, y, subset et weights dans l'objet modèle.                                                                                                                                |
| <b>trace</b>   | 0                 | Niveau de trace pour la sortie de l'ajustement (de 0 à 5).                                                                                                                           |
| <b>glm</b>     | none              | Famille GLM optionnelle : <b>gaussian</b> , <b>binomial</b> ou <b>poisson</b> .                                                                                                      |
| <b>degree</b>  | 1                 | Ordre maximal d'interaction. Fixer $\text{degree} \geq 2$ active automatiquement la validation croisée et fait apparaître la matrice Allowed Interactions (interactions autorisées). |
| <b>penalty</b> | 2                 | Pénalité GCV par noeud. Des valeurs plus élevées produisent des modèles plus simples.                                                                                                |

### 7.2 Passe avant

| Paramètre | Valeur par défaut | Description                                                                                                                                                      |
|-----------|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>nk</b> | auto              | Nombre maximal de termes avant élagage. Prérempli avec la valeur recommandée pour les données chargées (voir « Valeurs de paramètres recommandées » ci-dessous). |

| Paramètre             | Valeur par défaut | Description                                                                                                                                       |
|-----------------------|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>thresh</b>         | 0.001             | Seuil de la passe avant. Des valeurs plus petites autorisent davantage de termes.                                                                 |
| <b>minspan</b>        | auto              | Espacement minimal entre les nœuds. Prérempli avec la valeur recommandée. Les valeurs négatives fixent le nombre maximal de nœuds par prédicteur. |
| <b>endspan</b>        | auto              | Espacement aux extrémités — distance minimale entre un nœud et le bord des données. Prérempli avec la valeur recommandée.                         |
| <b>newvar.penalty</b> | 0                 | Pénalité pour l'introduction d'une nouvelle variable (favorise la réutilisation des prédicteurs déjà présents).                                   |
| <b>fast.k</b>         | 20                | Nombre de termes parents considérés dans l'algorithme fast MARS.                                                                                  |
| <b>fast.beta</b>      | 1                 | Contrôle le facteur de vieillissement de fast MARS.                                                                                               |

### 7.3 Interactions autorisées

Lorsque **degree** est fixé à 2 ou plus, une matrice d'interactions apparaît sous les paramètres de base. Il s'agit d'une grille triangulaire supérieure de cases à cocher, une pour chaque paire de prédicteurs. Une case cochée signifie que les deux prédicteurs sont autorisés à interagir ; la décocher interdit cette interaction spécifique.

Un clic sur le nom d'un prédicteur (étiquette de ligne ou de colonne) bascule toutes les interactions de cette variable. Les cases **Allow All** (tout autoriser) et **Clear All** (tout effacer), en haut, permettent un contrôle global. La matrice utilise des en-têtes fixes (« sticky ») afin que les étiquettes de lignes et de colonnes restent visibles lors du défilement.

Un message d'information le rappelle : « Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold). » (les termes d'interaction accroissent le risque de surajustement ; la validation croisée a été activée, en 10 plis).

### 7.4 Élagage

| Paramètre      | Valeur par défaut | Description                                                                                                           |
|----------------|-------------------|-----------------------------------------------------------------------------------------------------------------------|
| <b>pmethod</b> | backward          | Méthode d'élagage : <b>backward</b> , <b>none</b> , <b>exhaustive</b> , <b>forward</b> , <b>seqrep</b> ou <b>cv</b> . |
| <b>nprune</b>  | NULL              | Nombre maximal de termes après élagage. Laisser vide pour une sélection automatique.                                  |

### 7.5 Validation croisée

| Paramètre       | Valeur par défaut | Description                                                                                                 |
|-----------------|-------------------|-------------------------------------------------------------------------------------------------------------|
| <b>nfold</b>    | 10                | Nombre de plis de validation croisée. Fixer à 0 pour désactiver la CV.                                      |
| <b>ncross</b>   | 20                | Nombre de répétitions de la validation croisée.                                                             |
| <b>stratify</b> | on                | Stratifie les échantillons de CV de sorte que chaque pli présente une distribution de la réponse similaire. |

Lorsque  $\text{degree} \geq 2$ , la validation croisée est activée automatiquement (nfold fixé à 10) afin de mieux se prémunir contre le surajustement induit par les interactions.

## 7.6 Filtrage par sous-ensemble

Le champ de saisie **subset** accepte une expression R qui filtre les lignes utilisées pour l'ajustement du modèle. On peut saisir une expression directement (p. ex. `sale_age < 365 & area_id == 460`) ou utiliser le bouton **Build filter...** (construire un filtre) pour la composer visuellement.

**Boîte de dialogue Build Filter** — Cliquer sur « Build filter... » ouvre une boîte de dialogue guidée. Chaque ligne de condition comporte une liste déroulante de colonnes, un opérateur (<, >, <=, >=, ==, !=) et un champ de valeur qui s'adapte au type de la colonne : saisie numérique pour les nombres, sélecteur de date pour les dates et liste déroulante des valeurs uniques pour les colonnes de type caractère ou factor. Les conditions sont reliées par des connecteurs AND (&) ou OR (|). Un aperçu en bas de la boîte de dialogue affiche l'expression et le nombre de lignes correspondantes. Cliquer sur **Apply** (appliquer) insère l'expression dans le champ de saisie.

**Les colonnes de dates** doivent utiliser les enveloppes `as.Date("...")` ou `as.POSIXct("...")` dans les expressions saisies manuellement. La boîte de dialogue Build Filter s'en charge automatiquement.

**Le filtrage par sous-ensemble est non destructif** — les lignes exclues restent dans le jeu de données et reçoivent des prédictions dans l'export Excel.

***Conseil :** utilisez des parenthèses pour grouper les conditions OR lorsqu'elles sont combinées avec AND. Sans parenthèses,  $A \ \& \ B \ | \ C$  est évalué comme  $(A \ \& \ B) \ | \ C$ , ce qui peut ne pas correspondre à votre intention.*

## 7.7 Valeurs de paramètres recommandées

earthUI affiche une valeur recommandée sous les paramètres clés de la section 4. Ces recommandations se mettent à jour de manière réactive en fonction du nombre de lignes d'ajustement ( $n$ ) et des prédicteurs sélectionnés ( $p$ ). Les formules sont dérivées de l'article MARS de Friedman, des algorithmes internes de earth et de tests empiriques. Les champs **nk**, **minspan** et **endspan** sont préremplis avec leurs valeurs recommandées au chargement des données ; une modification manuelle est conservée.

### 7.7.1 Paramètres de la passe avant

**nk** (nombre maximal de termes avant élagage) :

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

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

La valeur par défaut de `earth`,  $\min(200, \max(20, 2p)) + 1$ , ne tient pas compte de la taille du jeu de données et peut se révéler trop faible pour de grands jeux de données, ce qui contraint la passe avant avant qu'elle n'ait exploré adéquatement tous les prédicteurs. Le terme  $\lfloor n/10 \rfloor$  autorise environ un terme pour 10 observations — une règle empirique standard pour éviter la surparamétrisation. Le plafond de 100 prévient les rendements décroissants.

**minspan** (nombre minimal d'observations entre les nœuds) :

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

**endspan** (nombre minimal d'observations depuis les bords des données) :

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

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

Le `minspan` calculé automatiquement par `earth` utilise l'équation 43 de Friedman :  $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$ , qui croît en  $\ln(np)$  et augmente trop lentement pour les grands jeux de données. À  $n=1500$ , elle ne donne que  $\approx 7$ , soit environ 245 emplacements de nœuds candidats par prédicteur continu — un nombre trop élevé, qui permet à la passe avant d'ajuster le bruit. La formule recommandée vise environ 100 candidats par prédicteur pour les grands  $n$ , tout en s'en remettant au calcul automatique de `earth` pour les petits  $n$  (où celui-ci est bien éprouvé). `Endspan` est fixé légèrement au-dessus de `minspan` afin d'apporter une protection supplémentaire aux bords, ce qui est utile lorsque les données présentent des queues de distribution peu fournies (situation courante en immobilier).

**penalty** (pénalité GCV par nœud) :

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

Il s'agit de la valeur par défaut de `earth` — 2 pour les modèles additifs, 3 pour les modèles avec interactions (la pénalité plus élevée compense l'espace de recherche plus vaste).

**newvar.penalty** (pénalité pour l'introduction d'un nouveau prédicteur) :

Recommandation : **0.1** lorsque les prédicteurs sont corrélés (p. ex. la surface habitable avec le nombre de chambres, le nombre de salles de bains, la superficie du terrain). Cette pénalité

oriente la passe avant vers la réutilisation des prédicteurs déjà présents dans le modèle plutôt que vers l'introduction d'alternatives corrélées. Le mécanisme : l'amélioration de la RSS apportée par chaque nouveau prédicteur est multipliée par  $1/(1 + \text{penalty})$ . Avec `newvar.penalty = 0.1`, une nouvelle variable doit améliorer la RSS d'au moins 10% de plus qu'un nœud supplémentaire sur une variable existante. On obtient ainsi des modèles plus simples, comportant moins de prédicteurs, sans affecter les estimations finales des coefficients (la pénalité est retirée après la sélection). Laisser à 0 si les prédicteurs ne sont pas corrélés. Noter que `earth` ne prend pas en charge un `newvar.penalty` non nul conjointement avec des poids d'observation — `earthUI` désactive le champ (en forçant 0) dès qu'une colonne de poids est désignée.

### 7.7.2 Paramètres d'élagage

**pmethod** : recommandation : **backward**. L'élagage arrière fondé sur le GCV, méthode par défaut, est déterministe — les mêmes données produisent toujours le même modèle, sans dépendance à des graines aléatoires. Ce point est important pour la reproductibilité, en particulier dans le travail d'évaluation immobilière. La méthode `cv` utilise la validation croisée pour l'élagage, ce qui introduit une dépendance à la graine aléatoire.

**nprune** : recommandation : **laisser vide** (NULL). Laisser le GCV sélectionner le nombre optimal de termes. Fixer `nprune` impose un plafond strict qui prime sur le jugement du GCV.

### 7.7.3 Paramètres de validation croisée

**nfold** (plis de CV) : recommandation par tranche de taille d'échantillon :

| $n$ (lignes d'ajustement) | nfold |
|---------------------------|-------|
| 0–299                     | 5     |
| 300–399                   | 6     |
| 400–499                   | 7     |
| 500–599                   | 8     |
| 600–699                   | 9     |
| 700+                      | 10    |

Un plus grand nombre de plis entraîne chaque pli sur davantage de données (une estimation moins biaisée) mais réduit le pli de test mis de côté. Le tableau progresse jusqu'au standard de 10 plis à mesure que l'échantillon croît, tout en maintenant chaque pli de test suffisamment grand (environ 50 lignes ou plus) pour une métrique par pli stable, et retombe à 5 plis pour les petits échantillons — où l'estimation du  $\text{CVR}^2$  demeure bruitée ; il convient alors d'augmenter `ncross` (p. ex. 15–20) pour moyenniser la variance. Le plafond de 10 traduit le fait que chaque pli supplémentaire représente un ajustement `earth` complet pour un gain diagnostique négligeable. Avec `pmethod = "backward"`, la validation croisée n'affecte pas le modèle — elle ne fait que calculer le  $\text{CVR}^2$  diagnostique et fournir des résidus pour le modèle de variance.

**ncross** (répétitions de CV) : recommandation par tranche de taille d'échantillon :

| $n$ (lignes d'ajustement) | ncross |
|---------------------------|--------|
| 0–199                     | 20     |
| 200–399                   | 16     |
| 400–699                   | 12     |
| 700–1199                  | 10     |
| 1200–2999                 | 7      |
| 3000+                     | 5      |

Les répétitions moyennent les aléas du partitionnement en plis afin de stabiliser le  $CVR^2$  ; les grands échantillons nécessitent moins de répétitions, car une seule validation croisée y est déjà plus stable. Les résidus de la CV répétée alimentent également le modèle de variance.

#### 7.7.4 Modèle de variance

**varmod.method** : recommandation : **lm** — la valeur par défaut de Milborrow : un modèle de variance linéaire robuste pour les intervalles de prédiction (la dispersion étant estimée à partir de résidus bruités, un modèle simple généralise mieux). Si **lm** ne converge pas sur vos données, essayez **rlm** ou **const** (homoscédastique, converge toujours), ou repliez-vous sur **earth** (qui gère aussi une dispersion non linéaire). Cela exige **nfold** > 1 et **ncross** > 1.

#### 7.7.5 Fichier de sortie earth

Après chaque ajustement réussi, earthUI écrit `<filename>_earth_output_<timestamp>.txt` dans le dossier de sortie. Ce fichier contient les termes du modèle, les statistiques récapitulatives ( $R^2$ ,  $GRSq$ ,  $CVRSq$ ), les détails du modèle de variance et le journal de trace. Un fichier est créé par ajustement, constituant un registre cumulatif permettant de comparer les configurations de paramètres.

### 7.8 Réglages par défaut

Un bouton radio en haut de la section 4 détermine quelles valeurs par défaut sont chargées :

- **Use last settings for input file** (utiliser les derniers réglages pour le fichier d'entrée) — restaure les réglages utilisés la dernière fois avec ce fichier particulier
- **Use default settings** (utiliser les réglages par défaut) — applique vos valeurs par défaut personnalisées enregistrées
- **Earth defaults** (valeurs par défaut de earth) — réinitialise tous les paramètres aux valeurs d'usine

Cliquer sur **Save current as default** (enregistrer comme valeurs par défaut), en bas de la section 4, mémorise les paramètres earth courants et la matrice Allowed Interactions ; le bouton correspondant de la section 3 mémorise la configuration de la cible, de la date d'effet et des prédicteurs. Les deux sont enregistrés par projet et par mission dans la base de données du projet.

## 8 Ajustement du modèle Earth

### 8.1 Le bouton Fit

La section 5 de la barre latérale contient le bouton **Fit Earth Model** (ajuster le modèle Earth) accompagné d'un champ **Random seed** (graine aléatoire). La graine rend reproductible l'affectation des plis de validation croisée — une liste déroulante rappelle les 10 dernières graines utilisées, et la graine employée est consignée dans l'onglet Earth Output après l'ajustement. Avant l'ajustement, earthUI valide votre configuration — une variable cible doit être sélectionnée et au moins un prédicteur doit être inclus. Dans les modes évaluation immobilière et marché, les colonnes de latitude et de longitude sont arrondies, et l'ancienneté de la vente est calculée à partir de la date d'effet et de la date de contrat (si elles sont désignées).

### 8.2 Fenêtre modale d'ajustement et sortie de trace

Lorsque vous cliquez sur Fit, une superposition modale sombre apparaît, comportant :

- **En-tête** — « Fitting Earth Model » avec un chronomètre affichant les secondes écoulées et un bouton **Abort** (interrompre) qui annule l'ajustement en arrière-plan
- **Journal de trace** — une zone de texte défilante à chasse fixe affichant en temps réel la sortie de la fonction `earth()` : informations sur le jeu de données, progression de la passe avant, plis de validation croisée et temps d'achèvement
- **Code couleur** — lignes ordinaires en vert, lignes de validation croisée en jaune, erreurs en rouge

earthUI ajuste les modèles de manière asynchrone au moyen de `callr::r_bg()`, qui exécute le calcul dans un processus R distinct. L'application reste ainsi réactive pendant les ajustements de longue durée. Un observateur d'interrogation cadencé à 300 ms lit la sortie du processus d'arrière-plan et la transmet en continu au journal de trace.

Lorsque l'ajustement se termine, un message « Done in X.Xs » apparaît, un bouton de fermeture (X) est ajouté à la fenêtre modale, et l'arrière-plan est retiré afin que les onglets de résultats puissent s'afficher à la demande. La fenêtre modale se ferme ensuite automatiquement une fois que les sorties visibles ont fini de s'afficher — ou vous pouvez la fermer plus tôt.

En cas de succès, une coche verte est apposée au bouton Fit. Si une erreur survient, le message d'erreur est affiché dans le journal de trace et une notification apparaît.

### 8.3 Journal d'ajustement

Chaque ajustement (réussi ou non) écrit un fichier journal dans votre dossier de sortie :

- **Nom de fichier** : `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Contenu** : horodatage, l'ensemble de la sortie de trace du processus d'ajustement, ainsi que tout message d'erreur
- **Emplacement** : le dossier de sortie du projet actif

***Astuce** : si `callr` n'est pas installé, earthUI revient à un ajustement synchrone avec une simple barre de progression. Installez `callr` pour bénéficier de l'expérience complète de sortie de trace.*

## 9 Téléchargement des données

Après l’ajustement, la section 6 de la barre latérale — intitulée « Download Estimated Target Variable(s) & Residuals » en mode général et « Download Estimated Sale Prices & Residuals » dans les modes évaluation immobilière et marché — télécharge un fichier Excel contenant les prédictions et les diagnostics. Cette sortie est utilisée à l’étape 7 (RCA) pour attribuer une note CQA (Condition-Quality-Appeal, soit état, qualité, attrait) au bien à évaluer. La sortie est triée selon `residual_sf` et `cqa_sf` afin de vous aider à apprécier la position du bien à évaluer dans le classement. Si le modèle est de bonne qualité, les biens devraient être classés du moins attrayant au plus attrayant sur la base des caractéristiques résiduelles qui ne sont pas entrées dans la régression. La valeur médiane devrait être approximativement 0, la moitié inférieure présentant des valeurs négatives et la moitié supérieure des valeurs positives. Vous devriez trouver les logements de la plus mauvaise qualité, ou « fixers » (biens à rénover), vers le bas du classement et les plus beaux logements en haut. Il existe généralement des exceptions correspondant à des anomalies telles que les saisies immobilières, les ventes à découvert (short sales), les ventes successorales (liées à un héritage) et les ventes rapides motivées par un changement d’emploi ou d’autres raisons. L’examen des anomalies révèle habituellement une raison pertinente expliquant l’anomalie de prix.

Le libellé du bouton s’adapte au mode d’utilisation :

- Général/Marché : **Download Output (Excel)**
- Évaluation immobilière : **Download Intermediate Output (Excel)**

Le format du nom de fichier est `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`.

### 9.1 Colonnes de sortie

Pour chaque variable cible, les colonnes suivantes sont ajoutées :

| Colonne                                    | Description                                                                                     |
|--------------------------------------------|-------------------------------------------------------------------------------------------------|
| <code>est_&lt;target&gt;</code>            | Prédiction du modèle, p. ex. <code>est_sale_price</code> (1 décimale)                           |
| <code>residual</code>                      | Valeur observée moins valeur prédite (1 décimale)                                               |
| <code>cqa</code>                           | Score Condition-Quality-Appeal (2 décimales, plage 0–10)                                        |
| <code>residual_sf</code>                   | Résidu divisé par la surface habitable (si désignée, 1 décimale)                                |
| <code>cqa_sf</code>                        | CQA calculé à partir du classement via <code>residual_sf</code> (2 décimales)                   |
| <code>&lt;variable&gt;_contribution</code> | Valeur de contribution par fonction g (1 décimale, une colonne par fonction g)                  |
| <code>basis</code>                         | Contribution à la valeur de l’ordonnée à l’origine, identique pour tous les biens (1 décimale)  |
| <code>calc_residual</code>                 | Valeur observée moins (basis + toutes les contributions) — colonne de vérification (1 décimale) |

Pour les modèles à cibles multiples, un suffixe `<i>` est ajouté afin de distinguer les colonnes de chaque cible.

### 9.2 Ordre des colonnes et mise en forme Excel

Les colonnes de classement sont placées **le plus à gauche** dans le fichier de sortie, dans cet ordre : `residual_sf`, `cqa_sf`, `residual`, `cqa`. Cela facilite le parcours de la liste triée et l’appréciation de la position du bien à évaluer dans la distribution des CQA.

Ces colonnes reçoivent une mise en forme numérique Excel :

| Colonne                  | Format                 | Exemple |
|--------------------------|------------------------|---------|
| <code>residual_sf</code> | Numérique, 2 décimales | 12.34   |
| <code>cqa_sf</code>      | Numérique, 2 décimales | 7.25    |
| <code>residual</code>    | Numérique, 0 décimale  | 15,200  |
| <code>cqa</code>         | Numérique, 2 décimales | 6.80    |

### 9.3 Scores CQA

Le score CQA (Condition-Quality-Appeal) classe le résidu de chaque ligne par rapport à l'ensemble des autres résidus. Pour une ligne donnée, le CQA est le pourcentage de lignes présentant un résidu signé plus petit, multiplié par 10. On obtient ainsi une échelle de 0 à 10 où :

- **CQA élevé** ( $\approx 9-10$ ) — le bien s'est vendu beaucoup plus cher que la prédiction du modèle (résidu positif important)
- **CQA faible** ( $\approx 0-1$ ) — le bien s'est vendu beaucoup moins cher que la prédiction (résidu négatif important)
- **CQA  $\approx 5$**  — le résidu est proche de la médiane de l'ensemble des résidus

Lorsqu'une colonne `living_area` est désignée, `cqa_sf` fournit le même classement sur la base des résidus par pied carré.

### 9.4 Tri et ligne du bien à évaluer

Dans les modes évaluation immobilière et marché, les lignes des biens comparables sont triées par `residual_sf` décroissant (ou `residual` si aucune surface habitable n'est désignée). La ligne du bien à évaluer (ligne 1) demeure en position 1 et n'est pas triée. En mode général, les lignes sont exportées dans leur ordre d'origine.

Ce tri, combiné aux colonnes de classement placées le plus à gauche, permet à l'évaluateur de parcourir rapidement les comparables, du plus surestimé au plus sous-estimé par le modèle, et d'apprécier la position du score CQA attribué au bien à évaluer dans cette distribution.

Les niveaux des variables facteurs dans les données de prédiction sont alignés sur les données d'entraînement avant l'appel à `predict()`. Les lignes comportant des niveaux de facteur non rencontrés à l'entraînement produiront des prédictions NA.

***Astuce :** la colonne `calc_residual` doit toujours être égale à `residual` (aux arrondis près). Un écart significatif entre les deux signale un problème de calcul méritant investigation.*

### 9.5 Export automatique vers mgcvUI

À chaque ajustement réussi avec `degree`  $\leq 2$ , `earthUI` enregistre automatiquement l'objet résultat complet sous forme de fichier `.rds` dans le dossier de sortie du projet. Le nom de fichier suit le modèle `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

Ce fichier peut être chargé par le module de post-traitement `mgcv` — inclus dans le package `earthUI` et lancé au moyen de `earthUI::launch_mgcv()` — via `readRDS()`. La routine GAM utilise les emplacements de nœuds et les fonctions de base du modèle `earth` comme points de départ pour les termes de lissage GAM, permettant une transition sans rupture de la modélisation MARS vers la modélisation GAM.

Les modèles avec `degree`  $> 2$  sont ignorés, car `mgcvUI` ne prend en charge que les interactions par paires. Lorsque `earthUI` est lancé en mode Trilogy (depuis l'interface Trilogy UI), une section supplémentaire **Lock Model Output** (verrouillage de la sortie du modèle) permet de verrouiller

le fichier `.rds` d'un ajustement donné — avec sa base de prédicteurs et ses paramètres CQA partagés — comme modèle earth canonique importé par `glmnetUI` et `mgcvUI`.

## 9.6 Import dans `glmnetUI`

Le même fichier `.rds` enregistré pour la routine GAM peut également être importé dans le **module de post-traitement elastic net** — lui aussi inclus dans le package `earthUI` et lancé au moyen de `earthUI::launch_glmnet()`. Dans l'application `glmnet`, rendez-vous dans **Section 2 : Import from earthUI** et utilisez le bouton Browse (parcourir) pour sélectionner le fichier `.rds`.

`glmnetUI` recourt à l'approche standard `earth::model.matrix()` : les fonctions de base du modèle earth (fonctions charnières, interactions et termes linéaires) deviennent les colonnes de la matrice de conception de `glmnet`. `glmnet` effectue ensuite une régression régularisée sur ces colonnes de base, en les sélectionnant et en les contractant via le lasso ou l'elastic net. Cette démarche combine la construction adaptative de la base par earth et la régularisation de `glmnet`.

### 9.6.1 Principaux avantages

- **Interactions interprétables** : les interactions d'earth fondées sur les fonctions charnières (y compris les termes à trois facteurs) ont une signification géométrique claire, contrairement aux interactions natives de `glmnet` par produits croisés, difficiles à expliquer devant un tribunal ou dans un contexte d'audit.
- **Non-linéarité automatique** : les fonctions charnières d'earth capturent des relations non linéaires qu'un modèle `glmnet` standard fondé sur les prédicteurs bruts ne détecterait pas.
- **Sélection de variables** : `glmnet` peut élaguer davantage la base earth, en éliminant les termes charnières qui ne contribuent pas.

### 9.6.2 Important : la colonne de pondération doit correspondre

#### Avertissement

Si `earthUI` utilise une colonne de pondération (p. ex. pour exclure la ligne du bien à évaluer avec un poids de 0), la même colonne de pondération doit être désignée dans la configuration des variables de `glmnetUI`. Si les paramètres de pondération diffèrent, les nombres de lignes ne concorderont pas et la base earth ne pourra pas être adjointe à la matrice de modèle de `glmnet`.

### 9.6.3 Flux de travail

1. Ajuster un modèle earth dans `earthUI` (degré quelconque).
2. Le fichier résultat `.rds` est enregistré automatiquement dans le dossier de sortie du projet.
3. Dans `glmnetUI`, ouvrir **Section 2 : Import from earthUI** et parcourir jusqu'au fichier `.rds`.
4. Vérifier que le **même fichier de données** et la **même colonne de pondération** sont utilisés dans les deux applications.
5. Configurer les paramètres de `glmnet` dans la section 5 (la matrice d'interaction est automatiquement désactivée lorsqu'un import `earthUI` est actif).
6. Cliquer sur **Fit Model**. Les colonnes de la base earth apparaissent dans les onglets Equation, Coefficients et Contributions.

## 10 Calculs RCA et téléchargement

Le bouton **Calculate RCA Adjustments & Download** (calcul des ajustements RCA et téléchargement) apparaît dans la section 7 de la barre latérale, visible uniquement en mode évaluation après l’ajustement d’un modèle. La RCA (Residual Constraint Approach) produit des ajustements dérivés du marché, par bien comparable, relatifs au bien à évaluer.

### 10.1 Ouverture de la boîte de dialogue RCA

Un clic sur le bouton ouvre une petite boîte de dialogue modale comportant :

- **Score type** (type de score) — boutons radio permettant de choisir **CQA** ou **CQA per SF** (l’option SF n’est disponible que lorsqu’une colonne `living_area` a été désignée). Si l’on choisit « CQA per SF », alors, sur la base du score CQA attribué au bien à évaluer, son score résiduel sera égal à sa surface habitable multipliée par le `residual_sf` correspondant au score CQA\_SF donné.
- **CQA value** (valeur CQA) — champ de saisie du score CQA du bien à évaluer (0.00–10.00). Ce score représente la position que vous attribuez au bien à évaluer dans la distribution de qualité des biens comparables.
- Bouton **Generate** (générer) — calcule la sortie RCA et la télécharge au format Excel. Le bouton reste désactivé tant qu’un score valide n’a pas été saisi.

### 10.2 Interpolation du score CQA

Lorsque vous cliquez sur Generate, earthUI interpole le résidu du bien à évaluer à partir des couples CQA/résidu des biens comparables :

1. Les scores CQA et les résidus des biens comparables sont triés par CQA croissant
2. Une interpolation linéaire (`stats::approx()`) fait correspondre la valeur CQA saisie à un résidu
3. En cas d’utilisation de CQA per SF, le résidu par SF est reconverti en résidu total par multiplication par la surface habitable du bien à évaluer
4. La valeur estimée du bien à évaluer est calculée comme suit : prédiction du modèle + résidu interpolé

### 10.3 Colonnes de sortie

Le fichier Excel RCA (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) comprend toutes les colonnes de sortie intermédiaires, auxquelles s’ajoutent :

| Colonne                                  | Description                                                                      |
|------------------------------------------|----------------------------------------------------------------------------------|
| <code>subject_value</code>               | Prédiction du modèle + résidu interpolé (ligne 1 uniquement)                     |
| <code>subject_cqa</code>                 | Le score CQA saisi (ligne 1 uniquement)                                          |
| <code>&lt;variable&gt;_adjustment</code> | Contribution du bien à évaluer moins contribution du comparable (par fonction g) |
| <code>residual_adjustment</code>         | Résidu du bien à évaluer moins résidu du comparable                              |
| <code>net_adjustments</code>             | Somme de tous les ajustements (contribution + résidu)                            |
| <code>gross_adjustments</code>           | Somme des valeurs absolues de tous les ajustements                               |
| <code>residual_pct</code>                | Ajustement résiduel en fraction du prix de vente du comparable                   |
| <code>net_adj_pct</code>                 | Ajustements nets en fraction du prix de vente du comparable                      |

| Colonne             | Description                                                  |
|---------------------|--------------------------------------------------------------|
| gross_adj_pct       | Ajustements bruts en fraction du prix de vente du comparable |
| adjusted_sale_price | Prix de vente du comparable + ajustements nets               |

Les colonnes d'ajustement indiquent, pour chaque bien comparable, la part de la différence que le modèle attribue à chaque variable. **adjusted\_sale\_price** est le prix de vente du comparable après application de tous les ajustements dérivés du modèle — un ensemble de prix de vente ajustés qui devraient se regrouper autour de la valeur estimée du bien à évaluer.

#### 10.4 L'onglet RCA Adjustments

Une fois la sortie RCA calculée, l'onglet **RCA Adjustments** (ajustements RCA) du panneau principal affiche des histogrammes des distributions Residual Adj %, Net Adj % et Gross Adj % sur l'ensemble des biens comparables (un jeu par cible pour les modèles multi-cibles), offrant un contrôle visuel rapide du degré de regroupement des ajustements.

***Conseil :** un score CQA de 5.00 place le bien à évaluer à la médiane des biens comparables. Un score supérieur à 5 indique un bien de qualité supérieure à la moyenne ; un score inférieur à 5, une qualité inférieure à la moyenne.*

## 11 Sales Comparison Grid

### 11.1 Vue d'ensemble et objet

La **Sales Comparison Grid** (grille de comparaison) est un classeur Excel généré en mode évaluation immobilière (section 8 de la barre latérale) après le calcul des ajustements RCA. Elle présente le bien à évaluer aux côtés des biens comparables sélectionnés dans un format de grille structuré, avec des formules Excel qui calculent automatiquement les ajustements ainsi qu'un prix de vente ajusté pour chaque comparable.

La grille est conçue pour le dossier de travail de l'évaluateur — elle combine les ajustements issus de la régression du modèle Earth avec des cellules modifiables dans lesquelles l'évaluateur peut ventiler le résidu CQA entre des caractéristiques spécifiques du bien. Le nom du fichier de sortie est **SalesGrid\_<YYYYMMDD\_HHMMSS>.xlsx**.

### 11.2 Fenêtre modale de sélection des comparables

Un clic sur « Generate Sales Grid & Download » ouvre une fenêtre modale dans laquelle on sélectionne les biens comparables à inclure :

- Les **Recommended comps** (comparables recommandés) sont cochés par défaut. Il s'agit de biens comparables présentant un pourcentage d'ajustement brut inférieur à 25% du prix de vente, triés par ancienneté de vente croissante (ventes les plus récentes en premier) et plafonnés à 30.
- Les **Additional comps** (comparables supplémentaires) sont listés en dessous, non cochés par défaut. Leurs ajustements bruts atteignent 25% ou plus, mais ils restent disponibles à la sélection (jusqu'à 50 affichés, triés par pourcentage d'ajustement brut croissant).
- Un maximum de **30 comparables** peut être sélectionné, produisant jusqu'à **10 feuilles** (3 comparables par feuille).

Après confirmation de la sélection, les comparables retenus sont triés par pourcentage d'ajustement brut croissant, de sorte que les biens comparables les plus similaires apparaissent sur la feuille 1.

### 11.3 Disposition de la grille

Chaque feuille comporte une disposition sur 20 colonnes accueillant le bien à évaluer plus 3 biens comparables. Les colonnes de chaque entité sont les suivantes :

| Entité            | Colonnes (5 par entité)                                                                |
|-------------------|----------------------------------------------------------------------------------------|
| Bien à évaluer    | Label, Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC)      |
| Chaque comparable | Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC), Adjustment |

Les lignes, de haut en bas, sont les suivantes :

1. **Titre** — nom de la feuille (p. ex. « Intermediate Sales Comparable Grid — Sheet 1 of 3 »)
2. **En-têtes** — en-têtes de colonnes « Subject » et comparables, avec l'adresse
3. **Adresse** — adresse complète du bien
4. **APN | MLS# | DOM | Subj.Prox** — numéro de parcelle, identifiant du listing, nombre de jours sur le marché et distance de Haversine (en miles) par rapport au bien à évaluer
5. **Sales Price | Concess. | Net SP** — prix de vente, concessions et formule du prix de vente net

6. Ligne d'en-tête **Regression Features**
7. **BASE VALUE** — l'ordonnée à l'origine du modèle
8. **Date of Sale | OffMkt | OnMkt** — date du contrat, ancienneté de la vente et DOM
9. **Lignes groupées** (conditionnelles) — groupes Location, Site Size et/ou Age
10. **Lignes de variables du modèle** — une ligne par prédictor (hors variables groupées)
11. **Ligne de séparation vide**
12. Ligne d'en-tête **Residual Features**
13. **CQA|Residual** — score CQA + formule du résidu restant
14. **Lignes de caractéristiques résiduelles** — lignes nommées + lignes vides pour la saisie de l'évaluateur
15. **Total VC / Net Adjustment**
16. **Net Adjustment %**
17. **Gross Adjustment %**
18. **Adjusted Sale Price** — ligne de formules
19. Pied de page **Copyright**

#### 11.4 Prix de vente, concessions et Net SP

La ligne 5 affiche trois valeurs pour chaque bien comparable :

- **Sale Price** — le prix de vente du comparable issu des données
- **Concessions** — la valeur de la colonne désignée comme **concessions** (0 si aucune n'est désignée)
- **Net SP** — une formule Excel : Sale Price – Concessions

La colonne du bien à évaluer affiche « N/A » pour le prix de vente (puisqu'il est inconnu) et, le cas échéant, la valeur des concessions si elle est disponible.

#### 11.5 Lignes groupées (localisation, site, âge)

Lorsque certains types spéciaux sont désignés et que ces variables figurent dans le modèle, earthUI crée des lignes groupées qui combinent les variables apparentées :

**Loc : Long | Lat | Area** — Apparaît lorsque l'une des variables **longitude**, **latitude** ou **area** figure dans le modèle. Affiche les valeurs factuelles de chaque variable constitutive, une contribution à la valeur combinée (somme des VC individuelles) et, pour les comparables, un ajustement (VC combinée du bien à évaluer – VC combinée du comparable). Les cellules VC sont stylisées avec un fond bleu clair.

**Site Size | Dimensions** — Apparaît lorsque **lot\_size** ou **site\_dimensions** figure dans le modèle. Même structure que le groupe Location.

**Actual Age | Effective Age** — Apparaît lorsque **actual\_age** ou **effective\_age** figure dans le modèle. Même structure.

Les variables consommées par les lignes groupées sont **exclues des lignes de variables individuelles du modèle** situées en dessous, ce qui évite tout double comptage dans les totaux d'ajustement.

#### 11.6 Lignes de variables du modèle

Sous les lignes groupées (ou directement sous BASE VALUE en l'absence de lignes groupées), une ligne par prédictor restant du modèle affiche :

- Les **valeurs factuelles** pour le bien à évaluer et chaque comparable (les valeurs réelles des données)
- La **contribution à la valeur (VC)** — la contribution de la fonction g à la valeur prédite pour cette ligne

- L'**ajustement** (comparables uniquement) — VC du bien à évaluer moins VC du comparable

### 11.7 Ligne CQA|Residual et lignes de caractéristiques résiduelles

La ligne **CQA|Residual** affiche le score CQA de chaque bien et contient une formule pour le **résidu restant** — la part du résidu non encore ventilée entre des caractéristiques spécifiques. La formule est :

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

En dessous se trouvent les **lignes de caractéristiques résiduelles** — 5 lignes nommées (Location, View/Appeal, Condition, Quality, Other, soit localisation, vue/attract, état, qualité, autre) plus 6 lignes vides. Ce sont des cellules de saisie dans lesquelles l'évaluateur inscrit les contributions à la valeur des caractéristiques non capturées par le modèle. À mesure que l'évaluateur saisit des valeurs, la formule du résidu restant (Remaining Residual) diminue automatiquement.

Pour chaque comparable, la colonne d'ajustement contient une formule : VC de la caractéristique du bien à évaluer — VC de la caractéristique du comparable.

### 11.8 Formule du prix de vente ajusté (Adjusted Sale Price)

La ligne **Adjusted Sale Price** contient des formules Excel :

- **Bien à évaluer** : somme de toutes les cellules de contribution à la valeur, depuis BASE VALUE jusqu'à la dernière ligne de caractéristique résiduelle. Elle représente la prédiction totale du modèle pour le bien à évaluer, augmentée des caractéristiques résiduelles ventilées par l'évaluateur.
- **Comparables** : Net SP + somme de toutes les cellules d'ajustement, depuis la première ligne groupée ou de variable jusqu'à la dernière ligne de caractéristique résiduelle. On obtient ainsi le prix de vente du comparable après tous les ajustements issus de la régression et saisis par l'évaluateur.

### 11.9 Protection des feuilles et cellules modifiables

Chaque feuille est **protégée** afin d'empêcher toute modification accidentelle des formules et des données. La protection verrouille :

- Toutes les cellules de formules (Net SP, résidu restant, ajustements, prix de vente ajusté, etc.)
- Toutes les cellules de données (adresses, prix de vente, valeurs factuelles, contributions à la valeur issues du modèle)
- Les libellés et les en-têtes

Les seules cellules **déverrouillées** (modifiables) sont les cellules de saisie des contributions à la valeur des caractéristiques résiduelles — les cellules situées sous la ligne CQA|Residual dans lesquelles l'évaluateur saisit ses ventilations. Elles sont stylisées avec un fond jaune clair pour indiquer qu'elles sont modifiables.

### 11.10 Utilisation de la grille dans Excel

1. **Ouvrir le fichier** dans Excel ou dans un tableur compatible.
2. **Examiner les ajustements de régression** — les ajustements dérivés du modèle sont pré-remplis et verrouillés.

3. **Ventiler le résidu** — dans les lignes de caractéristiques résiduelles (cellules jaunes), saisir votre appréciation de la part du résidu restant attribuable à chaque caractéristique (Location, View, Condition, Quality, etc.). La formule du résidu restant diminue au fur et à mesure de la ventilation.
4. **Vérifier l'Adjusted Sale Price** — cette formule se met à jour automatiquement à mesure que les valeurs des caractéristiques résiduelles sont saisies.
5. **Feuilles multiples** — si plus de 3 comparables ont été sélectionnés, naviguer entre les feuilles. Chaque feuille présente le même bien à évaluer dans la colonne de gauche, avec 3 comparables différents.

***Conseil :** l'objectif est de ventiler le résidu CQA entre les lignes de caractéristiques jusqu'à ce que le résidu restant soit proche de zéro. Cela démontre que l'on peut rendre compte de l'écart entre l'estimation par régression de chaque comparable et son prix de vente effectif.*

## 12 Téléchargement des rapports

Après l'ajustement d'un modèle, la génération de rapports suit un flux de travail Quarto en deux étapes : **Generate Quarto Report** (section 9 de la barre latérale en mode évaluation immobilière, 7 sinon) écrit une archive de rapport autonome, et **Convert Quarto Report** (section 10) en effectue le rendu — ou celui de tout autre fichier Quarto — vers les formats finaux.

### 12.1 Generate Quarto Report

Un clic sur **Generate Quarto Report** écrit une archive Quarto autonome dans le dossier de sortie du projet, à l'emplacement `<base>_qmd/` ; celle-ci contient la source `.qmd` renseignée, tous les graphiques pré-générés (PNG et PDF), les données du rapport et un modèle de référence Word. Aucun rendu n'a lieu à cette étape — l'archive peut être modifiée à la main, combinée à d'autres sources Quarto au moyen des inclusions Quarto, ou prévisualisée de manière autonome avec `quarto preview`.

### 12.2 Convert Quarto Report

La section 10 propose un sélecteur de fichiers (pointant par défaut vers l'archive la plus récemment générée du projet courant — mais on peut naviguer vers n'importe quel fichier `.qmd`) ainsi que des cases à cocher de format :

- **HTML** — un fichier `.html` avec rendu mathématique KaTeX, images intégrées et table des matières. Utilise le thème Bootstrap Flatly et la police Roboto Condensed.
- **Word** — un fichier `.docx` adapté à l'édition et à la diffusion. Comprend une table des matières, la numérotation des pages, et s'appuie sur un modèle de référence personnalisé assurant une mise en forme homogène.
- **PDF** — composé avec LuaLaTeX pour une sortie de qualité professionnelle. Le format du papier suit le paramètre régional (Letter ou A4). Comprend des pages en orientation paysage pour les grandes matrices d'interactions et utilise Roboto Condensed avec Latin Modern Math pour les équations. L'option PDF n'est proposée que lorsqu'une distribution LaTeX est détectée.

Cliquez sur **Convert** pour effectuer le rendu dans les formats sélectionnés. Le rendu s'exécute en arrière-plan — une boîte de dialogue modale affiche le temps écoulé et la progression de Quarto tandis que l'application reste réactive. Une notification confirme l'emplacement du fichier à l'achèvement.

Toutes les opérations (ajustement de modèles, téléchargements de données, calculs RCA, génération de Sales Grids et rendu de rapports) sont consignées, avec horodatages de début et de fin et durées écoulées, dans `<datafile>_earthui_log.txt` au sein du dossier de sortie, à des fins de débogage et de suivi des performances.

### 12.3 Contenu du rapport

Chaque rapport généré comprend les sections suivantes :

1. **Description du jeu de données** — nombre d'observations, variable(s) cible(s), prédicteurs utilisés, prédicteurs catégoriels
2. **Spécification du modèle** — degré, statut de la validation croisée, nombre de termes, prédicteurs présents dans le modèle
3. **Interactions autorisées** (degré  $\geq 2$  uniquement) — la matrice d'interactions avec coches, mise en forme selon le type de sortie

4. **Synthèse des résultats** —  $R^2$ , GRSq, GCV, RSS et CV  $R^2$  (si la validation croisée était activée)
5. **Équation du modèle** — l'équation complète du modèle Earth en notation LaTeX
6. **Coefficients et fonctions de base** — tableau de tous les termes, de leurs coefficients et des fonctions de base qui les définissent
7. **Importance des variables** — diagramme en barres et tableau ordonné des scores d'importance des prédicteurs
8. **Contributions des fonctions g** — graphiques pour chaque fonction g : courbes pour les termes univariés, graphiques en perspective 3D et graphiques de contours pour les interactions bivariées
9. **Matrice de corrélation** — carte thermique des corrélations entre prédicteurs
10. **Diagnostics** — graphiques « Residuals vs Fitted », « Normal Q-Q » et « Actual vs Predicted »
11. **Décomposition ANOVA** — tableau présentant la contribution de chaque fonction de base à la RSS
12. **Sortie Earth** — sortie texte brute de `earth::summary()`, y compris le détail de la passe d'élagage

Pour les modèles à réponses multiples, les sections 4–8 et 10 sont répétées pour chaque variable cible.

***Conseil** : les rapports nécessitent le package R `quarto` et une installation de Quarto. Pour la sortie PDF, une distribution LaTeX doit également être installée (p. ex. via `tinytex::install_tinytex()`) — à défaut, la case à cocher PDF est masquée.*

## 13 Les Post Processing Modules : Elastic Net et GAM

Le package earthUI est livré avec deux Post Processing Modules (modules de post-traitement) — une routine elastic net construite sur `glmnet` et une routine GAM construite sur `mgcv`. Chacune fait l’objet d’un traitement approfondi dans son propre article de ce numéro ; la présente section explique à quoi elles *servent* dans le flux de travail earthUI et comment les trois routines coopèrent.

### 13.1 Pourquoi des ajustements d’appui ?

Le modèle earth produit la conclusion de valeur. Les Post Processing Modules existent pour la **corroborer** : deux estimations supplémentaires du même bien à évaluer, produites par des moteurs méthodologiquement différents, ajustées sur les mêmes données de projet. Lorsqu’une ré-estimation elastic net de la base earth et un lissage GAM indépendant aboutissent à une valeur matériellement identique, la conclusion de l’évaluateur ne repose plus sur les hypothèses d’un seul modèle — un avantage significatif dans les contextes de revue d’évaluation et de contentieux. Lorsqu’elles sont en *désaccord*, ce désaccord est en lui-même un diagnostic, signalant le plus souvent des données clairessemées, un terme trop flexible ou un comparable aberrant méritant d’être réexaminé. Aucun des deux modules ne prime sur le modèle earth ; ils constituent des éléments de preuve à l’appui, délibérément présentés sans pondération.

### 13.2 Lancement et état partagé

Chaque routine est une application Shiny locale distincte :

- `earthUI::launch_glmnet()` — l’application elastic net, sur le port 7879.
- `earthUI::launch_mgcv()` — l’application GAM, sur le port 7880.

Les trois applications lisent la même arborescence de projet et la même base de données de paramètres, de sorte que le projet ouvert dans earthUI est immédiatement disponible dans les modules — de même que les champs à report automatique qu’earthUI enregistre pour lui : la Response (cible), l’Effective Date (date d’effet), le type et la valeur de CQA du bien à évaluer pour la RCA, et la désignation de la surface habitable. Vous ne les ressaisissez pas ; les modules les lisent, si bien que les mêmes valeurs circulent à travers les trois ajustements.

### 13.3 La routine elastic net

La routine `glmnet` importe les fonctions de base du modèle earth — ses charnières, ses interactions et ses termes linéaires — comme colonnes d’une matrice de conception et les ré-estime sous régularisation lasso/elastic-net (voir « `glmnetUI` Import » ci-dessus pour les mécanismes et la mise en garde relative à la colonne de pondération). Le résultat conserve les termes interprétables et géométriquement signifiants d’earth, tout en laissant le régularisateur rétrécir ou éliminer ceux qui ne justifient pas leur maintien. Voir l’article `glmnetUI` de ce numéro pour le traitement complet.

### 13.4 La routine GAM

La routine `mgcv` procède en sens inverse : plutôt que de réutiliser la base d’earth, elle ajuste de nouveaux termes de lissage pénalisés, en prenant les emplacements de nœuds du modèle earth comme points de départ (voir « `mgcvUI` Auto-Export » ci-dessus). Comme le GAM re-dérive les formes fonctionnelles sous un paradigme de lissage différent, sa concordance avec le modèle earth constitue une corroboration véritablement indépendante et non un simple écho. Les modèles earth avec  $\text{degree} \leq 2$  sont pris en charge. Voir l’article `mgcvUI` de ce numéro pour le traitement complet.

### 13.5 Un flux de travail pratique

1. Ajustez et affinez d'abord le **modèle earth**, dans earthUI, jusqu'à ce qu'il vous satisfasse. Chaque ajustement réussi ( $\text{degré} \leq 2$ ) exporte automatiquement le fichier **.rds** que consomment les modules.
2. Ouvrez les routines **glmnet** et **mgcv** sur le même projet, importez le résultat earth et ajustez. Travaillez dans une routine à la fois — les ajustements sont de longue durée et les routines se coordonnent via le projet partagé, si bien que le rythme raisonnable est séquentiel, et non parallèle.
3. Itérez librement. Chaque application conserve ses résultats pendant que vous passez de l'une à l'autre, et rien ne vous oblige à générer des rapports en cours de route.
4. Ce n'est que lorsque les trois résultats vous satisfont que vous revenez dans chaque routine pour générer son rapport Quarto. Chaque méthode enregistre son ajustement et sa conclusion de valeur dans le registre de trilogie du projet, ce qui maintient la traçabilité des trois rapports vers une même exécution comparative.

***Conseil :** les modules sont facultatifs à tous les niveaux : le flux de travail propre d'earthUI — ajustement, RCA, Sales Grids, rapports — est complet sans eux. Recourez-y lorsqu'une conclusion mérite corroboration : un bien à évaluer difficile, un segment de marché étroit, ou une mission susceptible de faire l'objet d'une revue.*

## 14 Jeu de données de démonstration : Appraisal\_1.csv

earthUI inclut un jeu de données MLS de démonstration permettant d'explorer le flux de travail d'évaluation immobilière. On le charge par programmation avec :

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

Ou bien, dans l'application Shiny, copiez-le dans le dossier `in/` d'un projet (ou joignez-le comme fichier de données initial lors de la création du projet) et sélectionnez-le dans la Section 2 (Import Data — importation des données).

### 14.1 Description

Le fichier contient 1,501 ventes résidentielles plus 1 bien à évaluer en ligne 1 (soit 1,502 lignes au total), issues d'un export MLS simulé. Les données représentent des ventes de maisons individuelles dans un marché à plusieurs zones, avec une gamme variée de surfaces, d'âges et de localisations.

Il ne s'agit pas de données réelles, mais elles s'appuient sur un quartier réaliste de Californie du Nord. Toutes les informations d'identification ont été modifiées ou supprimées.

## 14.2 Colonnes

| Colonne             | Type      | Type spécial  | Description                                                              |
|---------------------|-----------|---------------|--------------------------------------------------------------------------|
| weight              | numeric   | —             | Poids de l'observation (0 = exclue de l'ajustement)                      |
| id                  | numeric   | display_only  | Identifiant interne de l'enregistrement                                  |
| property_id         | numeric   | display_only  | Identifiant MLS du bien                                                  |
| listing_id          | character | display_only  | Numéro de listing MLS                                                    |
| parcel_number       | character | display_only  | Numéro de parcelle du cadastre du comté (APN)                            |
| street_address      | character | display_only  | Adresse du bien                                                          |
| city_name           | character | display_only  | Ville                                                                    |
| postal_code         | character | display_only  | Code postal (ZIP)                                                        |
| county_name         | character | display_only  | Comté                                                                    |
| contract_date       | Date      | contract_date | Date du contrat de vente (sert au calcul de sale_age)                    |
| sale_age            | numeric   | —             | Nombre de jours entre la date du contrat et la date d'effet (précalculé) |
| coe_date            | Date      | display_only  | Date de clôture de l'entiercement (close of escrow)                      |
| listing_status      | character | display_only  | Statut du listing (p. ex. "Sold")                                        |
| sale_price          | numeric   | (target)      | Prix de vente — variable réponse                                         |
| rent                | numeric   | —             | Loyer mensuel (pour les modèles à cibles multiples)                      |
| list_price          | numeric   | display_only  | Prix affiché du listing                                                  |
| original_list_price | numeric   | display_only  | Prix affiché initial                                                     |
| living_sqft         | numeric   | living_area   | Surface habitable brute en pieds carrés                                  |
| beds_total          | integer   | —             | Nombre de chambres                                                       |
| baths_total         | numeric   | —             | Nombre total de salles de bains (p. ex. 2.5 = 2 complètes + 1 demi)      |
| lot_size            | numeric   | lot_size      | Surface du terrain en pieds carrés                                       |
| area_id             | integer   | area          | Identifiant de zone MLS                                                  |
| area_text           | character | display_only  | Nom de la zone                                                           |
| age                 | numeric   | actual_age    | Âge du bien en années                                                    |
| year_built          | integer   | display_only  | Année de construction                                                    |
| latitude            | numeric   | latitude      | Latitude (arrondie à 3 décimales pour le modèle)                         |
| longitude           | numeric   | longitude     | Longitude (arrondie à 3 décimales pour le modèle)                        |
| latitude6           | numeric   | display_only  | Latitude en pleine précision                                             |
| longitude6          | numeric   | display_only  | Longitude en pleine précision                                            |
| garage_spaces       | integer   | —             | Nombre de places de garage                                               |
| fp_count            | integer   | —             | Nombre de cheminées                                                      |
| no_of_stories       | numeric   | —             | Nombre de niveaux                                                        |
| style               | character | —             | Style architectural                                                      |
| view                | character | —             | Type de vue (p. ex. "Neighborhood", "Hills")                             |
| days_on_market      | integer   | dom           | Nombre de jours sur le marché                                            |
| listing_date        | Date      | listing_date  | Date de mise en vente                                                    |
| sale_concessions    | numeric   | concessions   | Concessions du vendeur                                                   |

## 14.3 Démarrage rapide suggéré

1. Lancer earthUI : `earthUI::launch()`
2. Créer un projet dont le champ Purpose (finalité) est réglé sur **For Appraisal** dans l'assistant New Project, en joignant `Appraisal_1.csv` comme fichier de données initial
3. Sélectionner `Appraisal_1.csv` dans la Section 2 (Import Data)
4. Sélectionner `sale_price` comme cible
5. Attribuer les types spéciaux indiqués dans le tableau ci-dessus
6. Inclure les prédicteurs : `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (comme facteur), `age`, `latitude`, `longitude`, `garage_spaces`
7. Fixer le degré à 1, puis cliquer sur **Fit Earth Model**
8. Télécharger la sortie intermédiaire (étape 6) et examiner le classement CQA
9. Calculer les ajustements RCA (étape 7) avec un score CQA de ~5.00
10. Générer la Sales Comparison Grid (grille de comparaison des ventes) (étape 8)

**Conseil :** le jeu de données de démonstration constitue un bon moyen d'explorer toutes les fonctionnalités d'earthUI avant d'importer vos propres données MLS. Les prédicteurs suggérés ci-dessus produisent généralement un modèle dont le  $R^2$  dépasse 0.85.

## 15 Fonctionnalité expérimentale : extraction de texte en Prolog et règles de dérivation

La version 0.10.0 introduit un dispositif **optionnel et authentiquement expérimental** d'extraction de caractéristiques prêtes pour le modèle à partir des colonnes de texte descriptif d'un export MLS — la colonne « Public Remarks » avant tout, mais aussi les remarques d'agent et les listes de caractéristiques délimitées par des virgules, telles que les équipements des pièces ou les constructions existantes. Le moteur d'extraction est Prolog, et plus précisément les **grammaires à clauses définies (DCG, Definite Clause Grammars)** : les règles qui reconnaissent des expressions comme « move in ready », « 3 car garage » ou « owned solar » dans le texte des remarques sont écrites sous forme de règles de grammaire Prolog.

Cette fonctionnalité est encore en développement. Elle peut être utilisée *partiellement* pour l'extraction dès aujourd'hui — la grammaire fournie reconnaît déjà un ensemble utile de caractéristiques — mais son adaptation au vocabulaire de votre propre MLS exige de savoir lire et écrire du Prolog, et des DCG en particulier. Si ce n'est pas votre cas, ignorez entièrement ce chapitre : la fonctionnalité est désactivée par défaut, et le reste d'earthUI fonctionne exactement de la même manière sans elle.

### Avertissement

Il s'agit d'une fonctionnalité expérimentale en cours de développement actif. Sa grammaire d'extraction n'est ni complète ni validée pour un travail d'évaluation immobilière en production — les colonnes extraites doivent être vérifiées par rapport au texte source avant d'être utilisées dans un modèle. Un usage efficace de cette fonctionnalité requiert une connaissance pratique de Prolog et des DCG. Tout ce qu'elle fait est additif et réversible ; en cas de comportement anormal, désactivez l'option dans Settings et poursuivez normalement.

### 15.1 Ce qu'elle fait

Les listings MLS véhiculent une grande quantité d'informations pertinentes pour la valeur dans du texte libre et dans des champs de listes délimitées. Cette étape optionnelle transforme ce texte en colonnes ordinaires exploitables par votre régression, via deux rôles de colonne spéciaux (chapitre 6) :

- les **colonnes remarks** sont analysées par la grammaire DCG Prolog, qui produit une colonne de caractéristique par attribut extrait. Le préfixe de colonne est construit à partir des premières lettres du nom de la colonne source : **public\_remarks** donne des colonnes **pr\_\***, **agent\_remark** donne **ar\_\***, et ainsi de suite.
- les **colonnes list** (p. ex. **existing\_structures = "Shed, Workshop"**) sont scindées en colonnes TRUE/FALSE à encodage one-hot **<col>\_<value>**. Cette partie est du R pur et ne requiert aucun Prolog.

La grammaire fournie (**inst/prolog/mls\_remarks.pl**) extrait actuellement, entre autres : un niveau d'état gradué avec un **condition\_score** de 1–5 ; des indicateurs de vue et des types de vue ; des indicateurs de rénovation de cuisine et de salle de bains ; le parquet et d'autres revêtements de sol ; une liste et un décompte des améliorations ; les places et la surface de garage ; piscine, spa et autres aménagements extérieurs ; le nombre de cheminées ; les dépendances (shed, barn, workshop, casita, guest house, kennel, etc.) avec leurs tailles analysées à partir de dimensions telles que « 40x50 » ; le stationnement et le remisage de camping-cars (RV) ; la propriété ou la location des panneaux solaires, le stockage par batterie ; la surface du terrain (y compris la conversion acres-pieds carrés) ; des montants de prix liés au contexte de vente rattachés à

un sujet (réduction de prix, cotisations HOA, location ou rachat du solaire, Mello-Roos) ; des indices de localisation ; et un score de sentiment points-forts/points-faibles dans  $[-1, 1]$ .

## 15.2 Prérequis : vProlog (Trealla), et non SWI-Prolog

L'analyse des remarques s'exécute sur le package R **vProlog** — un moteur Prolog Trealla autonome. Il n'est pas disponible sur CRAN ; installez-le depuis le dépôt r-universe de l'auteur comme indiqué dans « Installation d'earthUI et de vProlog » au début de cet article (binaires précompilés pour Windows et macOS, sources à [github.com/wcraytor/vProlog](https://github.com/wcraytor/vProlog)). Aucune installation système de Prolog (SWI-Prolog ou autre) n'est requise ni utilisée. Si vProlog est absent, l'analyse des remarques est ignorée avec un message, tandis que la scission des colonnes `list` s'exécute quand même. La fonction utilitaire `vProlog_available()` indique si le moteur et la grammaire fournie sont tous deux présents.

## 15.3 Activation et exécution

1. Ouvrez **Settings** (engrenage) et cochez **Enable Prolog / list column processing** (activer le traitement Prolog / des colonnes de liste). L'option est désactivée par défaut, porte la mention « advanced feature — for users with Prolog expertise » (fonctionnalité avancée — pour les utilisateurs maîtrisant Prolog), est mémorisée par projet et apparaît dans la boîte de dialogue Project Information.
2. Dans le tableau des prédicteurs, attribuez le rôle **Special remarks** ou `list` aux colonnes concernées (plusieurs colonnes peuvent porter chacun de ces rôles).
3. Utilisez la section **Execute Prolog Processing** (exécuter le traitement Prolog), située entre les Sections 4 et 5 de la barre latérale. Elle propose deux étapes :

**1. Expand lists & remarks** — exécute la grammaire DCG sur chaque colonne `remarks` et scinde en one-hot chaque colonne `list`. L'expansion est *incrémentale* : seules les colonnes désignées depuis la dernière exécution sont traitées, et le bouton est désactivé dès que tout ce qui a été désigné a déjà été développé. Une barre de progression suit l'analyse.

**2. Execute derivation rules** — applique vos propres règles Prolog `derive/2` aux colonnes développées (voir ci-dessous). Les règles sont vérifiées syntaxiquement et validées par une exécution à blanc avant leur application.

## 15.4 Règles de dérivation

Pour chaque ligne de données, earthUI affirme les colonnes de la ligne sous forme de faits `cell(Name, Value)` et collecte toutes les solutions de vos règles `derive(Col, Val)` — chaque `Col` distinct devient une nouvelle colonne, et la *première* valeur trouvée par colonne l'emporte, ce qui autorise une clause de repli à suivre. Des prédicats utilitaires (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) facilitent l'agrégation et le traitement des NA, et des faits `drop/1` suppriment des colonnes après l'exécution des règles. Par exemple :

```
% 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).
```

Les règles résident dans un fichier `<project>_rules.pl` propre à chaque projet, initialisé à partir d'un gabarit entièrement commenté, et s'éditent via **Edit derivation rules...** (éditer les règles de dérivation) dans un éditeur de code flottant et déplaçable, qui peut aussi être ancré dans la barre latérale. Un bouton **Check** (vérifier) valide les règles (analyse syntaxique, vérification du chargement et exécution à blanc sur la ligne 1 de vos données) sans les exécuter.

## 15.5 Fichier de données actif versionné

Après l'une ou l'autre étape, le data frame augmenté est écrit dans le dossier d'entrée du projet sous le nom `<original_name>_<YYMMDD_HHMMSS>.xlsx` et devient le fichier de données **actif** du projet — la réouverture du projet charge les données augmentées, de sorte que les expansions ne sont jamais recalculées, et c'est ce même fichier que consomment glmnetUI et mgcvUI. Les colonnes sources de remarques brutes sont automatiquement écartées du modèle ; les colonnes générées ne sont *pas* incluses automatiquement comme prédicteurs. L'analyse de texte peut produire de nombreuses colonnes ; marquez donc celles que vous ne voulez pas (encore) dans la régression comme **Latent**, ou désactivez-les via  $\times$  (chapitre 6).

## 15.6 Interface programmatique et fichiers inclus

L'ensemble du pipeline est exporté pour un usage en traitement par lots : `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` orchestre les passes, en s'appuyant sur `remarks_to_columns()`, `split_list_column()`, `apply_user_rules()` et `validate_rules()`. Les sources Prolog fournies sous `inst/prolog/` sont `mls_remarks.pl` (la grammaire DCG et l'agrégation des caractéristiques), `mls_analyze.pl` (un pilote d'entrées-sorties de fichiers pour l'analyse par lots), `mls_rules.pl` (les prédicats utilitaires accessibles aux règles de dérivation) et `climate_regions_ca.pl` (ci-dessous).

## 15.7 Expérimentation connexe : a priori par région climatique

Un dispositif expérimental complémentaire traite les caractéristiques à *données clairsemées* — celles que la régression ne peut pas estimer parce qu'elles n'apparaissent que dans une poignée de biens comparables (un refroidisseur évaporatif, par exemple). Trois fonctions utilitaires R — `climate_region_for()`, `market_area_profile()` et `climate_feature_priors()` — classent la localisation d'un projet (comté + ville) dans l'une des sept régions climatiques californiennes reconnues par les acheteurs (une réduction pratique des 16 zones climatiques du bâtiment de la California Energy Commission) et renvoient des a priori ordinaux de valeur contributive pour les équipements de chauffage, ventilation et climatisation (chauffage, climatisation, refroidisseur évaporatif, ventilateur de maison entière, filtration) par région. Le module `Prolog climate_regions_ca.pl` reproduit la même classification pour un usage au sein des règles de dérivation ; les deux versants sont générés à partir d'une source R canonique unique, de sorte qu'ils ne peuvent pas diverger. Ces a priori sont destinés à *éclairer* l'évaluation des caractéristiques et à servir de contrôle de plausibilité des coefficients estimés — les affectations de région (en particulier pour les comtés s'étendant sur plusieurs régions) doivent être revues par l'évaluateur.

### 15.7.1 Au-delà de la Californie : packs juridictionnels et modules de zone de marché

La Californie est la première juridiction, non la frontière. earthUI est utilisé à l'international, et la connaissance des a priori climatiques est en cours de restructuration en **packs juridictionnels**

(« jurisdiction packs ») : des tables de faits autonomes, de forme purement déclarative (définitions de zones, affectations d'unités administratives, a priori de caractéristiques), indexées sur les codes de pays et de niveaux administratifs du système de projet, de sorte qu'un pack peut décrire un État américain, un Land allemand ou un petit pays tout entier. Une couche complémentaire de **modules de zone de marché** (« market-area modules ») capture le vocabulaire local et des a priori affinés pour une zone de marché nommée.

Les deux sont conçus pour être rédigés *sur le terrain* par des Valuation Engineers et des évaluateurs, dans leur propre langue — et, de façon réaliste, avec l'assistance d'un LLM : le monde ne compte pas beaucoup de programmeurs Prolog, aussi les artefacts de terrain sont-ils délibérément des tables plutôt que des programmes. Le format de pack est un schéma documenté accompagné d'un exemple complet (la Californie elle-même), expressément pour qu'un assistant tel que Claude puisse rédiger un pack valide à partir d'une description en langage courant d'un marché ; earthUI génère le Prolog sous-jacent à partir du pack, et un validateur bloquant (contrôles de schéma, de couverture, d'intégrité référentielle et d'exécution à blanc) doit être satisfait avant qu'un pack puisse être activé. Chaque pack rédigé sur le terrain porte une provenance obligatoire — l'auteur en titre, la date et l'autorité citée — de sorte que c'est le professionnel responsable, et non l'outil de rédaction, qui répond de la connaissance, et que les rapports peuvent citer le pack exact et la version utilisée.

**Conseil** : un schéma pratique : développez les remarques, consolidez les indicateurs granulaires *pr\_\** en une poignée de décomptes ou de booléens dérivés au moyen de règles *derive/2*, supprimez avec *drop/1* les indicateurs bruts ainsi consolidés, et ne conservez comme prédicteurs candidats que les colonnes dérivées. La liste des prédicteurs reste ainsi courte et la colinéarité maîtrisable.

## Références

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

# glmnetUI : une interface interactive pour le package glmnet (Lasso et Elastic Net)

*Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

William Bert Craytor

Volume 1, Numéro 1 · Été 2026

Publié le 1<sup>er</sup> juillet 2026

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

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

Cet article en libre accès est diffusé selon les termes de la [licence Creative Commons Attribution 4.0](#).

**Résumé.** glmnetUI est une interface graphique pour le package R glmnet, qui ajuste des modèles linéaires généralisés régularisés au moyen des pénalités Lasso, ridge et elastic-net. Elle propose trois modes d'utilisation — modélisation prédictive générale, évaluation immobilière et analyse de zone de marché — et guide l'utilisateur à travers l'importation des données, la configuration du modèle, l'ajustement par validation croisée, les chemins de régularisation des coefficients et les rapports téléchargeables. Cet article documente les exigences de format de données de glmnetUI, le flux de travail de modélisation, les affichages de sortie et la référence complète des fonctionnalités.

**Mots-clés :** Lasso, régression ridge, elastic net, régression régularisée, glmnet, R, interface graphique, évaluation immobilière

**Note de l'éditeur (juillet 2026) :** Peu avant la mise sous presse, glmnetUI a été intégré à earthUI (version 0.11.0 et ultérieures), qui fournit désormais le flux de travail elastic-net décrit ici aux côtés des flux de travail MARS et GAM. glmnetUI demeure disponible dans sa dernière version publiée (0.5.0) à l'adresse <https://github.com/wcraytor/glmnetUI> ; les développements futurs se poursuivent dans earthUI. Le présent article s'applique à glmnetUI tel que publié.

## Table des matières

|          |                                                     |          |
|----------|-----------------------------------------------------|----------|
| <b>1</b> | <b>Introduction</b>                                 | <b>3</b> |
| 1.1      | Historique et contexte                              | 3        |
| 1.1.1    | Le Lasso                                            | 3        |
| 1.1.2    | Régression ridge et elastic net                     | 4        |
| 1.1.3    | Le package glmnet                                   | 4        |
| 1.1.4    | Comparaison avec MARS (earth) et GAM (mgcv)         | 4        |
| 1.1.5    | Flux de travail recommandé                          | 5        |
| 1.2      | Qu'est-ce que glmnetUI ?                            | 6        |
| 1.3      | Trois modes Purpose                                 | 6        |
| 1.4      | Fonctionnalités propres à l'immobilier              | 7        |
| 1.5      | Premiers pas                                        | 7        |
| <b>2</b> | <b>Exigences relatives aux données d'entrée MLS</b> | <b>9</b> |
| 2.1      | Format et structure du fichier                      | 9        |
| 2.2      | Colonnes requises en mode évaluation                | 9        |
| 2.3      | Conventions de nommage des colonnes spéciales       | 10       |
| 2.4      | Qualité et complétude des données                   | 10       |
| 2.5      | Emplacement de la ligne du bien à évaluer           | 11       |

|          |                                                    |           |
|----------|----------------------------------------------------|-----------|
| <b>3</b> | <b>Mode General Purpose</b>                        | <b>12</b> |
| 3.1      | Vue d'ensemble                                     | 12        |
| 3.1.1    | Skip First Row                                     | 12        |
| 3.1.2    | Paramètres propres à chaque finalité               | 12        |
| 3.2      | Le flux de travail de la barre latérale            | 12        |
| 3.3      | Projets                                            | 13        |
| 3.4      | Importer depuis earthUI (facultatif)               | 14        |
| 3.5      | Onglets du panneau principal                       | 14        |
| 3.6      | Persistance des paramètres                         | 15        |
| 3.7      | Mode sombre                                        | 15        |
| 3.8      | Locale et paramètres régionaux                     | 15        |
| 3.8.1    | Enregistrer les valeurs par défaut                 | 15        |
| <b>4</b> | <b>Mode évaluation</b>                             | <b>16</b> |
| 4.1      | Traitement de la ligne du bien à évaluer           | 16        |
| 4.2      | Date d'effet et ancienneté de la vente             | 16        |
| 4.3      | Désignations spéciales de colonnes                 | 16        |
| 4.4      | Aperçu des ajustements RCA                         | 16        |
| <b>5</b> | <b>Mode analyse de zone de marché</b>              | <b>17</b> |
| 5.1      | Différences avec le mode évaluation                | 17        |
| 5.2      | Quand utiliser le mode marché                      | 17        |
| <b>6</b> | <b>Sélection des variables</b>                     | <b>18</b> |
| 6.1      | Variable cible                                     | 18        |
| 6.2      | Le tableau des prédicteurs                         | 18        |
| 6.3      | Détection des types de données et substitutions    | 18        |
| 6.4      | Signes attendus et application                     | 18        |
| 6.5      | Référence des types de colonnes spéciales          | 19        |
| <b>7</b> | <b>Choix des paramètres</b>                        | <b>20</b> |
| 7.1      | Alpha (paramètre de mélange)                       | 20        |
| 7.2      | Choix de lambda                                    | 20        |
| 7.3      | Famille                                            | 20        |
| 7.4      | Standardize                                        | 21        |
| 7.5      | Contraintes de signe                               | 21        |
| 7.6      | Relaxed lasso                                      | 21        |
| 7.7      | Graine aléatoire                                   | 21        |
| 7.8      | Matrice d'interactions                             | 22        |
| 7.9      | Exclusion de l'effet principal (interaction seule) | 22        |
| 7.10     | Paramètres avancés                                 | 22        |
| 7.11     | Valeurs par défaut des paramètres                  | 23        |
| <b>8</b> | <b>Ajustement du modèle</b>                        | <b>24</b> |
| 8.1      | Le bouton d'ajustement                             | 24        |
| 8.2      | Ce qui se passe pendant l'ajustement               | 24        |
| <b>9</b> | <b>Onglets de résultats</b>                        | <b>25</b> |
| 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        |
| <b>10</b> | <b>Téléchargement des données</b>                        | <b>28</b> |
| 10.1      | Colonnes de sortie                                       | 28        |
| 10.2      | Scores CQA                                               | 28        |
| <b>11</b> | <b>Calculs RCA et téléchargement</b>                     | <b>29</b> |
| 11.1      | Ouverture de la boîte de dialogue RCA                    | 29        |
| 11.2      | Interpolation du score CQA                               | 29        |
| 11.3      | Colonnes de sortie                                       | 29        |
| <b>12</b> | <b>Sales Comparison Grid</b>                             | <b>30</b> |
| 12.1      | Boîte de dialogue de sélection des comparables           | 30        |
| 12.2      | Structure du classeur                                    | 30        |
| 12.3      | Protection des feuilles et cellules résiduelles          | 30        |
| <b>13</b> | <b>Téléchargement des rapports</b>                       | <b>31</b> |
| 13.1      | Le flux de travail Quarto en deux étapes                 | 31        |
| 13.2      | Formats de rapport                                       | 31        |
| 13.3      | Contenu des rapports                                     | 31        |
| <b>14</b> | <b>Comparaison avec earthUI</b>                          | <b>33</b> |
| 14.1      | Principales différences                                  | 33        |
| 14.2      | Fonctionnalités communes                                 | 33        |
| 14.3      | Lequel utiliser, et quand                                | 33        |
| <b>15</b> | <b>Jeu de données de démonstration : Appraisal_1.csv</b> | <b>35</b> |
| 15.1      | Description                                              | 35        |
| 15.2      | Colonnes                                                 | 35        |
| 15.3      | Prise en main rapide suggérée                            | 36        |
| <b>16</b> | <b>Configuration requise et résolution des problèmes</b> | <b>37</b> |
| 16.1      | Plateformes prises en charge                             | 37        |
| 16.2      | Dépendances facultatives                                 | 37        |
| 16.3      | Remarques par plateforme                                 | 37        |
| 16.4      | Dégradation progressive                                  | 37        |
| 16.5      | Résolution des problèmes                                 | 37        |

## 1 Introduction

### 1.1 Historique et contexte

#### 1.1.1 Le Lasso

Le **Lasso** (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator, soit « opérateur de rétrécissement et de sélection par valeur absolue ») a été introduit par Robert Tibshirani en 1996. Le nom

est un jeu de mots délibéré sur le lasso du cow-boy — une corde qui contraint et qui sélectionne. Tibshirani l’a choisi parce que la méthode « lasso » littéralement les coefficients : elle contraint leur valeur absolue et en ramène certains exactement à zéro, sélectionnant ainsi les variables qui demeurent dans le modèle.

L’intuition de Tibshirani a consisté à ajouter une pénalité  $L_1$  (somme des valeurs absolues des coefficients) au critère des moindres carrés. Contrairement à la pénalité  $L_2$  de la régression ridge (Hoerl et Kennard, 1970), qui rétrécit les coefficients vers zéro sans jamais l’atteindre, la pénalité  $L_1$  produit des solutions *parcimonieuses* dans lesquelles de nombreux coefficients sont exactement nuls. Le Lasso est donc à la fois une méthode de régularisation et une méthode de sélection de variables.

### 1.1.2 Régression ridge et elastic net

La **régression ridge** a été proposée par Arthur Hoerl et Robert Kennard en 1970. Elle ajoute une pénalité  $L_2$  (somme des carrés des coefficients) qui rétrécit tous les coefficients proportionnellement mais conserve toutes les variables dans le modèle. Ridge est efficace lorsque les prédicteurs sont corrélés (multicolinéarité), mais elle ne simplifie pas le modèle.

L’**elastic net**, introduit par Hui Zou et Trevor Hastie en 2005, combine les deux pénalités. Le paramètre de mélange  $\alpha$  contrôle le dosage :  $\alpha = 1$  correspond au Lasso pur,  $\alpha = 0$  à la régression ridge pure, et les valeurs intermédiaires donnent un compromis. L’elastic net surmonte une limite du Lasso : lorsque les prédicteurs sont fortement corrélés, le Lasso tend à en sélectionner un seul et à ignorer les autres, tandis que l’elastic net regroupe les prédicteurs corrélés.

### 1.1.3 Le package glmnet

Le package R `glmnet` a été développé par Jerome Friedman, Trevor Hastie et Robert Tibshirani à l’université Stanford. Il met en œuvre la régularisation elastic net pour les modèles linéaires généralisés au moyen d’un algorithme de descente par coordonnées extrêmement efficace (FRIEDMAN, HASTIE et TIBSHIRANI 2010), étendu depuis aux modèles à risques proportionnels de Cox (SIMON et al. 2011) et à toutes les familles de modèles linéaires généralisés (TAY, NARASIMHAN et HASTIE 2023). Le package prend en charge les familles gaussian (linéaire), binomial (logistique), poisson et d’autres familles GLM.

Parmi les fonctionnalités clés de `glmnet` :

- **Chemin de régularisation** : ajuste le modèle sur toute une grille de valeurs de  $\lambda$  en une seule passe, ce qui est bien plus rapide qu’un ajustement séparé pour chaque  $\lambda$ .
- **Validation croisée** : `cv.glmnet()` sélectionne automatiquement l’intensité optimale de régularisation par validation croisée à  $k$  blocs.
- **Bornes sur les coefficients** : prise en charge de limites supérieures et inférieures sur les coefficients individuels (contraintes de signe).
- **Relaxed lasso** (lasso relâché) : réestime le modèle sur les variables sélectionnées avec une pénalisation réduite ou nulle, ce qui diminue le biais.

### 1.1.4 Comparaison avec MARS (earth) et GAM (mgcv)

L’application `earth` et ses deux Post Processing Modules (modules de post-traitement) `glmnet` et `mgcv` reposent sur des moteurs de modélisation différents. Le tableau suivant résume les principales différences :

| Caractéristique           | glmnet (elastic net)                                                     | earth (MARS)                                                                               | mgcv (GAM)                                                                      |
|---------------------------|--------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Type de modèle            | Linéaire (avec régularisation)                                           | Linéaire par morceaux (splines adaptatives)                                                | Non linéaire lisse (splines pénalisées)                                         |
| Non-linéarité             | Uniquement par interactions ou expansion de base                         | Automatique, par fonctions charnières à nœuds déduits des données                          | Automatique, par fonctions de lissage (thin plate, cubique, etc.)               |
| Sélection de variables    | Intégrée, via la pénalité $L_1$ (Lasso)                                  | Intégrée, par sélection pas à pas avant/arrière avec élagage GCV                           | Facultative, par double pénalité de rétrécissement ( <code>select=TRUE</code> ) |
| Interactions              | Produits croisés deux à deux (difficiles à interpréter)                  | Produits de charnières (interprétables, jusqu'au degré 3)                                  | Lissages par produit tensoriel ( <code>te()</code> , <code>ti()</code> )        |
| Interprétabilité          | Les coefficients sont des poids linéaires, faciles à expliquer un par un | Les g-fonctions montrent des effets partiels linéaires par morceaux, variable par variable | Courbes d'effets partiels lisses ; très intuitives                              |
| Domaine de prédilection   | Données de grande dimension, sélection de variables, petits échantillons | Détection automatique des non-linéarités et des interactions                               | Relations non linéaires lisses, modélisation souple                             |
| Ajustements RCA           | Ajustements linéaires par variable                                       | Ajustements linéaires par morceaux via les g-fonctions                                     | Ajustements lisses via les effets partiels                                      |
| Contrôle du surajustement | Lambda par validation croisée                                            | Élagage fondé sur le GCV                                                                   | Estimation du paramètre de lissage par REML/GCV                                 |

**MARS** (Multivariate Adaptive Regression Splines) a été introduit par Jerome Friedman en 1991. La méthode construit des modèles linéaires par morceaux en sélectionnant de manière adaptative des fonctions charnières  $\max(0, x - k)$  et la position de leurs nœuds à partir des données. L'implémentation en R est le package **earth** de Stephen Milborrow. MARS excelle à découvrir automatiquement les relations non linéaires et les interactions sans que l'utilisateur ait à les spécifier. Ses g-fonctions (termes de base regroupés par variable) fournissent des courbes d'effets partiels hautement interprétables.

Les **GAM** (Generalized Additive Models, modèles additifs généralisés) ont été formalisés par Hastie et Tibshirani (1990). L'implémentation en R, **mgcv** de Simon Wood, propose des splines de régression pénalisées avec sélection automatique du degré de lissage par REML ou GCV. Les GAM produisent des courbes d'effets partiels lisses, intuitives à visualiser et à expliquer. Lorsque **earthUI** exporte les positions de nœuds vers **mgcvUI**, les nœuds issus d'**earth** servent de points de départ aux termes de lissage du GAM, combinant le placement adaptatif des nœuds propre à MARS et l'estimation lisse propre aux GAM.

### 1.1.5 Flux de travail recommandé

Pour l'évaluation immobilière et les applications analogues qui exigent des modèles interprétables et défendables :

1. **Commencer par earthUI** pour découvrir les variables importantes, les relations non linéaires et les interactions présentes dans les données. La sélection automatique de base

d'earth et ses g-fonctions fournissent d'excellents éclairages initiaux.

2. **Affiner avec mgcvUI** si des effets partiels plus lisses sont souhaités. L'importation des nœuds d'earthUI dans mgcvUI assure une transition sans rupture des modèles linéaires par morceaux vers les modèles lisses.
3. **Utiliser glmnetUI** lorsqu'une sélection de variables agressive (Lasso) est nécessaire, lorsque le nombre de prédicteurs approche ou dépasse le nombre d'observations, ou lorsqu'on souhaite un modèle purement linéaire avec régularisation, pour des raisons de défendabilité.

## 1.2 Qu'est-ce que glmnetUI ?

glmnetUI est une interface graphique pour le package R **glmnet**. Elle s'exécute sous la forme d'une application Shiny locale — sans identifiant, sans serveur et sans compte. On la lance depuis R, on importe un jeu de données (CSV ou Excel), on configure le modèle et on l'ajuste de manière interactive.

Depuis la version actuelle, glmnetUI n'est plus un package installé séparément : il est livré à l'intérieur du package earthUI, dont il constitue le Post Processing Module elastic net, lancé par `earthUI::launch_glmnet()`. L'application elle-même — et tout ce qui est décrit dans le présent article — demeure inchangée ; seules l'installation et la commande de lancement diffèrent. Les trois routines (earth, glmnet, mgcv) partagent une même arborescence de projets, une même base de données de paramètres et une même copie de l'infrastructure de support.

L'application offre un flux de travail complet : importation des données, configuration des variables, ajustement du modèle, graphiques de diagnostic, importance des variables, équations du modèle et rapports téléchargeables aux formats Word, PDF ou HTML.

La **régression elastic net** combine deux techniques de régularisation :

- **Régression ridge** ( $\alpha = 0$ ) : rétrécit proportionnellement tous les coefficients vers zéro. Adaptée lorsqu'on estime que toutes les variables sont pertinentes.
- **Régression Lasso** ( $\alpha = 1$ ) : peut ramener des coefficients exactement à zéro, réalisant une sélection automatique de variables. Produit des modèles plus simples et plus interprétables.
- **Elastic net** ( $0 < \alpha < 1$ ) : un mélange des deux approches. Utile lorsque les prédicteurs sont corrélés.

## 1.3 Trois modes Purpose

Chaque projet glmnetUI possède un **Purpose** (finalité), choisi à sa création (section 1, *Project*). La finalité du projet actif détermine les outils et les éléments d'interface qui apparaissent ; pour travailler dans un autre mode, il faut créer ou ouvrir un projet doté de cette finalité. Il existe trois modes :

- **General** (général) — régression elastic net pour tout type de population ou de jeu de données. Il s'agit du mode par défaut. Il offre le flux de travail complet de régression régularisée, sans aucun ajout propre à un domaine particulier.
- **For Appraisal** (pour l'évaluation immobilière) — régression elastic net adaptée à l'évaluation immobilière. Ce mode ajoute des fonctionnalités propres à l'évaluation d'un bien individuel, notamment la prise en charge du bien à évaluer, les désignations spéciales de colonnes et la Residual Constraint Approach (RCA) (fr. : approche par contraintes résiduelles).
- **Market Area Analysis** (analyse de zone de marché) — régression elastic net adaptée aux études de zone de marché. Ce mode ajoute des fonctionnalités permettant d'analyser des groupes de biens dans un marché défini, avec exclusion facultative de la ligne du bien à évaluer.

Dans les trois modes, le moteur de modélisation central est identique — on ajuste toujours un modèle elastic net (glmnet). Le réglage Purpose détermine uniquement les outils et les éléments d'interface supplémentaires disponibles.

## 1.4 Fonctionnalités propres à l'immobilier

Lorsque **For Appraisal** ou **Market Area Analysis** est sélectionné, glmnetUI active plusieurs fonctionnalités conçues pour l'analyse immobilière :

- **Désignations spéciales de colonnes** — chaque prédicteur peut être doté d'un rôle spécial tel que `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Ces désignations déterminent le traitement de la colonne lors de l'ajustement et dans les sorties.
- **Arrondi de la latitude et de la longitude** — les colonnes désignées comme latitude ou longitude sont automatiquement arrondies à 3 décimales afin d'éviter le surajustement.
- **Colonne Sale Age** (ancienneté de la vente) — lorsqu'une colonne est désignée comme `contract_date` et qu'une Effective Date (date d'effet) est fournie, glmnetUI calcule une colonne `sale_age` (nombre de jours entre la date de vente et la date d'effet) et la substitue comme prédicteur.
- **Calculs RCA (évaluation uniquement)** — en mode évaluation, après l'ajustement du modèle, glmnetUI peut produire la sortie Residual Constraint Approach (RCA), qui fournit des ajustements par comparable, des synthèses d'ajustements nets et bruts, ainsi qu'un prix de vente ajusté pour le bien à évaluer.

***Conseil :** le mode General convient à tout jeu de données — financier, scientifique, technique ou immobilier. Les modes évaluation et zone de marché ne font qu'ajouter des fonctionnalités de commodité destinées aux professionnels de l'immobilier.*

## 1.5 Premiers pas

Pour utiliser glmnetUI :

1. **Installer earthUI** dans R : `install.packages("earthUI")` — la routine glmnet est incluse. (Voir « Installing earthUI and vProlog » dans l'article consacré à earthUI dans ce numéro pour la version de développement et les composants facultatifs.)
2. **Lancer l'application** : exécuter `earthUI::launch_glmnet()` dans R. L'application s'ouvre dans le navigateur web sur le port 7879. Il est également possible d'y accéder directement en se rendant à l'adresse <http://localhost:7879>. L'application mémorise le dernier mode Purpose utilisé et le restaure automatiquement.
3. **Créer ou ouvrir un projet** dans la section 1 de la barre latérale. Un projet fixe la finalité et le lieu de travail (pays, État, comté, ville) ainsi que ses dossiers d'entrée et de sortie. Les projets sont stockés sous une racine `regProj` partagée et sont communs aux applications sœurs earthUI et mgcvUI. C'est en ouvrant un projet doté d'une autre finalité que l'on change de mode.
4. **Importer les données** dans la section 2 en sélectionnant un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajouter les fichiers à ce dossier, puis cliquer sur **Refresh** (actualiser) pour les faire apparaître dans la liste.
5. **Importer depuis earthUI (facultatif)** — la section 3 permet d'importer un fichier de résultat earthUI `.rds` afin d'utiliser les fonctions de base charnières d'earth avec la régularisation de glmnet. Cette étape peut être ignorée en l'absence de résultat earthUI.
6. **Configurer les variables** — choisir la variable cible et les prédicteurs, définir les types de données, les signes attendus et attribuer les éventuels rôles spéciaux de colonnes.

7. **Régler les paramètres de glmnet** — alpha, lambda, famille, contraintes de signe, interactions et autres options.
8. **Ajuster le modèle** — cliquer sur « Fit Glmnet Model » (ajuster le modèle glmnet) et examiner les résultats dans le panneau principal.
9. **Exporter** — télécharger les prédictions au format Excel, générer et convertir un rapport Quarto ou, en mode évaluation, calculer les ajustements RCA et la Sales Comparison Grid (grille de comparaison).

Les paramètres sont automatiquement conservés dans le stockage local du navigateur et restaurés au rechargement du même fichier d'entrée.

## 2 Exigences relatives aux données d'entrée MLS

Pour les flux de travail d'évaluation immobilière et d'analyse de marché, les données d'entrée proviennent en général d'une exportation de Multiple Listing Service (MLS). Ce chapitre décrit la structure de fichier attendue et les colonnes que glmnetUI sait exploiter.

### 2.1 Format et structure du fichier

glmnetUI accepte les fichiers **CSV** et **Excel** (.xlsx, .xls). À l'import, les noms de colonnes sont automatiquement convertis en **snake\_case** — par exemple, « Living SqFt » devient **living\_sqft**, « Contract Date » devient **contract\_date** et « Sale Price » devient **sale\_price**. Cette normalisation garantit des références de colonnes cohérentes tout au long du flux de travail. Le séparateur CSV et le séparateur décimal utilisés à l'import sont déterminés par les paramètres de locale (voir le chapitre 3, « Locale & Regional Settings »).

Le fichier de données doit être un tableau plat comportant une ligne par bien et une colonne par attribut. La première ligne du fichier doit contenir les en-têtes de colonnes.

### 2.2 Colonnes requises en mode évaluation

Si glmnetUI fonctionne avec n'importe quel ensemble de colonnes, le flux de travail complet d'évaluation tire profit des colonnes suivantes :

| Colonne          | Type spécial           | Rôle                                                                              |
|------------------|------------------------|-----------------------------------------------------------------------------------|
| Sale Price       | (cible)                | Variable réponse du modèle                                                        |
| Contract Date    | <b>contract_date</b>   | Sert à calculer <b>sale_age</b> (jours depuis la date d'effet)                    |
| Listing Date     | <b>listing_date</b>    | Utilisée avec la date de contrat pour calculer le DOM en l'absence de colonne DOM |
| Days on Market   | <b>dom</b>             | Jours de mise en marché ; affiché dans les exportations                           |
| Concessions      | <b>concessions</b>     | Concessions de vente ; prix net = prix de vente – concessions                     |
| Living Area (SF) | <b>living_area</b>     | Active les résidus au pied carré ( <b>residual_sf</b> , <b>cqa_sf</b> )           |
| Lot Size         | <b>lot_size</b>        | Colonne de superficie du terrain                                                  |
| Site Dimensions  | <b>site_dimensions</b> | Regroupée avec la superficie du terrain                                           |
| Latitude         | <b>latitude</b>        | Arrondie à 3 décimales                                                            |
| Longitude        | <b>longitude</b>       | Arrondie à 3 décimales                                                            |
| Area ID          | <b>area</b>            | Identifiant de zone de marché ou de quartier                                      |
| Actual Age       | <b>actual_age</b>      | Âge du bien                                                                       |
| Effective Age    | <b>effective_age</b>   | Âge effectif du bien                                                              |
| Address          | <b>display_only</b>    | Affichée dans les exportations ; exclue du modèle                                 |

Les noms de colonnes du tableur peuvent être rédigés dans une langue étrangère — les noms « spéciaux » sont en anglais afin que le programme R puisse leur réserver un traitement particulier. Pour le reste, les noms de colonnes fournis apparaissent dans les modèles de régression, les graphiques et, s'il s'agit d'évaluations, dans les rapports produits.

Toutes les colonnes ne sont pas obligatoires. glmnetUI s'adapte : si une colonne manque, la fonctionnalité correspondante est simplement omise. Cela dit, pour les modèles de prix immobiliers, certaines colonnes sont vivement recommandées afin d'obtenir un ajustement acceptable :

1. **Sale Age** (ancienneté de la vente) — le nombre de jours entre la date de contrat de vente et la date d'effet de l'évaluation ou de l'analyse. Lorsqu'on utilise un historique de ventes pluriannuel, en particulier sur des périodes de plus de 5 ans, **sale\_age** joue souvent un rôle central dans l'estimation du prix de vente.
2. **Living Area** (surface habitable) — également désignée par « Living Sqft », « GLA » (gross living area) et autres. Il s'agit d'un autre déterminant majeur du prix de vente.
3. **Total Bath Count** (nombre total de salles de bains) — le nombre total de salles de bains complètes, aux trois quarts, à moitié et au quart. Par exemple, deux salles de bains complètes et une demi-salle de bains donnent la valeur 2,5.
4. **Garage Bays** ou **Garage Area** — le nombre de places de garage ou la surface du garage.
5. **Lot Size** (superficie du terrain) — la surface foncière du bien, généralement exprimée en pieds carrés ou en acres.
6. **Longitude**, **Latitude** et, si disponible, **Area ID**. Ces variables de localisation aident le modèle à rendre compte de la variation géographique des prix.

## 2.3 Conventions de nommage des colonnes spéciales

glmnetUI identifie les colonnes par leur **désignation de type spécial**, et non par leur nom. Les colonnes peuvent porter n'importe quel nom dans l'exportation MLS — seule compte l'attribution du bon type spécial dans le tableau Variable Configuration (chapitre 6).

Par exemple, votre MLS peut exporter la surface habitable sous le nom « GLA », « Living SqFt », « liv\_area » ou « gross\_living\_area ». Après l'import (au cours duquel le nom devient `snake_case`), il suffit de la désigner comme `living_area` dans la liste déroulante Special. glmnetUI l'utilisera alors pour les calculs de résidus au pied carré, quel que soit son nom d'origine.

## 2.4 Qualité et complétude des données

- **Valeurs manquantes (NA)** : les lignes comportant des NA dans une colonne prédictrice ou cible sont automatiquement supprimées avant l'ajustement.
- **Colonnes de dates** : les colonnes de type caractère correspondant aux formats de date courants sont automatiquement analysées à l'import. Les formats d'année à 2 chiffres sont privilégiés lorsqu'aucune année à 4 chiffres n'est détectée.
- **Colonnes numériques** : le prix de vente, la surface habitable, la superficie du terrain, les concessions et les champs analogues doivent être numériques. Si votre MLS exporte les prix avec des symboles monétaires ou des séparateurs de milliers (par exemple « \$350,000 » ou « 350.000,00 »), il peut être nécessaire de les nettoyer avant l'import.
- **Colonnes facteurs** : les colonnes de texte, TRUE/FALSE et les facteurs R sont automatiquement traités comme catégoriels — une régression ne peut les traiter que comme des facteurs. Une colonne **numérique** (par exemple un code de zone ou de style numérique) ne devient un facteur que si vous cochez sa case **Factor** ; glmnetUI ne déduit jamais le caractère catégoriel de l'étendue des valeurs d'une variable numérique, de sorte que des prédicteurs numériques discrets comme `bath_count` restent continus. Le rôle Special est indépendant de la décision relative au facteur.

***Conseil** : après l'import, examinez la colonne NA du tableau des prédicteurs. Les colonnes comportant de nombreuses valeurs manquantes peuvent entraîner la suppression de lignes. Il convient d'envisager d'exclure les colonnes à fort taux de NA ou de nettoyer les données avant l'import.*

## 2.5 Emplacement de la ligne du bien à évaluer

En mode **Appraisal**, la ligne 1 doit être le bien à évaluer. Toutes les lignes suivantes sont des ventes comparables. La ligne du bien à évaluer est exclue de l'ajustement du modèle. Après l'ajustement, le modèle produit néanmoins des prédictions pour cette ligne.

En mode **Market Area Analysis**, le placement du bien à évaluer en ligne 1 est facultatif.

En mode **General**, il n'existe aucun traitement particulier des lignes — toutes sont traitées de la même façon.

## 3 Mode General Purpose

### 3.1 Vue d'ensemble

Le mode General Purpose est le mode par défaut au lancement de glmnetUI. Il offre le flux de travail complet de régression elastic net pour tout jeu de données — et pas seulement en immobilier. glmnetUI peut ainsi servir à des données scientifiques, à des analyses financières, à des études techniques ou à tout problème de régression.

En mode General, l'interface omet les fonctionnalités propres à l'immobilier (colonnes spéciales, ancienneté de la vente, arrondi des coordonnées, RCA). La barre latérale est allégée et se concentre sur la sélection des variables, la configuration des paramètres, l'ajustement du modèle et l'exportation.

#### 3.1.1 Skip First Row

Une case à cocher **Skip first row** (ignorer la première ligne) apparaît en haut de la barre latérale dès qu'un projet General ou Market est ouvert. Lorsqu'elle est cochée, la ligne 1 est exclue de l'ajustement du modèle. C'est utile lorsque la ligne 1 contient une observation cible ou de référence qui ne doit pas influencer le modèle. En mode Appraisal, la ligne 1 (le bien à évaluer) est toujours exclue automatiquement.

#### 3.1.2 Paramètres propres à chaque finalité

Les paramètres (sélection des prédicteurs, paramètres du modèle, interactions) sont enregistrés séparément pour chaque combinaison de fichier d'entrée et de mode Purpose. Le passage d'un mode General, Appraisal ou Market à un autre préserve indépendamment les paramètres de chaque mode.

### 3.2 Le flux de travail de la barre latérale

La barre latérale est organisée en sections numérotées et repliables qui guident l'utilisateur depuis la sélection du projet jusqu'à l'exportation du rapport. La plupart des sections n'apparaissent qu'une fois un projet actif ouvert, et plusieurs sont propres à la finalité du projet. Le dossier de sortie est déduit du projet actif — il n'existe pas de champ distinct pour le dossier de sortie.

**1. Project** (projet) — créer ou ouvrir un projet. Un projet fixe la finalité et le lieu de travail (pays, État, comté, ville) ainsi que ses dossiers d'entrée (`<os>_in/`) et de sortie. Les projets sont stockés sous une racine `regProj` partagée (par défaut `~/regProj`, modifiable via la variable d'environnement `REGPROJ_ROOT`) et sont communs aux applications sœurs earthUI et mgcvUI, de sorte que les modèles des trois routines coexistent. Cliquer sur **Close Project** (fermer le projet) pour en changer.

**2. Import Data** (importer les données) — sélectionner un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajouter les fichiers au dossier via le Finder ou l'Explorateur, puis cliquer sur **Refresh** pour les lister. Pour les fichiers Excel comportant plusieurs feuilles, un sélecteur de feuille apparaît. Les noms de colonnes sont automatiquement convertis en `snake_case`.

**3. Import from earthUI (facultatif)** — importer un fichier de résultat earthUI `.rds` afin d'utiliser les fonctions de base charnières d'earth avec la régularisation elastic net de glmnet. Le bouton de navigation reprend le style du sélecteur de fichiers de la section 2. Cette étape est facultative — il suffit de l'ignorer en l'absence de résultat earthUI à importer.

**4. Variable Configuration** (configuration des variables) — sélecteur de variable cible, tableau des prédicteurs avec cases à cocher Include, Factor et Force, ainsi que des listes déroulantes

Sign. La colonne Special (pour désigner les dates de contrat, les coordonnées, etc.) n'apparaît qu'en modes Appraisal et Market. Voir le chapitre 6 pour le détail complet.

**5. glmnet Call Parameters** (paramètres d'appel de glmnet) — toute la configuration du modèle : alpha, lambda, famille, standardisation, contraintes de signe, relaxed lasso, graine aléatoire, matrice d'interactions et paramètres avancés (lambda.min.ratio, nlambdas, métrique de perte en validation croisée, seuil de convergence, nombre maximal d'itérations, ordonnée à l'origine). Voir le chapitre 7 pour la référence complète des paramètres.

**6. Fit Glmnet Model** (ajuster le modèle glmnet) — le bouton qui exécute le modèle. Les résultats apparaissent dans les onglets de droite.

**7. Download Output** (télécharger la sortie) — exporte les prédictions, les résidus, les scores CQA et les contributions par variable dans un fichier Excel placé dans le dossier de sortie du projet. Disponible dans tous les modes Purpose une fois le modèle ajusté.

**8. Calculate RCA Adjustments & Download** (calculer et télécharger les ajustements RCA) — calcule les ajustements RCA par variable pour chaque comparable et interpole le résidu du bien à évaluer via le CQA. N'apparaît qu'en modes Appraisal et Market, après l'ajustement. Voir le chapitre 11.

**9. Generate Sales Grid & Download** (générer et télécharger la grille de ventes) — génère la Sales Comparison Grid intermédiaire, en classant les comparables par pourcentage d'ajustement brut. N'apparaît qu'en mode Appraisal, après le calcul des ajustements RCA. Voir le chapitre 12.

**10. Generate Quarto Report** (générer le rapport Quarto) — écrit un ensemble de rapport Quarto (.qmd) autonome — avec tous les graphiques et tableaux — dans le dossier de sortie du projet. Disponible une fois le modèle ajusté.

**11. Convert Quarto Report** (convertir le rapport Quarto) — rend l'ensemble .qmd généré au format HTML, Word ou PDF. Le PDF requiert une installation LaTeX ; si aucune n'est détectée, l'option PDF est masquée. Voir le chapitre 13.

En modes Market et General, les étapes propres à l'évaluation sont masquées et les deux sections de rapport sont renumérotées en conséquence.

### 3.3 Projets

glmnetUI organise le travail en **projets** placés sous un dossier racine **regProj** partagé (la même arborescence que celle qu'utilisent earthUI et mgcvUI). Chaque projet possède un dossier d'entrée (<os>\_in) et des dossiers de sortie propres à chaque moteur (par exemple <os>\_out\_glmnet).

Pour créer un projet, cliquer sur + **New...** dans la section 1 et fournir :

- **Purpose** — General, Appraisal ou Market Area.
- **Location** (lieu) — la hiérarchie politique et administrative du projet. Les niveaux à renseigner dépendent du pays choisi, de sorte que les conventions de nommage étrangères sont prises en charge automatiquement — un projet américain utilise État/comté/ville, un projet français région/département/commune, un projet japonais préfecture/ville-ou-arrondissement/district, et ainsi de suite.
- **Project Name** (nom du projet) — une étiquette courte, limitée à 8 caractères (lettres, chiffres, \_ et -).

L'application n'utilise pas le nom saisi tel quel : elle construit un nom de dossier unique et horodaté en préfixant les codes de localisation et la date-heure de création, sous la forme <pays>\_<codes de niveaux administratifs>\_<AAAAAMJJ-HHMMSS>\_<votre nom> — par exemple, un projet américain nommé burl1 devient us\_ca\_081\_burlin\_20260604-131500\_burl1. La limite de 8 caractères ne s'applique qu'à la partie saisie. Seuls les projets nouvellement créés reçoivent le préfixe horodaté ; les projets déjà présents sur le disque conservent leur nom d'origine et restent pleinement accessibles.

### 3.4 Importer depuis earthUI (facultatif)

La section 3 de la barre latérale — **Import from earthUI** — permet d'importer un fichier de résultat earthUI `.rds`. glmnetUI peut ainsi utiliser les fonctions de base charnières d'earth (termes linéaires par morceaux) avec la régularisation elastic net de glmnet, combinant la détection automatique des non-linéarités propre à earth et la sélection de variables et les contraintes de coefficients propres à glmnet.

Lorsqu'un résultat earthUI est importé, ce sont les prédicteurs du modèle earth — et non les cases Include — qui pilotent l'ajustement glmnet, car glmnet s'ajuste sur l'expansion de base d'earth. Le tableau Variable Configuration l'indique explicitement : **Include** et **Type** sont verrouillés, les lignes correspondant aux prédicteurs non retenus par le modèle earth sont désactivées, et une note précise que l'ensemble des prédicteurs provient du modèle earth. **Force** et **Sign** restent modifiables pour les prédicteurs d'earth, puisqu'ils s'appliquent encore à la base earth au moment de l'ajustement.

Le bouton de navigation reprend le style du sélecteur de fichiers de la section 2. Cette étape est entièrement facultative — en l'absence de fichier de résultat earthUI, il suffit d'ignorer la section 3 et de passer à la section 4.

### 3.5 Onglets du panneau principal

Après l'ajustement, le panneau principal propose les onglets suivants :

| Onglet              | Contenu                                                                                                                                                                                                                                                    |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data Preview        | Aperçu des données importées. En mode évaluation, séparé en deux tableaux : Subject Property et Comparable Sales.                                                                                                                                          |
| Equation            | L'équation du modèle ajusté, affichée en LaTeX/MathJax, montrant chaque coefficient non nul et les termes d'interaction.                                                                                                                                   |
| Correlation         | Carte de chaleur des corrélations entre prédicteurs numériques (disponible avant l'ajustement).                                                                                                                                                            |
| Summary             | Indicateurs clés ( $R^2$ , $R^2$ ajusté, $GR^2$ , $R^2$ de validation croisée, RMSE, MAE) et indicateurs d'évaluation (COD, PRD, ratio médian).                                                                                                            |
| Coefficients        | Tableau des coefficients non nuls, avec avertissements en cas de violation de signe.                                                                                                                                                                       |
| Variable Importance | Diagramme en barres et tableau classé de l'importance des prédicteurs ( $ \beta  \times sd(x)$ ). Survol interactif plotly lorsque disponible.                                                                                                             |
| Contributions       | Graphiques d'effets partiels par variable, avec étiquettes de pente. Nuage de points et droite d'ajustement pour les prédicteurs numériques, boîtes à moustaches pour les facteurs, surface 3D, nuage de points ou carte de chaleur pour les interactions. |
| ANOVA               | Tableau de décomposition de la variance : SS par variable, % du SS du modèle, coefficient.                                                                                                                                                                 |
| Diagnostics         | Quatre graphiques : chemin de régularisation, erreur de validation croisée, valeurs observées vs prédites, résidus vs valeurs ajustées.                                                                                                                    |
| Glmnet Output       | Sortie brute du modèle : graine aléatoire utilisée, impression du modèle, lambda sélectionné, lambda.min/1se, gamma (si relaxed) et vecteur complet des coefficients.                                                                                      |
| Report              | Configuration du rapport et champs d'exportation.                                                                                                                                                                                                          |
| RCA Adjustments     | Résultats de l'analyse RCA (mode évaluation, après calcul).                                                                                                                                                                                                |

### 3.6 Persistance des paramètres

glmnetUI enregistre automatiquement la configuration dans le stockage local du navigateur, indexée par le nom du fichier d'entrée. Au rechargement du même fichier, tous les paramètres sont restaurés : sélection de la cible, cases des prédicteurs, types de données, signes attendus, paramètres de glmnet et matrice d'interactions. Le dernier mode Purpose utilisé est lui aussi conservé globalement et restauré au redémarrage de l'application. Trois options sont proposées :

- **Use last settings for input file** — restaurer les paramètres enregistrés pour ce fichier précis
- **Use default settings** — appliquer les valeurs par défaut globales enregistrées
- **glmnet defaults** — revenir aux valeurs d'usine ( $\alpha = 1$ ,  $\lambda_{1se}$  par validation croisée, 10 blocs, gaussian, standardisation activée)

Un bouton **Save current as default** (enregistrer comme valeur par défaut) enregistre tous les paramètres courants comme valeur par défaut globale.

### 3.7 Mode sombre

Cliquer sur l'icône lune/soleil dans le coin supérieur droit pour basculer entre le thème clair et le thème sombre. La préférence de thème est enregistrée dans le stockage local et persiste d'une session à l'autre. Tous les éléments de l'interface (tableaux, graphiques, cartes, boutons) s'adaptent au thème sélectionné via des variables CSS.

### 3.8 Locale et paramètres régionaux

glmnetUI prend en charge les conventions internationales de format numérique et CSV grâce à un système de locales fondé sur le pays. Le menu déroulant **Settings** de la barre de titre propose des sélecteurs **Country** (pays) et **Paper** (papier) pour 31 pays pris en charge. Chaque préreglage configure :

- **Séparateur CSV** — virgule (,) pour les États-Unis, le Royaume-Uni et le Japon, ou point-virgule (;) pour la majeure partie de l'Europe, où la virgule sert de séparateur décimal.
- **Séparateur décimal** — point (.) ou virgule (,).
- **Séparateur de milliers** — virgule (États-Unis/Royaume-Uni/Japon), point (Allemagne/Italie/Espagne), espace (Finlande/France/Pologne/pays baltes/Ukraine/Russie) ou apostrophe (Suisse).
- **Format de papier** — Letter (États-Unis/Canada/Mexique) ou A4 (partout ailleurs).

#### 3.8.1 Enregistrer les valeurs par défaut

Cliquer sur **Save as my default** pour enregistrer globalement vos préférences de locale. Ces valeurs par défaut s'appliquent à toutes les sessions futures, quel que soit le fichier de données chargé.

***Astuce :** le formatage des nombres sur les axes des graphiques et sur les étiquettes de pente s'adapte automatiquement à votre locale. La locale allemande utilise des points pour les milliers (200.000), la locale finlandaise des espaces (200 000), la locale suisse des apostrophes (200'000). Aucun symbole monétaire n'est affiché — glmnetUI est indépendant de la devise.*

## 4 Mode évaluation

Lorsque **For Appraisal** est sélectionné comme Purpose, glmnetUI se configure pour l'évaluation d'un bien individuel. Toutes les fonctionnalités décrites au chapitre 3 restent disponibles ; le présent chapitre ne traite que des ajouts propres à l'évaluation.

### 4.1 Traitement de la ligne du bien à évaluer

En mode évaluation, **la ligne 1 du jeu de données est le bien à évaluer** et toutes les lignes suivantes sont des ventes comparables. Le fichier d'entrée doit être organisé en conséquence (voir le chapitre 2). Le prix de vente du bien à évaluer peut être laissé vide ou renseigné avec n'importe quelle valeur — glmnetUI le traite automatiquement comme NA lors de l'ajustement.

Après l'import, l'onglet Data Preview se scinde en deux sections : **Subject Property** (ligne 1) et **Comparable Sales** (lignes 2 et suivantes). La ligne 1 est toujours exclue de l'ajustement du modèle. Après l'ajustement, le modèle produit néanmoins des prédictions pour cette ligne.

### 4.2 Date d'effet et ancienneté de la vente

En modes évaluation et marché, un champ **Effective Date** (date d'effet) apparaît dans la section Variable Configuration (initialisé à la date du jour). Si une colonne est désignée comme **contract\_date** dans la liste déroulante Special, glmnetUI calcule une colonne **sale\_age** — le nombre entier de jours entre la date de contrat de chaque vente et la date d'effet. Cette colonne est ajoutée comme prédicteur.

Lorsque la date d'effet change, **sale\_age** est automatiquement recalculée.

### 4.3 Désignations spéciales de colonnes

En modes évaluation et marché, une liste déroulante **Special** apparaît pour chaque prédicteur du tableau Variable Configuration. Voir le chapitre 6 pour la référence complète des types spéciaux et de leurs effets.

### 4.4 Aperçu des ajustements RCA

Le bouton **Calculate RCA Adjustments & Download** (section 8 de la barre latérale, visible en modes Appraisal et Market après l'ajustement) calcule des ajustements déduits du marché pour chaque comparable, relativement au bien à évaluer. Le flux de travail RCA complet est décrit au chapitre 11.

***Conseil :** le score CQA (état, qualité, attrait) que vous attribuez au bien à évaluer détermine la part de la distribution des résidus attribuée à celui-ci. Un score de 5.00 place le bien à évaluer à la médiane des biens comparables.*

## 5 Mode analyse de zone de marché

Lorsque **Market Area Analysis** est sélectionné comme Purpose, glmnetUI offre les mêmes fonctionnalités propres à l'immobilier que le mode évaluation (colonnes spéciales, ancienneté de la vente, arrondi des coordonnées), mais il est orienté vers l'analyse d'un groupe de biens plutôt que vers l'évaluation d'un bien unique.

### 5.1 Différences avec le mode évaluation

- **Le traitement de la ligne du bien à évaluer est souple** — en mode marché, la ligne 1 n'est pas automatiquement considérée comme un bien à évaluer.
- **Pas de Sales Comparison Grid** — l'étape « Calculate RCA Adjustments & Download » (section 8) est disponible en mode marché, mais l'étape Sales Grid (section 9) est réservée au mode évaluation.

### 5.2 Quand utiliser le mode marché

Le mode Market Area Analysis convient lorsqu'il s'agit de :

- construire un modèle de régression pour un quartier ou une zone de marché afin de comprendre les déterminants de la valeur ;
- analyser la façon dont des variables telles que la surface, l'âge, la superficie du terrain et la localisation influent sur les prix de vente au sein d'un groupe de biens ;
- préparer les éléments justificatifs d'une analyse des conditions de marché ou d'une délimitation de quartier ;
- travailler sur un jeu de données pour lequel les fonctionnalités de colonnes spéciales (ancienneté de la vente, arrondi des coordonnées) sont souhaitées sans que le flux de travail d'ajustements RCA soit nécessaire.

***Conseil :** le mode marché est également utile pour la régression immobilière générale lorsque vous souhaitez bénéficier des fonctionnalités de colonnes spéciales sans avoir besoin du flux de travail d'ajustements RCA.*

## 6 Sélection des variables

La section 4 de la barre latérale — **Variable Configuration** — est l'endroit où l'on choisit les colonnes qui participent au modèle et la façon dont elles sont traitées.

### 6.1 Variable cible

La liste déroulante **Target (response) variable** (variable cible ou réponse), en haut de la section 4, énumère toutes les colonnes du jeu de données. Sélectionner une colonne comme variable réponse (par exemple le prix de vente). La colonne cible est automatiquement exclue de la liste des prédicteurs.

### 6.2 Le tableau des prédicteurs

Sous le sélecteur de cible, un tableau énumère toutes les colonnes restantes avec les champs suivants :

| Colonne         | Description                                                                                                                |
|-----------------|----------------------------------------------------------------------------------------------------------------------------|
| <b>Variable</b> | Nom de la colonne                                                                                                          |
| <b>Type</b>     | Liste déroulante de type de données : <b>numeric</b> , <b>integer</b> , <b>character</b> , <b>Date</b> , <b>POSIXct</b>    |
| <b>Include</b>  | Case à cocher — inclure cette colonne comme prédicteur dans le modèle                                                      |
| <b>Factor</b>   | Case à cocher — traiter la variable comme un facteur catégoriel (crée des colonnes indicatrices dans la matrice du modèle) |
| <b>Force</b>    | Case à cocher — imposer la variable dans le modèle (facteur de pénalité = 0, jamais éliminée par le lasso)                 |
| <b>Special</b>  | Liste déroulante (évaluation et marché uniquement) — voir la référence des types de colonnes spéciales ci-dessous          |
| <b>Sign</b>     | Signe attendu du coefficient : positif, négatif ou indifférent                                                             |
| <b>NAs</b>      | Nombre de valeurs manquantes                                                                                               |

### 6.3 Détection des types de données et substitutions

glmnetUI détecte automatiquement les types de données à l'import. Les colonnes numériques, entières et de dates sont reconnues. Les colonnes de type caractère qui ressemblent à des dates (motifs de format de date courants) sont classées comme **Date**.

Toute détection peut être remplacée en modifiant la liste déroulante **Type**. Le changement de type modifie l'encodage de la colonne dans la matrice du modèle.

Les colonnes de texte, TRUE/FALSE et les facteurs R sont automatiquement traités comme catégoriels. Une colonne numérique ne devient un facteur que si vous cochez sa case **Factor** — glmnetUI ne déduit jamais le caractère catégoriel de l'étendue des valeurs d'une variable numérique, de sorte que des prédicteurs numériques discrets comme un nombre de salles de bains restent continus sauf indication contraire. Le réglage Factor est conservé via localStorage.

### 6.4 Signes attendus et application

Pour chaque prédicteur, la liste déroulante **Sign** précise si son coefficient est attendu positif, négatif ou dans l'une ou l'autre direction. Après l'ajustement :

- les coefficients dont le signe est inattendu sont **signalés dans le tableau Coefficients** ;

- des avertissements de signe apparaissent dans l'onglet Summary.

Lorsque la case **Enforce Sign Constraints** (imposer les contraintes de signe) est activée dans les paramètres de `glmnet` (chapitre 7), le modèle est contraint de produire des coefficients conformes aux signes attendus. Cela est mis en œuvre au moyen de `upper.limits` et `lower.limits` dans `glmnet`.

## 6.5 Référence des types de colonnes spéciales

En modes évaluation et marché, la liste déroulante **Special** propose les options suivantes :  
**Pondération :**

- **weight** — colonne de pondération des observations (disponible dans tous les modes ; une seule autorisée ; les lignes de poids = 0 sont exclues de l'ajustement)

**Types de date et d'heure :**

- **contract\_date** — déclenche le calcul automatique de **sale\_age** à partir de la date d'effet
- **listing\_date** — utilisée comme solution de repli pour calculer les jours de mise en marché
- **dom** — identifie la colonne des jours de mise en marché
- **sale\_age** — désigne une colonne d'ancienneté de vente existante, de sorte que n'importe quelle colonne puisse tenir ce rôle au lieu que `glmnetUI` la calcule à partir de **contract\_date**

**Types monétaires :**

- **concessions** — identifie les concessions de vente (crédits du vendeur, incitations à l'acheteur, etc.)

**Types de transaction :**

- **sale\_type** — identifie la colonne de type de vente ou de transaction

**Types de taille et de localisation :**

- **latitude** — valeurs automatiquement arrondies à 3 décimales
- **longitude** — même traitement d'arrondi que la latitude
- **area** — identifiant de zone de marché ou de quartier (par exemple `area_id`, `neighborhood`, `market_area`)
- **living\_area** — active les calculs de résidus au pied carré (**residual\_sf** et **cqa\_sf**)
- **lot\_size** — colonne de superficie du terrain
- **site\_dimensions** — regroupée avec la superficie du terrain (par exemple « 75x120 »)

**Types d'âge :**

- **actual\_age** — colonne d'âge du bien
- **effective\_age** — âge effectif du bien

**Types d'affichage :**

- **display\_only** — la colonne est incluse dans les exportations Excel mais entièrement exclue de l'ajustement du modèle. À utiliser pour les champs d'adresse, les numéros MLS, les identifiants de parcelle ou d'autres données de référence. Plusieurs colonnes peuvent porter cette désignation.

**Conseil :** les colonnes désignées comme **display\_only** demeurent dans le jeu de données et figurent dans l'exportation Excel, mais sont entièrement exclues de la table des prédicteurs. À utiliser pour les colonnes d'identifiants, les adresses ou d'autres champs de référence.

## 7 Choix des paramètres

La section 5 de la barre latérale — **glmnet Call Parameters** — donne accès à toutes les options de configuration du modèle elastic net. Chaque paramètre est accompagné d'une icône d'aide bleue (?) dont l'infobulle fournit une explication.

### 7.1 Alpha (paramètre de mélange)

Le paramètre alpha détermine le type de régularisation :

| Valeur d'alpha | Type        | Comportement                                                                  |
|----------------|-------------|-------------------------------------------------------------------------------|
| 0              | Ridge       | Rétrécit proportionnellement tous les coefficients ; n'en annule jamais aucun |
| 1              | Lasso       | Peut ramener des coefficients exactement à zéro (sélection de variables)      |
| Entre 0 et 1   | Elastic net | Mélange de ridge et de lasso                                                  |

Deux méthodes de choix d'alpha sont proposées :

- **Fixed** (fixe) — définir une valeur unique d'alpha à l'aide du curseur (par défaut : 1,0)
- **Grid Search** (recherche sur grille) — tester plusieurs valeurs d'alpha et retenir celle dont l'erreur de validation croisée est la plus faible. L'étendue et le nombre de valeurs sont configurables.

### 7.2 Choix de lambda

Lambda contrôle l'intensité de la régularisation :

| Méthode                   | Description                                                                          |
|---------------------------|--------------------------------------------------------------------------------------|
| <b>Validation croisée</b> | Recommandée. Teste de nombreuses valeurs de lambda par validation croisée à k blocs. |
| <b>Manuelle</b>           | Saisir une valeur précise de lambda si l'on a une raison de le faire.                |

En validation croisée, deux choix de lambda sont possibles :

- **lambda.1se** (par défaut) — le plus grand lambda situé à moins d'une erreur type du minimum. Produit des modèles plus simples et plus régularisés. Recommandé pour les travaux d'évaluation.
- **lambda.min** — le lambda qui minimise l'erreur de validation croisée. Produit le meilleur ajustement, mais un modèle potentiellement plus complexe.

Le paramètre **Number of CV Folds** (nombre de blocs de validation croisée) détermine la façon dont les données sont découpées pour la validation croisée (par défaut : 10 blocs).

### 7.3 Famille

Choisir la famille de distribution de la variable réponse :

| Famille         | Cas d'usage                                                          |
|-----------------|----------------------------------------------------------------------|
| <b>gaussian</b> | Réponses continues (par exemple le prix de vente). La plus courante. |
| <b>binomial</b> | Résultats binaires (par exemple vendu / non vendu).                  |
| <b>poisson</b>  | Données de comptage (par exemple le nombre de ventes).               |

## 7.4 Standardize

Lorsque la case est cochée (par défaut), tous les prédicteurs sont ramenés à une moyenne de 0 et un écart type de 1 avant l'ajustement. La pénalité traite ainsi tous les prédicteurs de la même façon, quelle que soit leur échelle d'origine. Les coefficients sont restitués à l'échelle initiale. Il est en général préférable de laisser cette option activée.

## 7.5 Contraintes de signe

Lorsque **Enforce Sign Constraints** est activé, les coefficients sont contraints de respecter les signes attendus définis dans le tableau des variables (chapitre 6) :

- un signe « positive » impose un coefficient  $\geq 0$  ;
- un signe « negative » impose un coefficient  $\leq 0$  ;
- les variables réglées sur « either » ne sont pas contraintes.

Cela est mis en œuvre au moyen des paramètres `upper.limits` et `lower.limits` de `glmnet()`.

## 7.6 Relaxed lasso

Le lasso ordinaire utilise la même pénalité pour la sélection des variables ET pour l'estimation des coefficients, ce qui peut rétrécir excessivement vers zéro des coefficients importants. Le **relaxed lasso** (lasso relâché) sépare ces deux étapes :

1. d'abord, sélectionner les variables par lasso (la pénalité détermine quels coefficients survivent) ;
2. ensuite, réestimer les variables retenues avec une pénalité réduite ou nulle, afin d'obtenir des estimations moins biaisées.

Le curseur **Gamma** règle le degré de relâchement :

| Gamma         | Effet                                                         |
|---------------|---------------------------------------------------------------|
| 0             | Entièrement relâché (réestimation MCO, aucun rétrécissement)  |
| 1             | Aucun relâchement (identique à <code>glmnet</code> ordinaire) |
| Intermédiaire | Mélange des ajustements relâché et pénalisé                   |

En validation croisée, gamma est choisi automatiquement pour une performance optimale.

## 7.7 Graine aléatoire

Un champ texte **Random seed** (graine aléatoire), prérempli avec un entier aléatoire, rend reproductible l'affectation des blocs de validation croisée — `set.seed()` est appelée avant l'ajustement, et la graine utilisée est indiquée dans l'onglet `Glmnet Output` ainsi que dans le message d'état de l'ajustement. Après chaque ajustement, une nouvelle graine est générée automatiquement (à partir de l'heure système, et non du générateur de nombres aléatoires de R). Les 5 dernières graines utilisées pour le fichier courant sont affichées sous forme de liens cliquables permettant de les rappeler aisément ; l'historique des graines est réinitialisé au chargement d'un nouveau fichier.

## 7.8 Matrice d'interactions

Une matrice triangulaire supérieure interactive permet de choisir quelles paires de variables sont autorisées à interagir. Chaque cellule contient une case à cocher — cochée, elle signifie que le terme d'interaction  $x_1:x_2$  est inclus dans la matrice du modèle.

- **Allow All** (tout autoriser) — cocher toutes les paires d'interaction
- **Clear All** (tout effacer) — décocher toutes les paires (par défaut, tout est décoché pour un nouveau fichier)
- **Cliquer sur le nom d'une variable** — basculer toutes les interactions de cette variable

Les sélections d'interactions sont enregistrées dans le localStorage du navigateur, par fichier d'entrée, et restaurées au rechargement du même fichier.

Dès qu'une interaction est activée, le rapport exporté comprend une matrice **Enabled Interactions** (interactions activées) — les prédicteurs sur les deux axes, une coche là où le terme produit  $x_1:x_2$  a été inclus — rendue en format paysage et calquée sur la matrice Allowed Interactions d'earthUI. Il convient de noter que le lasso peut néanmoins avoir ramené à zéro le coefficient d'un terme activé.

***Conseil :** avec de nombreux prédicteurs, le lasso annulera automatiquement les termes d'interaction inutiles. Partir de « Allow All » et laisser le modèle sélectionner constitue une approche raisonnable.*

### Avertissement

L'activation des interactions dans glmnet produit des ajustements qui peuvent être difficiles à interpréter et à expliquer devant un tribunal ou lors d'un audit. Les interactions doivent être employées avec prudence et il convient de vérifier que les ajustements obtenus sont raisonnables.

## 7.9 Exclusion de l'effet principal (interaction seule)

Un **clic droit** sur le nom d'une variable dans la matrice d'interactions l'empêche d'entrer dans le modèle comme effet principal. Un indicateur « 1 » en gras apparaît après le nom de la variable. Lorsqu'une variable est ainsi exclue :

- elle est **retirée** des termes principaux de la formule ;
- elle **peut encore** figurer dans des termes d'interaction ( $x_1:x_2$ ) ;
- un nouveau clic droit lève l'exclusion ;
- l'état est conservé via le localStorage, par fichier d'entrée.

Cas d'usage typique : lorsqu'on modélise des ajustements temporels qui doivent varier avec la taille du bien, on exclut `sale_age` de l'effet principal et on ne l'autorise qu'en interaction avec `living_area`. On obtient ainsi un terme d'interaction `sale_age:living_area` au lieu d'un coefficient `sale_age` uniforme — le taux d'appréciation dans le temps varie alors selon la taille de la maison.

## 7.10 Paramètres avancés

Une section repliable « Advanced » (avancé), au bas de glmnet Call Parameters, expose des réglages supplémentaires. Ceux-ci ont des valeurs par défaut judicieuses, mais restent visibles afin que tous les réglages du modèle puissent être documentés pour un tribunal ou un audit.

- **Lambda Min Ratio** — rapport entre le plus petit et le plus grand lambda du chemin de régularisation. Par défaut : 0,00001. Pour les petits jeux de données (< 50 observations), une valeur plus élevée (par exemple 0,01) peut améliorer les performances.

- **Number of Lambda Values** — nombre de valeurs de lambda du chemin. Par défaut : 100.
- **CV Loss Metric** — fonction de perte pour la validation croisée : MSE (par défaut pour gaussian), MAE (robuste aux valeurs aberrantes) ou déviance.
- **Convergence Threshold** — seuil de convergence de la descente par coordonnées. Par défaut : 1e-07.
- **Max Iterations** — nombre maximal d'itérations de la descente par coordonnées. Par défaut : 100 000. À augmenter si des avertissements de convergence apparaissent.
- **Fit Intercept** — inclure un terme constant (ordonnée à l'origine). Par défaut : activé. L'ordonnée à l'origine constitue la « valeur de base » dans la RCA en évaluation.

### 7.11 Valeurs par défaut des paramètres

Trois options déterminent l'initialisation des paramètres au chargement d'un fichier :

- **Use last settings for input file** — restaurer exactement les paramètres utilisés la dernière fois avec ce fichier
- **Use default settings** — appliquer les valeurs par défaut globales enregistrées
- **glmnet defaults** — revenir aux réglages d'usine ( $\alpha = 1$ ,  $\lambda_{1se}$  par validation croisée, 10 blocs, gaussian, standardisation activée)

Cliquer sur **Save current as default** pour enregistrer tous les paramètres courants comme valeur par défaut globale pour les fichiers futurs.

## 8 Ajustement du modèle

### 8.1 Le bouton d'ajustement

La section 6 de la barre latérale contient le bouton **Fit Glmnet Model**. Un clic exécute le modèle selon la configuration courante.

### 8.2 Ce qui se passe pendant l'ajustement

Lorsqu'on clique sur Fit, glmnetUI :

1. **Prépare les données** — supprime les lignes comportant des NA, applique les pondérations (si une colonne est désignée), exclut la ligne du bien à évaluer (en mode évaluation) et encode les variables facteurs sous forme de colonnes indicatrices.
2. **Construit la matrice du modèle** — crée la matrice de plan d'expérience, incluant les éventuels termes d'interaction autorisés par la matrice d'interactions.
3. **Applique les contraintes** — si l'application des signes est activée, définit `upper.limits` et `lower.limits` sur les coefficients.
4. **Exécute la validation croisée** — si la validation croisée est retenue, appelle `cv.glmnet()` pour trouver le lambda optimal. Si le relaxed lasso est activé, appelle `cv.glmnet(..., relax = TRUE)`.
5. **Ajuste le modèle final** — enregistre le modèle ajusté, les coefficients, les prédictions et les résidus.
6. **Met à jour tous les onglets de résultats** — Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA et Diagnostics sont renseignés.

Un message d'état, sous le bouton Fit, indique le nombre d'observations, les lignes exclues et le lambda retenu.

## 9 Onglets de résultats

### 9.1 Data Preview

Affiche les données importées sous forme de DataTable interactif. En modes évaluation et marché, l'aperçu est scindé en deux tableaux : **Subject Property** (ligne 1) et **Comparable Sales** (lignes 2 et suivantes).

Chaque cellule tient sur une seule ligne et est tronquée à la largeur de la colonne, de sorte que les champs de texte libre étendus (tels que les remarques sur le bien) n'étirent pas une ligne sur toute la hauteur de la page. **Un double-clic sur une cellule** ouvre une fenêtre contextuelle affichant son contenu complet, non tronqué.

### 9.2 Equation

Affiche l'équation du modèle ajusté, rendue en LaTeX via MathJax. Tous les coefficients non nuls sont présentés avec une mise en forme mathématique correcte :

- coefficients positifs précédés du signe +
- coefficients négatifs précédés du signe −
- termes d'interaction affichés sous la forme  $x_1 \times x_2$
- nombres formatés avec séparateurs de milliers et un nombre approprié de décimales

### 9.3 Correlation

Une carte de chaleur matricielle présentant les corrélations de Pearson entre tous les prédicteurs numériques et la variable réponse. Disponible immédiatement après l'import des données — aucun ajustement de modèle n'est requis.

- Dégradé de couleurs : bleu (−1) → blanc (0) → rouge (+1)
- Valeurs de corrélation imprimées dans chaque cellule
- Taille du texte et des axes adaptée au nombre de variables
- Utile pour repérer la multicollinéarité avant l'ajustement

### 9.4 Summary

Statistiques d'ajustement du modèle, affichées sous forme de cartes :

- **$R^2$**  — proportion de variance expliquée
- **$R^2$  ajusté** — ajusté selon le nombre de prédicteurs
- **$GR^2$**  —  $R^2$  généralisé, issu du rapport de déviance de glmnet au lambda retenu
- **$R^2$  de validation croisée** —  $R^2$  de validation croisée (lorsque la validation croisée est utilisée)
- **RMSE** — racine de l'erreur quadratique moyenne
- **MAE** — erreur absolue moyenne

En modes évaluation et marché, des cartes supplémentaires affichent :

- **Median Ratio** — médiane du rapport (prédit / observé)
- **COD** — coefficient de dispersion (plus il est faible, mieux c'est)
- **PRD** — différentiel lié au prix (Price-Related Differential)

Un **avertissement de surajustement** apparaît lorsque le  $R^2$  d'apprentissage dépasse de plus de 0,1 le  $R^2$  de validation croisée.

Sous les cartes, un tableau de coefficients avec avertissements de signe reprend, par commodité, le contenu de l'onglet Coefficients.

## 9.5 Coefficients

Un tableau de tous les coefficients non nuls indiquant :

- le nom de la variable
- la valeur du coefficient
- le signe attendu (issu de la configuration des variables)
- un indicateur d'avertissement si le sens du coefficient ne correspond pas aux attentes

## 9.6 Variable Importance

Amplitude standardisée du coefficient :  $|\beta| \times \text{sd}(x)$  pour chaque prédicteur, agrégée sur les colonnes indicatrices dans le cas des variables facteurs. Deux modes d'affichage :

- **plotly interactif** (lorsque le package `plotly` est installé) : diagramme en barres horizontales avec infobulles au survol indiquant l'importance exacte, le coefficient et le pourcentage relatif.
- **ggplot2 statique** (solution de repli) : diagramme en barres horizontales avec étiquettes d'axes formatées avec séparateurs de milliers.

Sous le graphique, un `DataTable` affiche Variable, Importance, Coefficient et % relatif.

## 9.7 Contributions

Graphiques d'effets partiels par variable, montrant la contribution de chaque prédicteur aux prédictions du modèle. La variable se choisit dans une liste déroulante.

**Pour les prédicteurs numériques** : un nuage de points des valeurs  $x$  de la variable en regard de sa contribution ( $\beta \times x$ ), avec un segment de droite rouge et une étiquette de pente. L'étiquette de pente indique l'effet marginal par unité (par exemple **+1,234.56/unit**), avec des unités adaptatives pour les variables de faible étendue comme la latitude et la longitude (par exemple **+0.12/0.001**). Le sous-titre indique la valeur de l'ordonnée à l'origine (valeur de base).

**Pour les variables facteurs** : une boîte à moustaches représentant la distribution des contributions par niveau du facteur.

**Pour les interactions à deux variables** : trois visualisations superposées — une surface 3D (`plotly` interactif lorsque disponible, `persp()` statique sinon) de la contribution conjointe, un nuage de points dont les points sont colorés selon la contribution, et une carte de chaleur de la contribution moyenne sur l'étendue des deux variables.

## 9.8 ANOVA

Tableau de décomposition de la variance indiquant, pour chaque prédicteur :

- **SS** — somme des carrés :  $\sum (\text{contribution}_j - \overline{\text{contribution}})^2$
- **% du SS du modèle** — pourcentage de la somme des carrés totale du modèle
- **Coefficient** — le coefficient dominant pour ce prédicteur

Le tableau comprend une ligne d'ordonnée à l'origine et une ligne TOTAL MODEL. Les termes d'interaction (contenant « : ») sont présentés comme des groupes distincts.

## 9.9 Diagnostics

Quatre graphiques de diagnostic à grands caractères (taille de base 16 pt), 15 graduations d'axe et étiquettes formatées avec séparateurs de milliers (sans notation scientifique) :

1. **Coefficient Path** (chemin de régularisation) — montre l'évolution de tous les coefficients à mesure que  $\log(\lambda)$  augmente. Les courbes sont colorées par variable. Aide à visualiser le chemin de régularisation et la sélection de variables.

2. **CV Error** (erreur de validation croisée) — erreur de validation croisée en fonction de  $\log(\lambda)$ . Les points représentent l'erreur moyenne de validation croisée ; les barres d'erreur indiquent  $\pm 1$  erreur type. Des lignes verticales pointillées marquent  $\lambda_{\min}$  (bleu) et  $\lambda_{1se}$  (vert).
3. **Actual vs Predicted** (observé vs prédit) — nuage de points des valeurs observées en regard des valeurs prédites. Les points devraient se regrouper autour de la droite de référence pointillée à 45 degrés. Une dispersion plus large signale une erreur de prédiction plus importante.
4. **Residuals vs Fitted** (résidus vs valeurs ajustées) — nuage de points des résidus en regard des valeurs ajustées. La dispersion devrait être aléatoire autour de la ligne zéro (pointillée). Toute structure suggère une mauvaise spécification du modèle (par exemple non-linéarité, hétéroscédasticité).

Chaque graphique dispose d'un bouton **Download PNG** permettant de l'enregistrer à 150 DPI.

### 9.10 Glmnet Output

Sortie brute du modèle indiquant : la graine aléatoire utilisée, l'impression du modèle, le lambda retenu,  $\lambda_{\min}/\lambda_{1se}$ , gamma (si le relaxed lasso est activé) et le vecteur complet des coefficients.

### 9.11 Report

Permet de renseigner les champs du rapport (nom de l'évaluateur, adresse du bien, date du rapport, numéro de dossier) et d'exporter directement depuis l'onglet au format Word (.docx) ou PDF. Pour le flux de travail complet de rapport fondé sur Quarto (HTML, Word ou PDF), voir le chapitre 13.

## 10 Téléchargement des données

Après l'ajustement, le bouton **Download Output (Excel)** (section 7 de la barre latérale, disponible dans tous les modes Purpose) exporte un fichier Excel contenant les prédictions et les diagnostics dans le dossier de sortie du projet. L'intitulé de la section est « Download Estimated Sale Prices & Residuals » en modes évaluation et marché, et « Download Estimated Target Variable(s) & Residuals » en mode General.

### 10.1 Colonnes de sortie

| Colonne                               | Description                                                           |
|---------------------------------------|-----------------------------------------------------------------------|
| <code>est_&lt;target&gt;</code>       | Prédiction du modèle (par exemple <code>est_sale_price</code> )       |
| <code>residual</code>                 | Observé – prédit                                                      |
| <code>cqa</code>                      | Score Condition-Quality-Appeal (état, qualité, attrait), échelle 0–10 |
| <code>residual_sf</code>              | Résidu / surface habitable (si <code>living_area</code> est désignée) |
| <code>cqa_sf</code>                   | CQA calculé à partir du classement selon <code>residual_sf</code>     |
| <code>&lt;var&gt;_contribution</code> | Contribution de chaque variable à la prédiction                       |
| <code>basis</code>                    | Valeur de l'ordonnée à l'origine (identique pour toutes les lignes)   |

Une coche blanche apparaît sur le bouton de téléchargement une fois l'opération réussie.

### 10.2 Scores CQA

Le CQA classe le résidu de chaque ligne par rapport à tous les autres, sur une échelle de 0 à 10 :

- **CQA élevé (~9–10)** : vendu bien plus cher que prédit — probablement un état, une qualité ou un attrait supérieurs
- **CQA faible (~0–1)** : vendu bien moins cher que prédit — probablement un état inférieur ou une vente en difficulté
- **CQA ~5** : proche de la médiane

En modes évaluation et marché, les lignes sont triées par `residual_sf` décroissant lorsqu'une colonne `living_area` est désignée.

## 11 Calculs RCA et téléchargement

Le flux de travail RCA (Residual Constraint Approach, approche par contraintes résiduelles) est disponible en modes **Appraisal** et **Market Area Analysis**, après l'ajustement du modèle. La Sales Comparison Grid qui en découle (chapitre 12) est réservée au mode évaluation.

### 11.1 Ouverture de la boîte de dialogue RCA

Cliquer sur le bouton **Calculate RCA Adjustments & Download** dans la section 8 de la barre latérale. Choisir :

- le type de score, **CQA** ou **CQA per SF** (CQA au pied carré). Si « CQA per SF » est retenu, le résidu du bien à évaluer est mis à l'échelle de la surface habitable.
- Saisir le score CQA du bien à évaluer (0,00–9,99, par défaut 5,00)

### 11.2 Interpolation du score CQA

1. Les scores CQA et les résidus des comparables sont triés
2. Une interpolation linéaire fait correspondre la valeur CQA choisie à un résidu
3. Valeur du bien à évaluer = prédiction du modèle + résidu interpolé

### 11.3 Colonnes de sortie

| Colonne             | Description                                                 |
|---------------------|-------------------------------------------------------------|
| subject_value       | Prédiction du modèle + résidu interpolé                     |
| <var>_adjustment    | Contribution du bien à évaluer – contribution du comparable |
| residual_adjustment | Résidu du bien à évaluer – résidu du comparable             |
| net_adjustments     | Somme de tous les ajustements                               |
| gross_adjustments   | Somme des valeurs absolues des ajustements                  |
| adjusted_sale_price | Prix de vente du comparable + ajustements nets              |

Une coche blanche apparaît sur le bouton une fois le calcul réussi.

## 12 Sales Comparison Grid

La Sales Comparison Grid (grille de comparaison des ventes) n'est disponible qu'en mode **Appraisal**, après le calcul des ajustements RCA (étape 8). Elle génère un classeur Excel mis en forme, prêt à être intégré aux rapports d'évaluation.

### 12.1 Boîte de dialogue de sélection des comparables

Un clic sur le bouton **Generate Sales Grid & Download** de la section 9 de la barre latérale ouvre une fenêtre modale énumérant toutes les ventes comparables issues de la sortie RCA. Les comparables sont répartis en deux groupes :

- **Recommended comps** (comparables recommandés, cochés d'office) : ajustement brut < 25 % du prix de vente, triés par ancienneté de la vente (les plus récents d'abord).
- **Other comps** (autres comparables, décochés par défaut) : les comparables restants, triés par ancienneté de la vente.

Sélectionner jusqu'à 30 comparables, puis cliquer sur **Generate Sales Grid** pour créer le classeur.

### 12.2 Structure du classeur

Le classeur produit comporte jusqu'à 10 feuilles, chacune contenant 3 comparables côte à côte :

- **Lignes d'en-tête** : adresses du bien à évaluer et des comparables, prix de vente, dates de vente et jours de mise en marché
- **Lignes groupées** : localisation (latitude, longitude, zone), taille du terrain (superficie, dimensions du site) et âge (âge réel, âge effectif), avec en-têtes de groupe et sous-éléments
- **Lignes de variables du modèle** : une ligne par coefficient non nul du modèle, indiquant la contribution du bien à évaluer, les contributions des comparables et les pourcentages d'ajustement
- **Lignes de caractéristiques résiduelles** : cellules vides et déverrouillées, destinées à la saisie par l'évaluateur des caractéristiques non captées par le modèle (par exemple état, qualité, vue)
- **Lignes de synthèse** : % d'ajustement net, % d'ajustement brut et prix de vente ajusté, calculés par des formules Excel actives

### 12.3 Protection des feuilles et cellules résiduelles

Chaque feuille est protégée par un mot de passe afin d'éviter les modifications accidentelles, mais les cellules de valeur des caractéristiques résiduelles sont explicitement déverrouillées. Les évaluateurs peuvent ainsi saisir des valeurs pour des caractéristiques absentes du modèle tout en préservant les calculs d'ajustement pilotés par les formules.

## 13 Téléchargement des rapports

### 13.1 Le flux de travail Quarto en deux étapes

L'exportation d'un rapport depuis la barre latérale suit un flux de travail Quarto en deux étapes :

1. **Generate Quarto Report** (section 10 de la barre latérale) — écrit un ensemble Quarto autonome (la source `.qmd` ainsi que les graphiques, `report_data.rds` et `reference.docx`) dans le dossier de sortie glmnet du projet.
2. **Convert Quarto Report** (section 11 de la barre latérale) — rend un fichier `.qmd` dans les formats de sortie choisis. Le chemin proposé par défaut est celui du dernier ensemble généré, mais il est possible de naviguer vers n'importe quel `.qmd`.

### 13.2 Formats de rapport

Trois formats sont proposés à l'étape de conversion :

- **HTML** — document HTML autonome. Fonctionne sur toutes les plateformes sans logiciel supplémentaire au-delà de pandoc. Format par défaut recommandé.
- **Word (.docx)** — document mis en forme. Adapté à l'édition et à la diffusion. Fonctionne sur toutes les plateformes.
- **PDF** — généré via Quarto avec LaTeX. Le format de papier suit le réglage de locale (Letter ou A4). **Requiert une installation LaTeX** (voir l'annexe consacrée à la configuration requise). Si aucune installation LaTeX n'est détectée, l'option PDF est automatiquement masquée dans les cases de format.

Les rapports sont générés au moyen de Quarto lorsqu'il est disponible, avec repli sur rmarkdown. L'onglet Report propose en outre des boutons d'export direct vers Word et vers PDF.

### 13.3 Contenu des rapports

Les rapports reprennent le contenu de tous les onglets :

- Description du jeu de données (nom du fichier, observations, réponse, prédicteurs, finalité)
- Spécification du modèle (alpha, lambda, famille, standardisation, relaxed, coefficients non nuls)
- Synthèse des résultats ( $R^2$ ,  $R^2$  ajusté,  $R^2$  généralisé,  $R^2$  de validation croisée, RMSE, MAE)
- Équation du modèle (rendue en LaTeX)
- Tableau des coefficients
- Importance des variables (diagramme en barres et tableau)
- Graphiques de contribution par variable, correspondant à l'onglet Contributions :
  - **Prédicteurs numériques** : nuage de points avec droite d'ajustement linéaire
  - **Variables facteurs** : boîte à moustaches par niveau
  - **Interactions** : nuage de points coloré selon la contribution, carte de chaleur de la contribution moyenne et instantané statique de la surface 3D (persp)
- Matrice Enabled Interactions (dès qu'une interaction est activée) — les prédicteurs sur les deux axes, avec une coche là où le terme produit  $x_1:x_2$  a été inclus, rendue en format paysage
- Carte de chaleur de la matrice de corrélation
- Tableau de décomposition ANOVA
- Graphiques de diagnostic :

- Chemin de régularisation
- Erreur de validation croisée
- Observé vs prédit
- Résidus vs valeurs ajustées
- Diagramme quantile-quantile normal
- Sortie du modèle

Toutes les étiquettes d'axes et les légendes de couleurs emploient des nombres formatés avec séparateurs de milliers (sans notation scientifique).

Une coche blanche apparaît sur le bouton de rapport une fois la génération réussie, et un chronomètre s'affiche pendant le rendu du rapport.

## 14 Comparaison avec earthUI

L'application earth et son Post Processing Module glmnet forment un ensemble d'outils unique pour la modélisation par régression. Ils partagent le même format de données, les mêmes types de colonnes spéciales, le même flux de travail RCA et les mêmes jeux de données de démonstration, mais reposent sur des moteurs de modélisation différents.

### 14.1 Principales différences

| Caractéristique              | glmnetUI                                   | earthUI                                       |
|------------------------------|--------------------------------------------|-----------------------------------------------|
| Méthode                      | Elastic net (glmnet)                       | MARS/Earth (earth)                            |
| Sélection de variables       | Automatique via la pénalité lasso          | Automatique via élagage avant/arrière         |
| Relations entre coefficients | Effets linéaires uniquement                | Linéaires par morceaux (fonctions charnières) |
| Interactions                 | Deux à deux, via la matrice d'interactions | Jusqu'au degré 3                              |
| Signes des coefficients      | Imposables par contraintes                 | Non imposables directement                    |
| Ajustement relâché           | Relaxed lasso disponible                   | Sans objet                                    |
| Réglage d'alpha              | Recherche sur grille disponible            | Sans objet (pas de paramètre de mélange)      |
| Sales Comparison Grid        | Disponible (étape 9)                       | Disponible (étape 8)                          |
| Graphiques de g-fonctions    | Sans objet                                 | Visualisation des termes groupés              |
| Formats de rapport           | HTML, Word, PDF                            | HTML, Word, PDF                               |

### 14.2 Fonctionnalités communes

Les deux outils offrent :

- l'import CSV et Excel avec conversion des noms de colonnes en snake\_case ;
- la prise en charge des locales de 31 pays, avec séparateur CSV, séparateur décimal et format de papier ;
- les désignations spéciales de colonnes (contract\_date, living\_area, area, latitude, longitude, etc.) ;
- le calcul automatique de sale\_age à partir de la date d'effet ;
- le basculement entre mode sombre et mode clair ;
- la persistance des paramètres via localStorage ;
- la carte de chaleur des corrélations, l'importance des variables, les contributions, l'ANOVA et les diagnostics ;
- le score CQA et le flux de travail d'ajustements RCA ;
- la Sales Comparison Grid avec sélection des comparables et sortie Excel mise en forme ;
- la compatibilité avec les mêmes jeux de données de démonstration (Appraisal\_1.csv).

### 14.3 Lequel utiliser, et quand

- **Utiliser glmnetUI** lorsqu'on souhaite un modèle linéaire simple et interprétable, doté d'une sélection automatique de variables. L'elastic net est particulièrement adapté lors-

qu'on dispose de nombreux prédicteurs et que l'on souhaite laisser le modèle retenir les plus importants.

- **Utiliser earthUI** lorsqu'on s'attend à des relations non linéaires (par exemple lorsque l'effet de la surface habitable sur le prix varie selon la taille) ou lorsque des fonctions charnières sont nécessaires pour capter des effets de seuil. Les modèles earth peuvent en outre capter des interactions d'ordre supérieur (degrés 2 et 3). earthUI propose par ailleurs une fonctionnalité expérimentale d'extraction des remarques fondée sur Prolog/DCG, sans équivalent dans glmnetUI.
- **Utiliser les deux** pour comparer modèles linéaires et non linéaires sur les mêmes données. Si les deux produisent des  $R^2$  voisins, le modèle glmnet, plus simple, peut être préféré pour son interprétabilité.

## 15 Jeu de données de démonstration : Appraisal\_1.csv

### 15.1 Description

Le jeu de données de démonstration est fourni avec glmnetUI (et partagé avec earthUI). Il se localise ainsi :

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

Le fichier contient 1 502 ventes résidentielles (plus 1 bien à évaluer en ligne 1) issues d'une exportation MLS simulée. Les données représentent des ventes de maisons individuelles dans un marché comportant plusieurs zones, avec une gamme variée de tailles, d'âges et de localisations.

Il ne s'agit pas de données réelles, mais elles s'inspirent d'un quartier réaliste du nord de la Californie. Toutes les informations d'identification ont été modifiées ou supprimées.

### 15.2 Colonnes

| Colonne          | Type      | Type spécial  | Description                                                              |
|------------------|-----------|---------------|--------------------------------------------------------------------------|
| weight           | numeric   | weight        | Pondération de l'observation (0 = exclue de l'ajustement)                |
| id               | numeric   | display_only  | Identifiant interne d'enregistrement                                     |
| property_id      | numeric   | display_only  | Identifiant MLS du bien                                                  |
| listing_id       | character | display_only  | Numéro d'annonce MLS                                                     |
| parcel_number    | character | display_only  | Numéro de parcelle du service du cadastre (APN)                          |
| street_address   | character | display_only  | Adresse du bien                                                          |
| city_name        | character | display_only  | Ville                                                                    |
| postal_code      | character | display_only  | Code postal                                                              |
| county_name      | character | display_only  | Comté                                                                    |
| contract_date    | Date      | contract_date | Date du contrat de vente (permet de calculer sale_age)                   |
| sale_age         | numeric   | —             | Nombre de jours entre la date de contrat et la date d'effet              |
| sale_price       | numeric   | (cible)       | Prix de vente — variable réponse                                         |
| living_sqft      | numeric   | living_area   | Surface habitable brute en pieds carrés                                  |
| beds_total       | integer   | —             | Nombre de chambres                                                       |
| baths_total      | numeric   | —             | Nombre total de salles de bains (par exemple 2,5 = 2 complètes + 1 demi) |
| lot_size         | numeric   | lot_size      | Superficie du terrain en pieds carrés                                    |
| area_id          | integer   | area          | Identifiant de zone MLS                                                  |
| age              | numeric   | actual_age    | Âge du bien en années                                                    |
| latitude         | numeric   | latitude      | Latitude (arrondie à 3 décimales pour le modèle)                         |
| longitude        | numeric   | longitude     | Longitude (arrondie à 3 décimales pour le modèle)                        |
| garage_spaces    | integer   | —             | Nombre de places de garage                                               |
| days_on_market   | integer   | dom           | Jours de mise en marché                                                  |
| listing_date     | Date      | listing_date  | Date de mise en annonce                                                  |
| sale_concessions | numeric   | concessions   | Concessions du vendeur                                                   |

### 15.3 Prise en main rapide suggérée

1. Lancer la routine glmnet : `earthUI::launch_glmnet()`
2. Créer un nouveau projet (section 1) avec la finalité **For Appraisal**
3. Placer **Appraisal\_1.csv** dans le dossier d'entrée du projet et le sélectionner dans la section 2 (cliquer sur **Refresh** si nécessaire)
4. Sélectionner **sale\_price** comme variable cible
5. Attribuer les types spéciaux comme indiqué dans le tableau ci-dessus
6. Inclure les prédicteurs : **sale\_age**, **living\_sqft**, **baths\_total**, **lot\_size**, **area\_id** (en facteur), **age**, **latitude**, **longitude**, **garage\_spaces**
7. Conserver  $\alpha = 1$  (Lasso), validation croisée avec `lambda.1se`
8. Cliquer sur **Fit Glmnet Model**
9. Examiner les onglets Summary, Coefficients et Diagnostics
10. Télécharger la sortie (étape 7) et examiner le classement CQA
11. Calculer les ajustements RCA (étape 8) avec un score CQA d'environ 5,00
12. Générer une Sales Comparison Grid (étape 9) avec les comparables recommandés

## 16 Configuration requise et résolution des problèmes

### 16.1 Plateformes prises en charge

glmnetUI fonctionne sous **macOS**, **Windows** et **Linux**. RStudio Desktop (2023.06 ou ultérieur) est vivement recommandé — il embarque **pandoc**, nécessaire aux rapports.

### 16.2 Dépendances facultatives

- **LaTeX** (TinyTeX, MiKTeX ou MacTeX) : uniquement pour la génération de rapports PDF. En l'absence de détection, l'option PDF est automatiquement masquée. Installation : `tinytex::install_tinytex()`
- **sysfonts + showtext** : police Roboto Condensed pour les graphiques. Repli automatique sur la police sans empattement du système si elle est indisponible ou hors ligne.
- **earth** ( $\geq 5.3.0$ ) : nécessaire uniquement pour la chaîne d'import depuis earthUI.

### 16.3 Remarques par plateforme

**macOS** : le moins de difficultés. Installer TinyTeX pour le PDF : `tinytex::install_tinytex()`

**Windows** : fonctionne bien avec RStudio. Les machines d'entreprise verrouillées peuvent imposer des restrictions sur le répertoire temporaire — l'application émet alors un avertissement dans la console mais poursuit son exécution sans persistance des paramètres.

**Linux** : peut nécessiter des bibliothèques système : `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

### 16.4 Dégradation progressive

| Composant manquant                                     | Comportement                                                                                                                   |
|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Pas de LaTeX                                           | Option PDF masquée. HTML et Word restent disponibles.                                                                          |
| Pas d'accès internet / échec du chargement des polices | Police sans empattement du système utilisée. Message consigné dans la console.                                                 |
| Pas de RSQLite / système de fichiers en lecture seule  | Les paramètres ne sont pas conservés. L'application fonctionne normalement.                                                    |
| Répertoire temporaire non accessible en écriture       | La génération du rapport échoue avec un message d'erreur explicite. Définir TMPDIR vers un emplacement accessible en écriture. |

### 16.5 Résolution des problèmes

« **PDF option not available** » : exécuter `tinytex::install_tinytex()` dans R, puis redémarrer l'application.

« **Settings will not persist** » : vérifier les permissions du répertoire de données utilisateur (macOS : `~/Library/Application Support/R/glmnetUI/`, Linux : `~/.local/share/R/glmnetUI/`, Windows : `%APPDATA%/R/data/glmnetUI/`).

**Port 7879 déjà utilisé** : exécuter `lsof -ti:7879 | xargs kill` (macOS/Linux) ou `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell).

## Références

- FRIEDMAN, Jerome, Trevor HASTIE et Robert TIBSHIRANI (2010). « Regularization Paths for Generalized Linear Models via Coordinate Descent ». In : *Journal of Statistical Software* 33.1, p. 1-22. DOI : [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01).
- SIMON, Noah et al. (2011). « Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent ». In : *Journal of Statistical Software* 39.5, p. 1-13. DOI : [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).
- TAY, J. Kenneth, Balasubramanian NARASIMHAN et Trevor HASTIE (2023). « Elastic Net Regularization Paths for All Generalized Linear Models ». In : *Journal of Statistical Software* 106.1, p. 1-31. DOI : [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).

# mgcvUI : une interface interactive pour le package mgcv (modèles additifs généralisés)

*Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

William Bert Craytor

Volume 1, Numéro 1 · Été 2026

Publié le 1<sup>er</sup> juillet 2026

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

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

Cet article en libre accès est diffusé selon les termes de la [licence Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/).

**Résumé.** mgcvUI est une interface graphique pour le package R mgcv, qui ajuste des modèles additifs généralisés (GAM) au moyen de splines de régression pénalisées avec sélection automatique du degré de lissage. Elle propose trois modes d'utilisation — modélisation prédictive générale, évaluation immobilière et analyse de zone de marché — et guide l'utilisateur à travers l'importation des données, la spécification des termes lisses, l'ajustement du modèle, les graphiques de diagnostic et d'effets, ainsi que les rapports téléchargeables. Cet article documente les exigences de format de données de mgcvUI, le flux de travail de modélisation, les affichages de sortie et la référence complète des fonctionnalités.

**Mots-clés :** modèles additifs généralisés, GAM, splines pénalisées, lissage, mgcv, R, interface graphique, évaluation immobilière

**Note de l'éditeur (juillet 2026) :** Peu avant la mise sous presse, mgcvUI a été intégré à earthUI (version 0.11.0 et ultérieures), qui fournit désormais le flux de travail GAM décrit ici aux côtés des flux de travail MARS et elastic-net. mgcvUI demeure disponible dans sa dernière version publiée (0.3.0) à l'adresse <https://github.com/wcraytor/mgcvUI> ; les développements futurs se poursuivent dans earthUI. Le présent article s'applique à mgcvUI tel que publié.

## Table des matières

|          |                                                         |          |
|----------|---------------------------------------------------------|----------|
| <b>1</b> | <b>Introduction</b>                                     | <b>4</b> |
| 1.1      | Historique et contexte                                  | 4        |
| 1.1.1    | Modèles additifs généralisés                            | 4        |
| 1.1.2    | Le package mgcv                                         | 4        |
| 1.1.3    | Comparaison avec MARS (earth) et l'elastic net (glmnet) | 5        |
| 1.1.4    | La chaîne earthUI-mgcvUI                                | 5        |
| 1.1.5    | Flux de travail recommandé                              | 6        |
| 1.2      | Qu'est-ce que mgcvUI ?                                  | 6        |
| 1.3      | Trois modes d'utilisation                               | 7        |
| 1.4      | Fonctionnalités propres à l'immobilier                  | 7        |
| 1.5      | Premiers pas                                            | 7        |
| <b>2</b> | <b>Exigences relatives aux données d'entrée MLS</b>     | <b>9</b> |
| 2.1      | Format et structure du fichier                          | 9        |
| 2.2      | Colonnes requises en mode évaluation                    | 9        |
| 2.3      | Conventions de nommage des colonnes spéciales           | 10       |
| 2.4      | Qualité et complétude des données                       | 10       |
| 2.5      | Emplacement de la ligne objet                           | 10       |

|          |                                           |           |
|----------|-------------------------------------------|-----------|
| <b>3</b> | <b>Mode General</b>                       | <b>11</b> |
| 3.1      | Vue d'ensemble                            | 11        |
| 3.2      | Le flux de travail de la barre latérale   | 11        |
| 3.3      | Importation depuis earthUI (facultatif)   | 12        |
| 3.4      | Onglets du panneau principal              | 12        |
| 3.5      | Persistance des réglages                  | 13        |
| 3.6      | Mode sombre                               | 13        |
| 3.7      | Locale et réglages régionaux              | 13        |
| <b>4</b> | <b>Mode Appraisal</b>                     | <b>15</b> |
| 4.1      | Traitement de la ligne objet              | 15        |
| 4.2      | Date d'effet et âge de la vente           | 15        |
| 4.3      | Désignations de colonnes spéciales        | 15        |
| 4.4      | Aperçu des ajustements RCA                | 15        |
| <b>5</b> | <b>Mode Market Area Analysis</b>          | <b>16</b> |
| 5.1      | Différences avec le mode évaluation       | 16        |
| 5.2      | Quand utiliser le mode marché             | 16        |
| <b>6</b> | <b>Sélection des variables</b>            | <b>17</b> |
| 6.1      | Variable cible                            | 17        |
| 6.2      | Transformation de la réponse              | 17        |
| 6.3      | Le tableau des prédicteurs                | 17        |
| 6.4      | Détection et forçage des types de données | 18        |
| 6.5      | Référence des types de colonnes spéciales | 18        |
| <b>7</b> | <b>Choix des paramètres</b>               | <b>20</b> |
| 7.1      | Préréglages de paramètres                 | 20        |
| 7.2      | Famille                                   | 20        |
| 7.3      | Méthode                                   | 20        |
| 7.4      | Gamma (multiplicateur de lissage)         | 21        |
| 7.5      | Validation croisée                        | 21        |
| 7.6      | Select (rétrécissement)                   | 21        |
| 7.7      | Base par défaut                           | 21        |
| 7.8      | k par défaut (dimension de la base)       | 21        |
| 7.9      | Type de produit tensoriel                 | 22        |
| 7.10     | Interactions autorisées                   | 22        |
| 7.11     | Interactions facteur $\times$ terme lisse | 22        |
| 7.12     | Paramètres avancés                        | 23        |
| <b>8</b> | <b>Ajustement du modèle</b>               | <b>24</b> |
| 8.1      | Le bouton d'ajustement                    | 24        |
| 8.2      | Ce qui se passe pendant l'ajustement      | 24        |
| 8.3      | Ajustement en arrière-plan et progression | 24        |
| <b>9</b> | <b>Onglets de résultats</b>               | <b>25</b> |
| 9.1      | Data                                      | 25        |
| 9.2      | Equation                                  | 25        |
| 9.3      | Summary                                   | 25        |
| 9.4      | Variable Importance                       | 25        |
| 9.5      | Graphiques de contribution                | 26        |
| 9.5.1    | Termes lisses univariés — $s(x)$          | 26        |

|           |                                                               |           |
|-----------|---------------------------------------------------------------|-----------|
| 9.5.2     | Termes lisses par facteur — $s(x, \text{by} = \text{factor})$ | 26        |
| 9.5.3     | Termes d'interaction — $ti(x_1, x_2)$ et $te(x_1, x_2)$       | 26        |
| 9.5.4     | Termes paramétriques — linéaires et facteurs                  | 26        |
| 9.6       | Correlation                                                   | 26        |
| 9.7       | Diagnostics                                                   | 27        |
| 9.8       | RCA Adjustments                                               | 27        |
| 9.9       | Anova                                                         | 27        |
| 9.10      | Mgev Output                                                   | 27        |
| 9.11      | Sign Check                                                    | 27        |
| 9.12      | Concurvity                                                    | 27        |
| 9.13      | Basis Check                                                   | 28        |
| <b>10</b> | <b>Téléchargement des données</b>                             | <b>29</b> |
| 10.1      | Colonnes de sortie                                            | 29        |
| 10.2      | Scores CQA                                                    | 29        |
| <b>11</b> | <b>Calculs RCA et téléchargement</b>                          | <b>30</b> |
| 11.1      | Ouverture de la boîte de dialogue RCA                         | 30        |
| 11.2      | Interpolation du score CQA                                    | 30        |
| 11.3      | Colonnes de sortie                                            | 30        |
| <b>12</b> | <b>Grille de comparaison des ventes</b>                       | <b>31</b> |
| 12.1      | Boîte de dialogue de sélection des comparables                | 31        |
| 12.2      | Structure du classeur                                         | 31        |
| 12.3      | Protection des feuilles et cellules résiduelles               | 31        |
| <b>13</b> | <b>Téléchargement des rapports</b>                            | <b>32</b> |
| 13.1      | Étape 1 : générer le rapport Quarto                           | 32        |
| 13.2      | Étape 2 : convertir le rapport Quarto                         | 32        |
| 13.3      | Contenu du rapport                                            | 32        |
| <b>14</b> | <b>Export des fonctions du modèle</b>                         | <b>34</b> |
| 14.1      | Fonctions R                                                   | 34        |
| 14.2      | Fonctions Python                                              | 34        |
| 14.3      | En-tête C++                                                   | 34        |
| 14.4      | Base de données SQLite                                        | 34        |
| <b>15</b> | <b>Comparaison avec earthUI et glmnetUI</b>                   | <b>36</b> |
| 15.1      | Principales différences                                       | 36        |
| 15.2      | Fonctionnalités communes                                      | 36        |
| 15.3      | Flux de travail recommandé                                    | 37        |
| <b>16</b> | <b>Jeu de données de démonstration : Appraisal_1.csv</b>      | <b>38</b> |
| 16.1      | Description                                                   | 38        |
| 16.2      | Colonnes                                                      | 38        |
| 16.3      | Prise en main rapide suggérée                                 | 39        |
| <b>17</b> | <b>Configuration requise et dépannage</b>                     | <b>40</b> |
| 17.1      | Plateformes prises en charge                                  | 40        |
| 17.2      | Version de R                                                  | 40        |
| 17.3      | Packages R requis                                             | 40        |
| 17.4      | Dépendances facultatives                                      | 41        |
| 17.5      | Remarques propres à chaque plateforme                         | 41        |

|        |                                                  |    |
|--------|--------------------------------------------------|----|
| 17.5.1 | macOS                                            | 41 |
| 17.5.2 | Windows                                          | 41 |
| 17.5.3 | Linux                                            | 41 |
| 17.6   | Dégradation progressive                          | 42 |
| 17.7   | Dépannage                                        | 42 |
| 17.7.1 | « PDF option is not available »                  | 42 |
| 17.7.2 | « Settings will not persist »                    | 42 |
| 17.7.3 | « Diagnostics plot unavailable »                 | 43 |
| 17.7.4 | « Report error : number of items to replace... » | 43 |
| 17.7.5 | Le port 7880 est déjà utilisé                    | 43 |
| 17.7.6 | L'installation des packages échoue sous Linux    | 43 |

## 1 Introduction

### 1.1 Historique et contexte

#### 1.1.1 Modèles additifs généralisés

Les **modèles additifs généralisés** (GAM, *Generalized Additive Models*) ont été formalisés par Trevor Hastie et Robert Tibshirani dans leur monographie de 1990. Les GAM étendent le modèle linéaire en remplaçant chaque terme linéaire  $\beta_j x_j$  par une fonction lisse  $f_j(x_j)$ , de sorte que le modèle devient :

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

Chaque  $f_j$  est estimée à partir des données au moyen d'une spline de régression pénalisée. Les fonctions lisses permettent au modèle de capter des relations non linéaires sans que l'utilisateur ait à spécifier une forme fonctionnelle à l'avance. La pénalité contrôle l'ondulation de chaque terme lisse, évitant le surajustement tout en laissant assez de souplesse pour suivre les structures réelles.

#### 1.1.2 Le package mgcv

Le package R `mgcv` (Mixed GAM Computation Vehicle) a été développé par Simon Wood à l'université de Bath (WOOD 2017). Publié pour la première fois en 2001, il est aujourd'hui l'implémentation R de référence des GAM et est fourni avec toute installation de R. Les méthodes sous-jacentes sont décrites dans une série d'articles portant sur l'estimation stable de paramètres de lissage multiples (WOOD 2004), les splines de régression à plaque mince (WOOD 2003), l'estimation REML rapide et stable (WOOD 2011) et la sélection de modèles lisses généraux (WOOD, PYA et SÄFKEN 2016). Parmi les fonctionnalités clés :

- **Sélection automatique du degré de lissage** : le paramètre de lissage de chaque terme est estimé dans le cadre même de l'ajustement du modèle, par REML (maximum de vraisemblance restreint), GCV (validation croisée généralisée) ou ML (maximum de vraisemblance).
- **Types de bases multiples** : splines de régression à plaque mince (tp), splines de régression cubiques (cr), P-splines (ps) et B-splines (bs), chacune ayant des propriétés différentes.
- **Lissages par produit tensoriel** : `te()` et `ti()` permettent des interactions lisses entre variables sans subir le fléau de la dimension.
- **Sélection de variables** : l'option `select = TRUE` ajoute une pénalité supplémentaire capable de ramener des termes lisses à zéro, ce qui revient à effectuer une sélection de variables.

- **Diagnostics de concurvit ** : mesure l’analogue de la multicolin arit  pour les termes lisses.

### 1.1.3 Comparaison avec MARS (earth) et l’elastic net (glmnet)

L’application earth et ses deux Post Processing Modules (modules de post-traitement) glmnet et mgcv reposent sur des moteurs de mod lisation diff rents. Le tableau suivant r sume les principales diff rences :

| Caract ristique                  | mgcv (GAM)                                                                      | earth (MARS)                                                                               | glmnet (elastic net)                                                     |
|----------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|
| <b>Type de mod le</b>            | Non lin aire lisse (splines p nalis es)                                         | Lin aire par morceaux (splines adaptatives)                                                | Lin aire (avec r gularisation)                                           |
| <b>Non-lin arit </b>             | Automatique, par fonctions de lissage (thin plate, cubique, etc.)               | Automatique, par fonctions charni res   n uds d duits des donn es                          | Uniquement par interactions ou expansion de base                         |
| <b>S lection de variables</b>    | Facultative, par double p nalit  de r tr cissement ( <code>select=TRUE</code> ) | Int gr e, par s lection pas   pas avant/arri re avec  lagage GCV                           | Int gr e, via la p nalit  $L_1$ (Lasso)                                  |
| <b>Interactions</b>              | Lissages par produit tensoriel ( <code>te()</code> , <code>ti()</code> )        | Produits de charni res (interpr tables, jusqu’au degr  3)                                  | Produits crois s deux   deux (difficiles   interpr ter)                  |
| <b>Interpr tabilit </b>          | Courbes d’effets partiels lisses ; tr s intuitives                              | Les g-fonctions montrent des effets partiels lin aires par morceaux, variable par variable | Les coefficients sont des poids lin aires, faciles   expliquer un par un |
| <b>Domaine de pr dilection</b>   | Relations non lin aires lisses, mod lisation souple                             | D tection automatique des non-lin arit s et des interactions                               | Donn es de grande dimension, s lection de variables, petits  chantillons |
| <b>Ajustements RCA</b>           | Ajustements lisses via les effets partiels                                      | Ajustements lin aires par morceaux via les g-fonctions                                     | Ajustements lin aires par variable                                       |
| <b>Contr le du surajustement</b> | Estimation du param tre de lissage par REML/GCV                                 |  lagage fond  sur le GCV                                                                   | Lambda par validation crois e                                            |

**MARS** (Multivariate Adaptive Regression Splines) a  t  introduit par Jerome Friedman en 1991. La m thode construit des mod les lin aires par morceaux en s lectionnant de mani re adaptative des fonctions charni res et la position de leurs n uds   partir des donn es. L’impl mentation R est le package **earth** de Stephen Milborrow.

La r gression **elastic net**, mise en  uvre dans le package **glmnet** de Friedman, Hastie et Tibshirani (2010), combine les p nalit s Lasso ( $L_1$ ) et ridge ( $L_2$ ) pour effectuer simultan ment la s lection de variables et le r tr cissement des coefficients.

### 1.1.4 La cha ne earthUI–mgcvUI

Un flux de travail puissant associe la d couverte automatique des n uds par earthUI   l’estimation lisse de mgcvUI. Lorsque earthUI exporte un fichier de r sultats `.rds` et que mgcvUI l’importe :

1. **Earth découvre les nœuds** — MARS détermine, pour chaque variable, des positions de charnières (points de rupture) déduites des données.
2. **mgcvUI les utilise comme ancrages** — les nœuds d'earth deviennent la base des splines de régression cubiques (base `cr`) du GAM, ce qui donne aux termes lisses une longueur d'avance grâce à des positions de nœuds reflétant la structure réelle des données.
3. **mgcv affine l'ajustement** — l'estimation par splines pénalisées lisse l'ajustement linéaire par morceaux d'earth et produit des courbes d'effets partiels lisses, tout en préservant la forme générale découverte par earth.

Cette chaîne fait le lien entre modélisation exploratoire (earth) et modélisation confirmatoire (GAM), en combinant le meilleur des deux approches.

### 1.1.5 Flux de travail recommandé

Pour l'évaluation immobilière et les applications similaires exigeant des modèles interprétables et défendables :

1. **Commencez par earthUI** pour découvrir les variables importantes, les relations non linéaires et les interactions présentes dans vos données. La sélection automatique de base d'earth et ses g-fonctions fournissent d'excellents éclairages initiaux.
2. **Affinez avec mgcvUI** si vous souhaitez des effets partiels plus lisses. Importez les nœuds d'earthUI dans mgcvUI pour passer sans heurt des modèles linéaires par morceaux aux modèles lisses.
3. **Utilisez glmnetUI** lorsque vous avez besoin d'une sélection de variables agressive (Lasso), lorsque le nombre de prédicteurs approche ou dépasse le nombre d'observations, ou lorsque vous voulez un modèle purement linéaire avec régularisation, pour des raisons de défendabilité.

## 1.2 Qu'est-ce que mgcvUI ?

mgcvUI est une interface graphique pour le package R `mgcv`. Elle s'exécute comme une application Shiny locale — sans identification, sans serveur et sans compte utilisateur. Vous la lancez depuis R, vous importez un jeu de données (CSV ou Excel), vous configurez votre modèle et vous ajustez un GAM de manière interactive.

Dans la version actuelle, mgcvUI n'est plus un package installé séparément : il est livré à l'intérieur du package earthUI, dont il constitue le Post Processing Module (module de post-traitement) GAM, démarré par `earthUI::launch_mgcv()`. L'application elle-même — et tout ce qui est décrit dans cet article — demeure inchangée ; seules l'installation et la commande de lancement diffèrent. Les trois routines (earth, glmnet, mgcv) partagent une même arborescence de projet, une même base de données de réglages et un même exemplaire de l'infrastructure de support.

L'application offre un flux de travail complet : importation des données, configuration des variables avec spécification des termes lisses, ajustement du modèle avec traitement en arrière-plan, graphiques de diagnostic, courbes d'effets partiels lisses et rapports téléchargeables aux formats Word, PDF ou HTML.

Les **modèles additifs généralisés** remplacent la relation linéaire  $\beta x$  par une fonction lisse  $f(x)$  pour chaque prédicteur. Cela signifie :

- **Aucune forme fonctionnelle à spécifier** — ce sont les données qui déterminent la forme de chaque relation.
- **Des effets partiels lisses** — la contribution de chaque prédicteur est une courbe lisse et non une droite, ce qui permet de voir facilement comment une variable influe sur la réponse selon ses différentes valeurs.

- **Sélection automatique du degré de lissage** — `mgcv` détermine par REML ou GCV le degré d'ondulation de chaque courbe, en équilibrant qualité d'ajustement et surajustement.

### 1.3 Trois modes d'utilisation

Chaque projet a un **objet** (*purpose*), choisi lors de sa création à la section 1 de la barre latérale (**Project**, projet) puis verrouillé. L'objet du projet actif détermine les outils et éléments d'interface qui apparaissent ; pour travailler dans un autre mode, créez ou ouvrez un projet ayant cet objet. Les trois modes sont :

- **General** (général) — régression GAM pour tout type de population ou de jeu de données. C'est le mode par défaut. Il offre le flux de travail GAM complet, sans ajout propre à un domaine.
- **For Appraisal** (pour l'évaluation) — régression GAM adaptée à l'évaluation immobilière. Ajoute des fonctionnalités propres à l'évaluation d'un bien unique : traitement du bien objet, désignations de colonnes spéciales et approche par contrainte des résidus (RCA, *Residual Constraint Approach*).
- **Market Area Analysis** (analyse de zone de marché) — régression GAM adaptée aux études de zone de marché. Ajoute des fonctionnalités permettant d'analyser des groupes de biens dans un marché défini, avec traitement facultatif d'une ligne objet.

Dans les trois modes, le moteur de modélisation est identique — vous ajustez toujours un GAM au moyen de `mgcv::gam()`. Le réglage de l'objet détermine seulement quels outils et éléments d'interface supplémentaires sont disponibles.

### 1.4 Fonctionnalités propres à l'immobilier

Lorsque **For Appraisal** ou **Market Area Analysis** est sélectionné, `mgcvUI` active plusieurs fonctionnalités conçues pour l'analyse immobilière :

- **Désignations de colonnes spéciales** — chaque prédicteur peut recevoir un rôle spécial tel que `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Ces désignations déterminent la manière dont la colonne est traitée lors de l'ajustement et dans les sorties.
- **Colonne Sale Age** (âge de la vente) — lorsqu'une colonne est désignée `contract_date` et qu'une date d'effet est fournie, `mgcvUI` calcule une colonne `sale_age` (nombre de jours entre la date de vente et la date d'effet) et la substitue comme prédicteur.
- **Calculs RCA (mode évaluation uniquement)** — en mode évaluation, après l'ajustement du modèle, `mgcvUI` peut produire les sorties de l'approche par contrainte des résidus (RCA) : ajustements par comparable, synthèses des ajustements nets et bruts et prix de vente ajusté du bien objet.

**Astuce :** Le mode General convient à tout jeu de données — financier, scientifique, technique ou immobilier. Les modes évaluation et marché ne font qu'ajouter des commodités destinées aux professionnels de l'immobilier.

### 1.5 Premiers pas

Pour utiliser `mgcvUI` :

1. **Installez earthUI** dans R : `install.packages("earthUI")` — la routine `mgcv` y est incluse. (Voir « Installing earthUI and vProlog » dans l'article `earthUI` du présent numéro pour la version de développement et les composants facultatifs.)

2. **Lancez l'application** : exécutez `earthUI::launch_mgcv()` dans R. L'application s'ouvre dans votre navigateur sur le port 7880.
3. **Créez ou ouvrez un projet** à la section 1 de la barre latérale. Un projet fixe l'objet (General, For Appraisal ou Market Area Analysis) et le lieu de travail (pays, état, comté, ville) ainsi que ses dossiers d'entrée et de sortie. Les projets sont stockés sous une racine `regProj` commune et sont partagés avec les applications sœurs earthUI et glmnetUI. Ouvrir un projet ayant un autre objet est la façon de changer de mode.
4. **Importez vos données** à la section 2 en sélectionnant un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajoutez des fichiers à ce dossier, puis cliquez sur **Refresh** (actualiser) pour les lister.
5. **Importez depuis earthUI (facultatif)** — la section 3 permet d'importer un fichier de résultats `.rds` d'earthUI afin d'amorcer le GAM avec les nœuds découverts par earth. Ignorez cette étape si vous n'avez pas de résultat earthUI.
6. **Configurez les variables** — choisissez votre cible et vos prédicteurs, définissez les types de données, attribuez les éventuels rôles de colonne spéciaux et marquez les variables comme facteur ou linéaire.
7. **Réglez les paramètres mgcv** — famille, méthode, gamma, type de base, k, type de produit tensoriel, interactions et autres options.
8. **Ajustez le modèle** — cliquez sur « Fit Mgcv GAM Model » (ajuster le modèle GAM mgcv) et examinez les résultats dans le panneau principal.
9. **Exportez** — téléchargez les prédictions au format Excel, générez et convertissez un rapport Quarto ou, en mode évaluation, calculez les ajustements RCA et une grille de ventes.

Les réglages sont automatiquement conservés dans une base de données SQLite et restaurés lorsque vous rechargez le même fichier d'entrée.

## 2 Exigences relatives aux données d'entrée MLS

Pour les flux de travail d'évaluation immobilière et d'analyse de marché, vos données d'entrée proviennent généralement d'un export de Multiple Listing Service (MLS). Ce chapitre décrit la structure de fichier attendue et les colonnes que mgcvUI sait exploiter.

### 2.1 Format et structure du fichier

mgcvUI accepte les fichiers **CSV** et **Excel** (.xlsx, .xls). À l'importation, les noms de colonnes sont automatiquement convertis en **snake\_case** — par exemple, « Living SqFt » devient **living\_sqft**, « Contract Date » devient **contract\_date** et « Sale Price » devient **sale\_price**. Cette normalisation garantit des références de colonnes cohérentes dans tout le flux de travail. Le séparateur CSV et le symbole décimal utilisés à l'importation sont déterminés par les réglages de locale (voir le chapitre 3, « Locale & Regional Settings »).

Votre fichier de données doit être un tableau plat comportant une ligne par bien et une colonne par attribut. La première ligne du fichier doit contenir les en-têtes de colonnes.

### 2.2 Colonnes requises en mode évaluation

Bien que mgcvUI fonctionne avec n'importe quel ensemble de colonnes, le flux de travail complet d'évaluation tire profit de la présence des colonnes suivantes :

| Colonne          | Type spécial           | Rôle                                                                              |
|------------------|------------------------|-----------------------------------------------------------------------------------|
| Sale Price       | (cible)                | Variable réponse du modèle                                                        |
| Contract Date    | <b>contract_date</b>   | Sert à calculer <b>sale_age</b> (jours écoulés depuis la date d'effet)            |
| Listing Date     | <b>listing_date</b>    | Utilisée avec la date de contrat pour calculer le DOM en l'absence de colonne DOM |
| Days on Market   | <b>dom</b>             | Jours sur le marché ; affichée dans les exports                                   |
| Concessions      | <b>concessions</b>     | Concessions de vente ; SP net = prix de vente – concessions                       |
| Living Area (SF) | <b>living_area</b>     | Active les résidus au pied carré ( <b>residual_sf</b> , <b>cqa_sf</b> )           |
| Lot Size         | <b>lot_size</b>        | Colonne de superficie du terrain                                                  |
| Site Dimensions  | <b>site_dimensions</b> | Regroupée avec la superficie du terrain                                           |
| Latitude         | <b>latitude</b>        | Arrondie à 3 décimales                                                            |
| Longitude        | <b>longitude</b>       | Arrondie à 3 décimales                                                            |
| Area ID          | <b>area</b>            | Identifiant de zone de marché / quartier                                          |
| Actual Age       | <b>actual_age</b>      | Âge du bien                                                                       |
| Effective Age    | <b>effective_age</b>   | Âge effectif du bien                                                              |
| Address          | <b>display_only</b>    | Affichée dans les exports ; exclue du modèle                                      |

Les noms de colonnes du tableur peuvent être dans une langue étrangère — les noms « spéciaux » sont en anglais afin que le programme R puisse leur appliquer un traitement particulier. Pour le reste, les noms de colonnes fournis apparaissent tels quels dans les modèles de régression, les graphiques et, si vous réalisez des évaluations, les rapports de sortie.

Toutes les colonnes ne sont pas obligatoires. mgcvUI s'adapte — si une colonne est absente, la fonctionnalité correspondante est simplement omise. Cependant, pour les modèles de prix immobiliers, certaines colonnes sont vivement recommandées :

1. **Sale Age** — le nombre de jours entre la date de vente contractuelle et la date d'effet. Lorsqu'on utilise un historique de ventes pluriannuel, `sale_age` joue souvent un rôle central.
2. **Living Area** — appelée aussi « Living Sqft » ou « GLA » (*gross living area*, surface habitable brute).
3. **Total Bath Count** — le nombre total de salles de bains complètes, aux trois quarts, à moitié et au quart.
4. **Garage Bays** ou **Garage Area** — nombre de places de garage ou surface du garage.
5. **Lot Size** — la superficie du terrain du bien.
6. **Longitude, Latitude** et, si disponible, **Area ID** pour la variation géographique des prix.

## 2.3 Conventions de nommage des colonnes spéciales

mgcvUI identifie les colonnes par leur **désignation de type spécial**, et non par leur nom. Vous pouvez nommer vos colonnes comme vous le souhaitez dans l'export MLS — ce qui compte, c'est de leur attribuer le bon type spécial dans le tableau de configuration des variables (chapitre 6).

## 2.4 Qualité et complétude des données

- **Valeurs manquantes (NA)** : les lignes comportant des valeurs NA dans un prédicteur ou dans la colonne cible sont automatiquement supprimées avant l'ajustement. mgcvUI restreint les données aux seules colonnes utilisées par le modèle, de sorte que des NA dans des colonnes inutilisées n'entraînent aucune perte de ligne.
- **Colonnes numériques** : le prix de vente, la surface habitable, la superficie du terrain et les champs analogues doivent être numériques.
- **Colonnes facteur** : les colonnes texte, TRUE/FALSE et facteur R sont automatiquement traitées comme catégorielles — elles ne peuvent entrer dans le modèle qu'en tant que facteurs. Une colonne **numérique** (par exemple un code numérique de zone ou de style) ne devient un facteur que si vous cochez sa case **Factor** ; mgcvUI ne déduit jamais le caractère catégoriel de l'étendue des valeurs d'une variable numérique, si bien que des prédicteurs numériques discrets comme `bath_count` restent continus (lissés). Le rôle spécial est indépendant de la décision relative au facteur.

**Astuce :** Examinez la colonne NA du tableau des prédicteurs après l'importation. Les colonnes comportant de nombreuses valeurs manquantes peuvent provoquer la suppression de lignes. Les colonnes présentant  $\geq 50\%$  de NA sont signalées en rouge et automatiquement exclues.

## 2.5 Emplacement de la ligne objet

En mode **Appraisal**, la **ligne 1 doit être le bien objet**. Toutes les autres lignes sont des ventes comparables. La ligne objet est exclue de l'ajustement du modèle. Après l'ajustement, le modèle produit malgré tout des prédictions pour la ligne objet.

En mode **Market Area Analysis**, placer l'objet en ligne 1 est facultatif.

En mode **General**, il n'y a pas de traitement particulier des lignes — toutes sont traitées de la même façon.

## 3 Mode General

### 3.1 Vue d'ensemble

Le mode General est le mode par défaut au lancement de mgcvUI. Il offre le flux de travail GAM complet pour n'importe quel jeu de données — pas seulement immobilier. Vous pouvez utiliser mgcvUI pour des données scientifiques, une analyse financière, des études techniques ou tout problème de régression où l'on s'attend à des relations non linéaires lisses.

En mode General, l'interface omet les fonctionnalités propres à l'immobilier (colonnes spéciales, âge de la vente, RCA). La barre latérale est simplifiée et centrée sur la sélection des variables, la configuration des paramètres, l'ajustement du modèle et l'export.

### 3.2 Le flux de travail de la barre latérale

La barre latérale est organisée en sections numérotées et repliables qui vous guident depuis le choix du projet jusqu'à l'export du rapport. La plupart des sections n'apparaissent qu'une fois un projet actif ouvert, et plusieurs sont propres à l'objet du projet. Le dossier de sortie découle du projet actif — il n'y a pas de champ distinct pour le dossier de sortie.

**1. Project** (projet) — créez ou ouvrez un projet. Un projet fixe l'objet et le lieu de travail (pays, état, comté, ville) ainsi que ses dossiers d'entrée (`<os>_in/`) et de sortie. Les projets sont stockés sous une racine `regProj` commune (par défaut `~/regProj`, modifiable au moyen de la variable d'environnement `REGPROJ_ROOT` ou du menu Settings) et sont partagés avec les applications sœurs `earthUI` et `glmnetUI`, de sorte que les modèles des trois applications coexistent. Cliquez sur **Close Project** (fermer le projet) pour changer.

**2. Import Data** (importer les données) — sélectionnez un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajoutez des fichiers au dossier via le Finder ou l'Explorateur, puis cliquez sur **Refresh** pour les lister. Pour les fichiers Excel comportant plusieurs feuilles, un sélecteur de feuille apparaît. Les noms de colonnes sont automatiquement convertis en `snake_case`.

**3. Import from earthUI (facultatif)** — importez un fichier de résultats `.rds` d'`earthUI` pour amorcer le GAM avec les positions de nœuds découvertes par `earth` (et pour préremplir la sélection des variables). Cette étape est facultative — ignorez-la si vous n'avez pas de résultat `earthUI` à importer.

**4. Variable Configuration** (configuration des variables) — sélecteur de variable cible, transformation de la réponse (aucune, log, log10), tableau des prédicteurs avec cases à cocher Include, Factor et Linear. La colonne Special n'apparaît qu'en modes Appraisal et Market. Voir le chapitre 6 pour tous les détails.

**5. MgcV Call Parameters** (paramètres de l'appel mgcv) — toute la configuration du modèle : préréglages de paramètres, famille, méthode, gamma, validation croisée, select, type de base, k, type de produit tensoriel, matrice d'interactions et paramètres avancés. Voir le chapitre 7 pour la référence complète des paramètres.

**6. Fit MgcV GAM Model** (ajuster le modèle GAM mgcv) — le bouton qui lance le modèle. Les résultats apparaissent dans les onglets de droite.

**7. Download Output** (télécharger les sorties) — exporte les prédictions, les résidus, les scores CQA et les contributions par variable dans un fichier Excel placé dans le dossier de sortie du projet. Disponible dans tous les modes une fois le modèle ajusté.

**8. Calculate RCA Adjustments & Download** (calculer et télécharger les ajustements RCA) — calcule les ajustements RCA par variable pour chaque comparable et interpole le résidu de l'objet par CQA. N'apparaît qu'en modes Appraisal et Market, après l'ajustement.

**9. Generate Sales Grid & Download** (générer et télécharger la grille de ventes) — génère la grille de comparaison des ventes, en recommandant les comparables d'après le pourcentage d'ajustement brut. N'apparaît qu'en mode Appraisal.

**10. Generate Quarto Report** (générer le rapport Quarto) — écrit un ensemble de rapport Quarto (`.qmd`) autonome — avec tous les graphiques et tableaux — dans le dossier de sortie du projet. Disponible une fois le modèle ajusté.

**11. Convert Quarto Report** (convertir le rapport Quarto) — rend l'ensemble `.qmd` généré en HTML, Word ou PDF. Le PDF exige une installation LaTeX ; si aucune n'est détectée, l'option PDF est masquée. Voir le chapitre 13.

Les étapes propres à un objet sont omises (et les étapes restantes renumérotées) dans les autres modes : en mode Market, l'étape de la grille de ventes est absente et, en mode General, les étapes RCA et grille de ventes le sont toutes deux.

### 3.3 Importation depuis earthUI (facultatif)

La section 3 de la barre latérale — **Import from earthUI** — permet d'importer un fichier de résultats `.rds` d'earthUI. Cela active la chaîne earth-mgcv : les positions de nœuds déduites des données par earth deviennent des points d'ancrage pour les termes lisses du GAM.

Lorsqu'un résultat earthUI est importé, mgcvUI :

- affiche une synthèse indiquant la variable cible, le R-carré, le nombre de termes et les positions de nœuds par variable ;
- sélectionne automatiquement le pré-réglage de paramètres « Earth Pipeline » (base de splines de régression cubiques,  $\gamma = 1,4$ ) ;
- fixe la variable cible pour qu'elle corresponde à celle d'earth ;
- précoche Include pour les variables prédictrices d'earth ;
- précoche Factor pour les variables catégorielles d'earth et Linear pour ses prédicteurs linéaires ;
- préremplit les interactions détectées par earth (surlignées en jaune dans la matrice d'interactions et verrouillées contre toute modification manuelle) ;
- fournit un bouton **Export Knots CSV** permettant de télécharger les nœuds d'earth au format CSV.

Le tableau de configuration des variables affiche alors une note expliquant que les sélections de variables (Include, Linear, Factor) et les nœuds de splines ont été *amorçés* à partir du modèle earth. Contrairement à glmnetUI, où l'ajustement glmnet est lié à la base d'earth, mgcvUI ajuste le GAM à partir de vos sélections courantes — earth ne fournit qu'une longueur d'avance, et tous les contrôles restent modifiables.

Un bouton **Clear** (effacer) supprime les données earth importées et rétablit le mode autonome.

### 3.4 Onglets du panneau principal

Après l'importation des données, le panneau principal propose les onglets suivants (les onglets dépendant du modèle se remplissent après l'ajustement) :

| Onglet              | Contenu                                                                                                                                                     |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data                | Tableau interactif (DataTable) des données importées. En modes évaluation et marché, scindé en tableaux Bien objet et Ventes comparables.                   |
| Equation            | Formule GAM complète rendue en MathJax, avec un tableau des définitions des fonctions lisses.                                                               |
| Summary             | Métriques clés ( $R^2$ , $R^2$ de validation croisée, déviance expliquée, AIC, n, nombre de termes lisses) et tableaux des termes lisses et paramétriques.  |
| Variable Importance | Diagramme en barres horizontales des statistiques F (termes lisses) et des valeurs $ t $ (termes paramétriques), avec tableau classé.                       |
| Contribution        | Courbes interactives (plotly) d'effets partiels lisses, avec bandes de confiance et repères des nœuds earth.                                                |
| Correlation         | Carte de chaleur des corrélations entre prédicteurs numériques (disponible avant l'ajustement).                                                             |
| Diagnostics         | Résidus vs valeurs ajustées, diagramme quantile-quantile, histogramme, réponse vs valeurs ajustées (via <b>gratia</b> ). Nuage de points observé vs prédit. |
| RCA Adjustments     | Histogrammes de l'analyse RCA (modes évaluation/marché, après calcul).                                                                                      |
| Anova               | Tableau ANOVA combinant les termes paramétriques et lisses.                                                                                                 |
| MgcV Output         | Texte brut : durée d'exécution, formule, synthèse du modèle, famille, concurvit.                                                                            |
| Sign Check          | Contrôle de cohérence de direction entre earth et le GAM (lorsque earth a été importé).                                                                     |
| Concurvity          | Matrices de concurvit globale et par paires, avec code couleur.                                                                                             |
| Basis Check         | Tests d'adéquation de la dimension de base fournis par <b>gam.check()</b> .                                                                                 |

### 3.5 Persistance des réglages

mgcvUI enregistre automatiquement votre configuration dans une base de données SQLite, indexée par le nom du fichier d'entrée. Lorsque vous rechargez le même fichier, tous les réglages sont restaurés : choix de la cible, cases à cocher des prédicteurs, types de données, paramètres mgcv et matrice d'interactions. L'objet est une propriété du projet actif plutôt qu'un réglage global, et le dernier fichier d'entrée utilisé est mémorisé par projet puis chargé automatiquement à la réouverture du projet.

### 3.6 Mode sombre

Cliquez sur l'icône lune/soleil dans le coin supérieur droit pour basculer entre les thèmes Nord Light et Nord Dark. La préférence de thème est enregistrée dans le localStorage et persiste d'une session à l'autre. Tous les éléments de l'interface s'adaptent au thème sélectionné.

### 3.7 Locale et réglages régionaux

mgcvUI prend en charge les conventions internationales de formatage des nombres et des CSV au moyen d'un système de locale fondé sur le pays. Le menu **Settings** de la barre de titre propose des sélecteurs **Country** (pays) et **Paper** (papier) pour plus de 30 pays. Chaque préréglage configure :

- le **séparateur CSV** — virgule (,) pour les États-Unis, le Royaume-Uni et le Japon, ou point-virgule (;) pour la majeure partie de l'Europe ;
- le **symbole décimal** — point (.) ou virgule (,) ;

- **le séparateur de milliers** — virgule (États-Unis/Royaume-Uni/Japon), point (Allemagne/Italie/Espagne), espace (Finlande/France/Pologne/pays baltes) ou apostrophe (Suisse) ;
- **le format de papier** — Letter (États-Unis/Canada/Mexique) ou A4 (partout ailleurs).

Cliquez sur **Save** (enregistrer) pour conserver globalement vos préférences de locale. La section Import Data comporte également un sélecteur **Locale** propre à chaque fichier, pour les importations ponctuelles suivant d'autres conventions. Le même menu Settings contient le champ **regProj root folder** (dossier racine regProj), qui déplace l'arborescence de projet partagée par mgcvUI, earthUI et glmnetUI.

***Astuce :** Le formatage des nombres sur les axes des graphiques et sur les étiquettes de pente s'adapte automatiquement à votre locale. La locale allemande utilise le point pour les milliers (200.000), la finlandaise l'espace (200 000) et la suisse l'apostrophe (200'000). Aucun symbole monétaire n'est affiché — mgcvUI est indépendant de la devise.*

## 4 Mode Appraisal

Lorsque l'objet du projet actif est **For Appraisal**, mgcvUI se configure pour l'évaluation d'un bien unique. Toutes les fonctionnalités décrites au chapitre 3 restent disponibles ; le présent chapitre ne traite que des ajouts propres à l'évaluation.

### 4.1 Traitement de la ligne objet

En mode évaluation, **la ligne 1 de votre jeu de données est le bien objet** et toutes les autres lignes sont des ventes comparables. Votre fichier d'entrée doit être organisé en conséquence (voir le chapitre 2). Le prix de vente de l'objet peut être laissé vide ou fixé à n'importe quelle valeur — mgcvUI le traite automatiquement comme NA lors de l'ajustement.

Après l'ajustement, le modèle produit des prédictions pour la ligne objet, ce qui constitue la base du flux de travail d'ajustement RCA.

### 4.2 Date d'effet et âge de la vente

En modes évaluation et marché, un champ **Effective Date** (date d'effet) apparaît dans la section Variable Configuration (avec la date du jour par défaut). Si vous désignez une colonne comme `contract_date` dans le menu Special, mgcvUI calcule une colonne `sale_age` — le nombre entier de jours entre la date de contrat de chaque vente et la date d'effet. Cette colonne est ajoutée comme prédicteur.

Lorsque la date d'effet change, `sale_age` est automatiquement recalculée.

### 4.3 Désignations de colonnes spéciales

En modes évaluation et marché, un menu **Special** apparaît pour chaque prédicteur dans le tableau de configuration des variables. Voir le chapitre 6 pour la référence complète des types spéciaux et de leurs effets.

### 4.4 Aperçu des ajustements RCA

Le bouton **Calculate RCA Adjustments & Download** (visible en modes évaluation et marché après l'ajustement) calcule, pour chaque comparable, des ajustements dérivés du marché par rapport à l'objet. Le flux de travail RCA complet est décrit au chapitre 11.

***Astuce :** Le score CQA que vous attribuez à l'objet détermine la part de la distribution des résidus qui lui est imputée. Un score de 5,00 place l'objet à la médiane des comparables.*

## 5 Mode Market Area Analysis

Lorsque l'objet du projet actif est **Market Area Analysis**, mgcvUI offre les mêmes fonctionnalités propres à l'immobilier que le mode évaluation (colonnes spéciales, âge de la vente), mais il est orienté vers l'analyse d'un groupe de biens plutôt que vers l'évaluation d'un objet unique.

### 5.1 Différences avec le mode évaluation

- **Le traitement de la ligne objet est souple** — en mode marché, la ligne 1 n'est pas automatiquement considérée comme un bien objet.
- **Pas de grille de ventes** — l'étape « Generate Sales Grid & Download » n'est pas disponible.

### 5.2 Quand utiliser le mode marché

Le mode Market Area Analysis convient lorsque vous :

- construisez un modèle GAM pour un quartier ou une zone de marché afin de comprendre les déterminants non linéaires de la valeur ;
- analysez la manière dont des variables telles que la surface, l'âge, la superficie du terrain et la localisation influent sur les prix de vente au sein d'un groupe de biens ;
- préparez les éléments justificatifs d'une analyse des conditions de marché ou d'une délimitation de quartier ;
- travaillez sur un jeu de données pour lequel vous souhaitez les fonctionnalités de colonnes spéciales (âge de la vente) sans avoir besoin du flux de travail de la grille de ventes.

**Astuce :** Le mode marché est également utile pour la régression immobilière générale lorsque vous souhaitez les fonctionnalités de colonnes spéciales sans avoir besoin du flux de travail d'évaluation d'un bien unique.

## 6 Sélection des variables

La section 4 de la barre latérale — **Variable Configuration** — est l'endroit où vous choisissez quelles colonnes participent au modèle et comment elles sont traitées.

### 6.1 Variable cible

Le menu **Target (response) variable** (variable cible ou réponse) situé en haut de la section 4 liste chaque colonne numérique de votre jeu de données. Sélectionnez une colonne comme variable réponse (par exemple le prix de vente). La colonne cible est automatiquement exclue de la liste des prédicteurs.

### 6.2 Transformation de la réponse

Sous le sélecteur de cible, un menu **Response Transform** (transformation de la réponse) propose trois options :

- **None (raw)** (aucune) — utilise la réponse telle quelle. Approprié lorsque la relation entre les prédicteurs et la réponse est approximativement additive sur l'échelle d'origine.
- **Log (natural)** (logarithme naturel) — applique une transformation logarithmique népérienne avant l'ajustement. Après transformation inverse, les ajustements deviennent proportionnels (%) plutôt qu'absolus (\$). Particulièrement utile pour les prix de vente, où un ajustement de 10 000 \$ n'a pas du tout la même portée selon le niveau de prix.
- **Log10** — applique une transformation logarithmique en base 10.

***Astuce :** La transformation logarithmique rend les ajustements temporels proportionnels (%) au lieu d'absolus (\$). Pour les modèles immobiliers couvrant une large plage de prix, les transformations logarithmiques produisent souvent des résidus mieux comportés.*

Lorsqu'une transformation logarithmique est sélectionnée, les valeurs  $\leq 0$  de la réponse sont automatiquement écartées. Toutes les sorties (prédictions, contributions, résidus) subissent automatiquement la transformation inverse pour revenir à l'échelle d'origine.

### 6.3 Le tableau des prédicteurs

Sous le sélecteur de cible, un tableau défilant liste toutes les colonnes restantes. Les colonnes de cases à cocher utilisent des en-têtes verticaux pivotés, par souci de compacité. L'ordre des colonnes est le suivant :

| Colonne         | Description                                                                                                                                       |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Variable</b> | Nom de la colonne (police à chasse fixe, tronqué par des points de suspension). Cellule encadrée.                                                 |
| <b>Type</b>     | Menu du type de données : numeric, integer, character, factor, logical, Date, POSIXct. Menu encadré.                                              |
| <b>Include</b>  | Case à cocher — inclure cette colonne comme prédicteur du modèle. En-tête vertical.                                                               |
| <b>Factor</b>   | Case à cocher — traiter comme facteur catégoriel (crée un coefficient distinct par modalité). En-tête vertical.                                   |
| <b>Linear</b>   | Case à cocher — imposer une relation linéaire (pas de terme lisse, simplement $\beta x$ ). En-tête vertical.                                      |
| <b>Special</b>  | Menu (modes évaluation/marché uniquement) — voir la référence des types de colonnes spéciales ci-dessous. Menu encadré.                           |
| <b>NAs</b>      | Nombre et pourcentage de valeurs manquantes, avec code couleur : rouge ( $\geq 50\%$ ), orange ( $\geq 20\%$ ), gris ( $> 0$ ). Cellule encadrée. |

**Facteur, terme lisse ou terme linéaire** : par défaut, les variables numériques incluses reçoivent un terme lisse  $f(x)$ . Cocher **Factor** crée un terme catégoriel (un coefficient par modalité). Cocher **Linear** crée un simple terme linéaire  $\beta x$  au lieu d'un terme lisse. Les variables marquées à la fois Factor et Linear sont traitées comme facteurs.

#### 6.4 Détection et forçage des types de données

mgcvUI détecte automatiquement les types de données à l'importation. Les colonnes numériques, entières, logiques, facteur et date sont reconnues. Vous pouvez forcer n'importe quelle détection en modifiant le menu **Type**. Le changement de type modifie la manière dont la colonne est traitée dans le modèle.

#### 6.5 Référence des types de colonnes spéciales

En modes évaluation et marché, le menu **Special** propose les options suivantes :

##### **Pondération :**

- **weight** — colonne de poids des observations (une seule autorisée ; les lignes de poids nul sont exclues de l'ajustement)

##### **Types de date et d'heure :**

- **contract\_date** — déclenche le calcul automatique de **sale\_age** à partir de la date d'effet
- **sale\_age** — identifie une colonne d'âge de vente déjà calculée (jours écoulés depuis la date d'effet), utilisée telle quelle
- **listing\_date** — utilisée comme solution de repli pour calculer le nombre de jours sur le marché
- **dom** — identifie la colonne des jours sur le marché

##### **Type d'enregistrement :**

- **sale\_type** — identifie la colonne de type d'enregistrement qui distingue l'objet des ventes comparables

##### **Types monétaires :**

- **concessions** — identifie les concessions de vente (crédits du vendeur, incitations à l'acheteur, etc.)

##### **Types de taille et de localisation :**

- **latitude** — valeurs automatiquement arrondies à 3 décimales

- `longitude` — même arrondi que la latitude
- `area` — identifiant de zone de marché ou de quartier
- `living_area` — active les calculs de résidus au pied carré (`residual_sf` et `cqa_sf`)
- `lot_size` — colonne de superficie du terrain
- `site_dimensions` — regroupée avec la superficie du terrain

**Types d'âge :**

- `actual_age` — colonne de l'âge du bien
- `effective_age` — âge effectif du bien

**Types d'affichage :**

- `display_only` — la colonne est incluse dans les exports Excel mais entièrement exclue de l'ajustement du modèle. À utiliser pour les champs d'adresse, les numéros MLS, les identifiants de parcelle ou d'autres données de référence.

***Astuce :** Les colonnes désignées `display_only` demeurent dans votre jeu de données et apparaissent dans l'export Excel, mais elles sont entièrement exclues du tableau des prédicteurs. À utiliser pour les colonnes d'identifiant, les adresses ou d'autres champs de référence.*

## 7 Choix des paramètres

La section 5 de la barre latérale — **Mgcv Call Parameters** — donne accès à toutes les options de configuration du GAM. Chaque paramètre est accompagné d'une icône d'aide bleue (?) affichant une infobulle explicative.

### 7.1 Préréglages de paramètres

Un menu situé en haut propose deux préréglages qui fixent des valeurs par défaut sensées pour les flux de travail courants :

| Préréglage                   | Base                    | Réglages                                  |
|------------------------------|-------------------------|-------------------------------------------|
| Standalone (découverte)      | tp (plaque mince)       | k = auto (10), gamma = 1,2, select = TRUE |
| Earth Pipeline (raffinement) | cr (régression cubique) | k = auto (10), gamma = 1,4, select = TRUE |

Le préréglage **Earth Pipeline** est automatiquement sélectionné lorsque des nœuds earthUI sont importés. La base de splines de régression cubiques (**cr**) est indispensable à l'intégration des nœuds d'earth, car elle seule permet de spécifier des positions de nœuds exactes.

### 7.2 Famille

Choisissez la famille de distribution de votre variable réponse :

| Famille                 | Cas d'usage                                                                     |
|-------------------------|---------------------------------------------------------------------------------|
| <b>gaussian</b>         | Réponses continues (par exemple un prix de vente). La plus courante.            |
| <b>Gamma</b>            | Réponses continues positives dont la variance est proportionnelle à la moyenne. |
| <b>poisson</b>          | Données de comptage (par exemple un nombre de ventes).                          |
| <b>binomial</b>         | Résultats binaires.                                                             |
| <b>inverse.gaussian</b> | Données positives fortement asymétriques à droite.                              |

### 7.3 Méthode

La méthode d'estimation du paramètre de lissage :

| Méthode        | Description                                                                                         |
|----------------|-----------------------------------------------------------------------------------------------------|
| <b>REML</b>    | Maximum de vraisemblance restreint (par défaut, recommandé). La plus robuste face au surajustement. |
| <b>GCV.Cp</b>  | Validation croisée généralisée. Tend à sélectionner des modèles légèrement plus complexes.          |
| <b>ML</b>      | Maximum de vraisemblance. Proche de REML, mais peut sous-lisser.                                    |
| <b>P-REML</b>  | Variante REML de Pearson.                                                                           |
| <b>P-ML</b>    | Variante ML de Pearson.                                                                             |
| <b>GACV.Cp</b> | Validation croisée approchée généralisée.                                                           |

## 7.4 Gamma (multiplicateur de lissage)

Le paramètre gamma multiplie les degrés de liberté effectifs dans le critère de lissage, ce qui favorise des ajustements plus lisses (moins ondulés). La valeur par défaut est 1,2.

| Valeur      | Effet                                                                           |
|-------------|---------------------------------------------------------------------------------|
| 1,0         | Lissage standard                                                                |
| 1,2–1,4     | Légèrement plus lisse que par défaut (recommandé)                               |
| 2,0 et plus | Beaucoup plus lisse — utile pour des données bruitées ou de petits échantillons |

Des valeurs de gamma plus élevées protègent du surajustement en pénalisant davantage la complexité. Le préréglage Earth Pipeline emploie 1,4 pour un lissage supplémentaire lors du raffinement des nœuds d'earth.

## 7.5 Validation croisée

Lorsque la case est cochée (par défaut), mgcvUI exécute après l'ajustement une validation croisée à 10 blocs afin de calculer un R-carré de validation croisée. Cela fournit une estimation honnête du pouvoir prédictif hors échantillon.

## 7.6 Select (rétrécissement)

Lorsque la case est cochée (par défaut), une pénalité supplémentaire est ajoutée à chaque terme lisse, capable de le ramener à zéro. Cela permet une sélection automatique de variables — les termes lisses sans importance sont de fait retirés du modèle. Mis en œuvre par `select = TRUE` dans `gam()`.

## 7.7 Base par défaut

Le type de fonction de base spline des termes lisses :

| Base      | Nom complet                         | Description                                                                               |
|-----------|-------------------------------------|-------------------------------------------------------------------------------------------|
| <b>tp</b> | Spline de régression à plaque mince | Par défaut. Isotrope (aucun placement de nœuds requis). Bon choix polyvalent.             |
| <b>cr</b> | Spline de régression cubique        | Permet un placement explicite des nœuds. Indispensable à l'intégration des nœuds d'earth. |
| <b>ps</b> | P-spline                            | Base de B-splines à nœuds équidistants avec pénalité de différences. Efficace en calcul.  |
| <b>bs</b> | B-spline                            | Base de B-splines souple.                                                                 |

## 7.8 k par défaut (dimension de la base)

La dimension de la base  $k$  détermine l'ondulation maximale de chaque terme lisse. La valeur 0 (par défaut) signifie « automatique » — mgcvUI utilise  $k = 10$  ou le nombre de nœuds d'earth, selon ce qui convient.

- **k petit (3–5)** : courbes très lisses, presque linéaires.
- **k moyen (8–15)** : souplesse modérée (généralement suffisante).
- **k grand (20 et plus)** : très souple — risque de surajustement en l'absence d'une régularisation forte.

**Astuce :** Les degrés de liberté effectifs (EDF) affichés dans l'onglet Summary indiquent combien de degrés de liberté chaque terme lisse consomme réellement. Si l'EDF est proche de  $k - 1$ , la base est peut-être trop petite — augmentez  $k$ . Si l'EDF est bien inférieur à  $k$ , le terme lisse est correctement pénalisé.

## 7.9 Type de produit tensoriel

Pour les interactions entre variables continues, deux types de produits tensoriels sont disponibles :

| Type      | Fonction                | Description                                                                                                                                                                                    |
|-----------|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ti</b> | <code>ti(x1, x2)</code> | Interaction tensorielle — ne modélise que l'effet d'interaction, les effets principaux étant traités séparément par des termes <code>s()</code> univariés. À préférer pour l'interprétabilité. |
| <b>te</b> | <code>te(x1, x2)</code> | Produit tensoriel — modélise l'effet conjoint complet, effets principaux compris. Plus difficile à décomposer pour les ajustements RCA.                                                        |

## 7.10 Interactions autorisées

Une section repliable **Allowed Interactions** (interactions autorisées) affiche une matrice triangulaire supérieure de cases à cocher pour tous les prédicteurs lisses inclus (ni linéaires, ni facteurs). Chaque case active un lissage par produit tensoriel entre le couple de variables correspondant.

- **Allow All / Clear All** — boutons de bascule globale.
- **Cliquez sur un nom de variable** — bascule toutes les interactions de cette variable.
- **Clic droit sur un nom de variable** — « Block from main effect » (exclure de l'effet principal) — marque la variable pour un usage en interaction seulement (pas de terme lisse univarié).

Lorsque des nœuds earthUI sont importés, les interactions détectées par earth sont précochées et verrouillées (surlignées sur fond jaune). Si le modèle earth utilisait `degree = 1` (sans interactions), un message d'information est affiché.

### Avertissement

L'activation d'interactions tensorielles accroît sensiblement la complexité du modèle. N'utilisez des interactions que lorsque vous disposez d'une connaissance du domaine ou d'indices détectés par earth qui les étayent. Vérifiez que les ajustements obtenus sont raisonnables et explicables.

## 7.11 Interactions facteur $\times$ terme lisse

Une matrice distincte **Factor-by-Smooth Interactions** apparaît lorsque des variables facteur et des variables lisses sont toutes deux incluses. Chaque case crée un terme `s(x, by = factor)` — une courbe lisse distincte pour chaque modalité du facteur.

### 7.12 Paramètres avancés

Une section repliable « Advanced » (avancé) expose des réglages supplémentaires :

- **Optimizer** (optimiseur) — algorithme d'estimation du paramètre de lissage : `outer/newton` (par défaut), `outer/bfgs` ou `efs` (Fellner-Schall étendu).
- **Scale** (échelle) — paramètre d'échelle. 0 (par défaut) = estimation à partir des données. Fixez une valeur connue pour les modèles à échelle imposée.
- **Discrete covariate method** (méthode à covariables discrètes) — ajustement discrétisé rapide pour les grands jeux de données (`discrete = TRUE`).
- **Threads** (fils d'exécution) — nombre de fils parallèles (1–32). Effectif seulement lorsque `discrete = TRUE`.

## 8 Ajustement du modèle

### 8.1 Le bouton d'ajustement

La section 6 de la barre latérale contient le bouton **Fit MgcV GAM Model**. Un clic lance le modèle avec votre configuration courante.

### 8.2 Ce qui se passe pendant l'ajustement

Lorsque vous cliquez sur Fit, mgcvUI :

1. **Valide la configuration** — vérifie la présence d'au moins un prédicteur et d'au moins 10 lignes complètes. Avertit si moins de 50% des lignes sont complètes (en indiquant les colonnes responsables de la plupart des NA).
2. **Prépare les données** — restreint aux seules colonnes utiles au modèle, supprime les lignes comportant des NA, applique les poids (s'ils sont désignés), exclut la ligne objet (en mode évaluation) et applique les transformations de la réponse.
3. **Construit la formule du GAM** — crée des termes lisses  $s(x, bs="cr", k=10)$  pour les prédicteurs numériques, des termes facteur pour les variables catégorielles, des termes linéaires pour les variables marquées Linear et des produits tensoriels pour les interactions sélectionnées. Les nœuds d'earth sont intégrés comme positions de nœuds explicites dans les termes à base `cr`.
4. **Traite les cas limites** — écarte les variables constantes, convertit les prédicteurs binaires en facteurs, plafonne  $k$  au nombre de valeurs distinctes moins un, complète les nœuds d'earth par des quantiles des données lorsqu'il en faut davantage.
5. **Ajuste le GAM** — appelle `mgcv::gam()` avec la formule construite, la famille, la méthode, gamma, select et les autres paramètres.
6. **Exécute la validation croisée** (si elle est activée) — validation croisée à 10 blocs calculant le R-carré hors échantillon.
7. **Met à jour tous les onglets de résultats** — Summary, Equation, graphiques de contribution, Diagnostics et tous les autres onglets sont remplis.

### 8.3 Ajustement en arrière-plan et progression

Lorsque le package `callr` est disponible, l'ajustement du modèle s'exécute dans un processus d'arrière-plan, de sorte que l'application reste réactive. Une surcouche de progression au style terminal sur fond sombre affiche :

- un **chronomètre** indiquant les secondes/minutes écoulées ;
- un **journal de trace en couleurs** : bleu pour l'état, jaune pour la progression des blocs de validation croisée, rouge pour les erreurs, vert pour les résultats ;
- un bouton **Abort** (interrompre) permettant de tuer le processus d'ajustement au besoin ;
- un bouton **Close** (fermer) qui apparaît une fois l'exécution terminée.

Si `callr` n'est pas disponible, l'ajustement s'exécute de manière synchrone (l'application se fige jusqu'à la fin de l'ajustement).

Après un ajustement réussi, une coche blanche apparaît sur le bouton Fit et une ligne d'état affiche : « R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X. »

## 9 Onglets de résultats

### 9.1 Data

Affiche les données importées sous forme de tableau interactif (DataTable) avec défilement horizontal et 15 lignes par page. En modes évaluation et marché, l'onglet est scindé en un tableau **Subject Property** (bien objet, ligne 1) et un tableau **Comparable Sales** (ventes comparables, les autres lignes); en mode General, un seul tableau est affiché.

Chaque cellule tient sur une seule ligne et est tronquée à la largeur de la colonne, de sorte que les champs de texte libre étendus (tels que les commentaires sur un bien) n'étirent pas une ligne sur toute la hauteur de la page. Un simple clic sélectionne une ligne; un **double-clic sur n'importe quelle cellule** ouvre une fenêtre affichant son contenu intégral, non tronqué.

### 9.2 Equation

Affiche le modèle GAM ajusté en deux parties :

**Équation du modèle** (rendue par MathJax) :

- Le membre de gauche présente la réponse avec la transformation adéquate (log, log10) et la fonction de lien.
- Le membre de droite présente la valeur de l'ordonnée à l'origine, les termes lisses sous la forme  $f_i(\text{variable})$ , les termes facteur regroupés par variable avec le nombre de modalités, et les termes linéaires avec la valeur de leur coefficient.
- La famille et la méthode sont affichées au-dessus de l'équation.

Tableau **Smooth Function Definitions** (définitions des fonctions lisses) :

- chaîne de formule du terme (par exemple `s(living_sqft, bs="cr", k=8)`)
- nom de la variable
- type (s, te, ti ou linéaire)
- type de fonction de base
- valeur de  $k$
- positions des nœuds d'earth (en cas d'importation)

### 9.3 Summary

Six cartes de métriques affichées en haut :

- **R-squared** — proportion de déviance expliquée
- **CV R-squared** — R-carré de validation croisée (si la validation croisée est activée)
- **Dev. Explained %** — pourcentage de déviance expliquée
- **AIC** — critère d'information d'Akaike
- **n** — nombre d'observations utilisées
- **Smooths** — nombre de termes lisses

Sous les cartes :

- **Tableau Smooth Terms** (termes lisses) — terme, EDF (degrés de liberté effectifs), Ref.df, statistique F, valeur p
- **Tableau Parametric Terms** (termes paramétriques) — terme, estimation, erreur type, valeur t, valeur p

### 9.4 Variable Importance

Un diagramme en barres horizontales montrant l'importance relative de chaque terme du modèle :

- **Termes lisses** : classés par statistique F (barres bleues)
- **Termes paramétriques** : classés par valeur  $|t|$  (barres vertes)

Sous le graphique, un DataTable indique le terme, le type, l'EDF, la statistique et la valeur p, triés par statistique décroissante.

## 9.5 Graphiques de contribution

Chaque terme du modèle qui contribue à la valeur prédite dispose d'un graphique de contribution interactif **plotly**, présenté dans une carte encadrée. L'onglet affiche des graphiques pour tous les types de termes :

### 9.5.1 Termes lisses univariés — $s(x)$

- **Courbe bleue** avec bande de confiance à 95%
- **Nuage de points gris** représentant les résidus partiels (pour les modèles sans transformation logarithmique)
- **Lignes verticales rouges en tirets** aux positions des nœuds d'earth (si earth a été importé)
- **Infobulle au survol** : valeur de la variable, contribution (\$), IC à 95%, taux (pente par unité)

Pour les **modèles à transformation logarithmique**, l'axe des ordonnées subit la transformation inverse pour exprimer des contributions en dollars, selon la formule  $\bar{y} \times (e^{f(x)} - 1)$ , ce qui rend les courbes directement interprétables en dollars.

Pour les variables de **latitude et longitude**, la pente est affichée par tranche de 0,001 degré plutôt que par degré, cet incrément géographique étant plus parlant.

### 9.5.2 Termes lisses par facteur — $s(x, \text{by} = \text{factor})$

Une courbe colorée par modalité du facteur, superposées sur un même graphique. Utile pour voir comment l'effet d'un prédicteur varie d'un groupe à l'autre (par exemple selon les quartiers).

### 9.5.3 Termes d'interaction — $ti(x_1, x_2)$ et $te(x_1, x_2)$

**Carte de chaleur** représentant la surface de contribution en 2D. Échelle de couleurs : rouge (négatif) → blanc (nul) → bleu (positif). Le survol affiche les valeurs des deux variables ainsi que le montant de la contribution.

### 9.5.4 Termes paramétriques — linéaires et facteurs

- **Termes linéaires numériques** : nuage de points des données avec la droite ajustée superposée
- **Termes facteur** : diagramme en barres de la contribution moyenne par modalité

## 9.6 Correlation

Une matrice en carte de chaleur des corrélations de Pearson entre tous les prédicteurs numériques. Disponible dès l'importation des données — aucun ajustement de modèle n'est requis.

- Dégradé de couleurs : bleu (−1) → blanc (0) → rouge (+1)
- Valeurs de corrélation imprimées dans chaque cellule
- Utile pour repérer la concurvité (l'analogue lisse de la multicollinéarité) avant l'ajustement

## 9.7 Diagnostics

Deux ensembles de visualisations de diagnostic :

**Panneau de diagnostic** (via `gratia::appraise`) : quatre graphiques — résidus vs valeurs ajustées, diagramme quantile-quantile, histogramme des résidus et réponse vs valeurs ajustées.

Nuage de points **Actual vs Predicted** (observé vs prédit) : valeurs observées comparées aux valeurs prédites, avec une droite de référence en tirets à 45 degrés. Pour les modèles à transformation logarithmique, les valeurs subissent la transformation inverse vers l'échelle d'origine.

## 9.8 RCA Adjustments

Trois histogrammes affichés après le calcul RCA (modes évaluation/marché uniquement) :

- **Residual Adjustment %** — distribution des ajustements de résidu en pourcentage du prix de vente
- **Net Adjustment %** — distribution des ajustements nets
- **Gross Adjustment %** — distribution des ajustements bruts

Chaque histogramme indique la moyenne, la médiane et l'écart type dans son sous-titre, avec des droites de référence en tirets pour la moyenne et la médiane.

## 9.9 Anova

Un tableau ANOVA combinant les termes paramétriques et lisses issus de la synthèse du GAM. Il indique le terme, le type (paramétrique ou lisse) et toutes les statistiques associées.

## 9.10 Mgcvc Output

Sortie en texte brut affichant :

- la durée de l'ajustement
- la formule du GAM
- la sortie d'impression du modèle
- la sortie complète de `summary(model)`
- la famille et la fonction de lien
- les mesures globales de concurvité

## 9.11 Sign Check

Lorsque des nœuds earthUI ont été importés, cet onglet compare la direction de chaque variable dans le modèle earth avec sa direction dans le GAM :

- **earth\_direction** : croissante, décroissante ou mixte (d'après le signe des fonctions charnières)
- **gam\_direction** : croissante, décroissante ou mixte (estimée à partir de la dérivée du terme lisse sur une grille de 100 points)
- **consistent** : TRUE/FALSE
- **warning** : description de toute discordance

Une ligne d'état affiche « All directions consistent » (toutes les directions sont cohérentes) ou le nombre de variables présentant des incohérences.

## 9.12 Concurvity

La concurvité est l'analogue lisse de la multicolinéarité. Cet onglet affiche :

- **Tableau Overall Concurvity** (concurvité globale) : lignes pour les mesures worst/observed/estimate, colonnes pour chaque terme lisse. Code couleur : blanc ( $<0,5$ ), jaune ( $0,5-0,8$ ), rouge ( $>0,8$ ).
- **Tableau Pairwise Concurvity** (concurvité par paires) : matrice des pires cas par paire, avec le même code couleur.

Des valeurs supérieures à  $\sim 0,8$  dans la ligne « worst » indiquent qu'un terme lisse peut être bien approché par les autres termes lisses, ce qui suggère une redondance.

### 9.13 Basis Check

Sortie de `mgcv::gam.check()` : tests d'adéquation de la dimension de base pour chaque terme lisse. Si le  $k$  d'un terme lisse est trop petit, ce test le signale par une valeur  $p$  faible, ce qui suggère d'augmenter  $k$ .

## 10 Téléchargement des données

Après l'ajustement, le bouton **Download Output (Excel)** (section 7 de la barre latérale) exporte un fichier Excel contenant les prédictions et les diagnostics dans le dossier de sortie mgcv du projet actif.

### 10.1 Colonnes de sortie

| Colonne                         | Description                                                                                                                           |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| <b>basis</b>                    | Valeur de l'ordonnée à l'origine après transformation inverse (identique pour toutes les lignes)                                      |
| <b>&lt;var&gt;_contribution</b> | Contribution de chaque terme lisse à la prédiction. Pour les modèles log, répartie proportionnellement en valeurs monétaires.         |
| <b>est_&lt;target&gt;</b>       | Prédiction du modèle (par exemple <b>est_sale_price</b> ), après transformation inverse depuis l'échelle logarithmique le cas échéant |
| <b>residual</b>                 | Observé – prédit                                                                                                                      |
| <b>calc_residual</b>            | Vérification : observé – somme des contributions (après transformation inverse)                                                       |
| <b>cqa</b>                      | Score CQA (Condition-Quality-Appeal : état, qualité, attrait), sur une échelle de 0 à 10                                              |
| <b>residual_sf</b>              | Résidu / surface habitable (si <b>living_area</b> est désignée)                                                                       |
| <b>cqa_sf</b>                   | CQA calculé à partir du classement selon <b>residual_sf</b>                                                                           |

En mode évaluation, la ligne 1 (l'objet) a **residual**, **cqa** et **cqa\_sf** fixés à NA. Les lignes sont triées par **residual\_sf** décroissant lorsque une colonne **living\_area** est désignée.

Une coche blanche apparaît sur le bouton de téléchargement une fois l'opération réussie.

### 10.2 Scores CQA

Le CQA classe le résidu de chaque ligne par rapport à tous les autres sur une échelle de 0 à 10 :

- **CQA élevé (~9–10)** : vendu bien au-dessus du prix prédit — état, qualité ou attrait probablement supérieurs
- **CQA faible (~0–1)** : vendu bien au-dessous du prix prédit — état probablement inférieur ou vente en difficulté
- **CQA ~5** : proche de la médiane

## 11 Calculs RCA et téléchargement

Le flux de travail RCA (*Residual Constraint Approach*, approche par contrainte des résidus) est disponible en modes **Appraisal** et **Market Area Analysis**, après l'ajustement du modèle. Le bien objet doit se trouver en ligne 1.

### 11.1 Ouverture de la boîte de dialogue RCA

Cliquez sur le bouton **Calculate RCA Adjustments & Download** à la section 8 de la barre latérale. Une boîte de dialogue modale apparaît, comportant :

- **Score type** (type de score) : CQA ou CQA par pied carré (si **living\_area** est désignée). L'option par pied carré met à l'échelle le résidu de l'objet selon la surface habitable.
- **Subject CQA Score** (score CQA de l'objet) : accepte des valeurs de 0,00 à 10,00. Le champ est initialement vide ; lorsqu'un modèle earthUI a été importé, le type de score et la valeur utilisés dans earthUI sont repris depuis les réglages du projet partagé.

Cliquez sur **Generate** (générer) pour calculer et télécharger. Le fichier Excel est écrit dans le dossier de sortie mgcv du projet actif (<os>\_out\_mgcv).

### 11.2 Interpolation du score CQA

1. Les scores CQA et les résidus des comparables sont triés
2. Une interpolation linéaire (**approx()**) fait correspondre votre valeur de CQA à un résidu
3. Valeur de l'objet = prédiction du modèle + résidu interpolé
4. Les lignes de poids nul sont traitées comme si elles avaient le même résidu que l'objet

### 11.3 Colonnes de sortie

| Colonne             | Description                                                       |
|---------------------|-------------------------------------------------------------------|
| subject_value       | Prédiction du modèle + résidu interpolé                           |
| subject_cqa         | Score CQA saisi par l'utilisateur                                 |
| <var>_adjustment    | Contribution de l'objet – contribution du comparable (en dollars) |
| residual_adjustment | Résidu de l'objet – résidu du comparable                          |
| net_adjustments     | Somme de tous les ajustements                                     |
| gross_adjustments   | Somme des valeurs absolues des ajustements                        |
| residual_adj_pct    | Ajustement de résidu en % du prix de vente                        |
| net_adj_pct         | Ajustements nets en % du prix de vente                            |
| gross_adj_pct       | Ajustements bruts en % du prix de vente                           |
| adjusted_sale_price | Prix de vente observé + ajustements nets                          |

Une coche blanche apparaît sur le bouton une fois le calcul réussi.

## 12 Grille de comparaison des ventes

La grille de comparaison des ventes n'est disponible qu'en mode **Appraisal**, après le calcul des ajustements RCA. Elle génère un classeur Excel mis en forme, adapté à l'inclusion dans un rapport d'évaluation.

### 12.1 Boîte de dialogue de sélection des comparables

Un clic sur le bouton **Generate Sales Grid & Download** ouvre une boîte de dialogue modale listant toutes les ventes comparables. Les comparables sont répartis en deux groupes :

- **Recommended comps** (comparables recommandés, précochés) : ajustement brut < 25% du prix de vente, triés par âge de la vente (les plus récents d'abord). Maximum 30 comparables.
- **Other comps** (autres comparables, décochés par défaut) : les comparables restants, triés par pourcentage d'ajustement brut.

Sélectionnez vos comparables, puis cliquez sur **Generate Sales Grid** pour créer le classeur.

**Types spéciaux requis** : `contract_date` et `living_area` doivent être désignés. **Types spéciaux recommandés** : `latitude`, `longitude` et `lot_size` — la grille fonctionne sans eux, mais certains champs resteront vides.

### 12.2 Structure du classeur

Le classeur généré contient plusieurs feuilles, chacune présentant 3 comparables côte à côte :

- **Lignes d'en-tête** : adresses de l'objet et des comparables, prix de vente, dates de vente et jours sur le marché
- **Lignes groupées** : localisation (latitude, longitude, zone), taille du terrain (superficie, dimensions du site) et âge (âge réel, âge effectif), avec en-têtes de groupe et sous-éléments
- **Lignes des variables du modèle** : une ligne par terme lisse indiquant la contribution de l'objet, celles des comparables et les montants d'ajustement
- **Lignes des caractéristiques résiduelles** : cellules vides et déverrouillées permettant à l'évaluateur de saisir les caractéristiques non captées par le modèle (état, qualité, vue, etc.)
- **Lignes de synthèse** : ajustement net en %, ajustement brut en % et prix de vente ajusté, avec des formules Excel actives

### 12.3 Protection des feuilles et cellules résiduelles

Chaque feuille est protégée afin d'éviter les modifications accidentelles, mais les cellules de valeur des caractéristiques résiduelles sont explicitement déverrouillées. Les évaluateurs peuvent ainsi saisir des valeurs pour les caractéristiques absentes du modèle tout en préservant les calculs d'ajustement pilotés par formules.

## 13 Téléchargement des rapports

La génération de rapports suit un flux Quarto en deux étapes, qui remplace le rapport en une seule étape des versions antérieures.

### 13.1 Étape 1 : générer le rapport Quarto

Le bouton **Generate Quarto Report** (disponible une fois le modèle ajusté) écrit un ensemble `.qmd` autonome dans le dossier de sortie `mgcv` du projet actif. L'ensemble contient la source Quarto, les fichiers de graphiques prérendus (PNG et PDF), une charge utile `report_data.rds` regroupant tous les tableaux et les métadonnées du modèle, ainsi qu'un modèle de style `reference.docx` pour la sortie Word. Parce que l'ensemble est autonome, vous pouvez modifier la source `.qmd` ou la combiner avec d'autres documents Quarto avant le rendu.

Les graphiques sont prérendus lors de la génération de l'ensemble ; si un graphique échoue, une image de remplacement est écrite et le reste de l'ensemble est généré normalement.

### 13.2 Étape 2 : convertir le rapport Quarto

La section **Convert Quarto Report** rend un fichier `.qmd` vers un ou plusieurs formats de sortie au moyen de `quarto::quarto_render()`. Elle sélectionne par défaut le dernier ensemble généré, mais le bouton **Browse** (parcourir) permet de choisir n'importe quel fichier `.qmd`. Trois formats sont disponibles :

- **HTML** — document HTML autonome. Fonctionne sur toutes les plateformes sans logiciel supplémentaire. Choix recommandé par défaut.
- **Word (.docx)** — rendu avec le modèle de style `reference.docx` fourni. Adapté à l'édition et à la diffusion.
- **PDF** — **exige une installation LaTeX** (voir la configuration requise ci-dessous). Si aucune installation LaTeX n'est détectée, l'option PDF est automatiquement masquée dans les cases de format.

Les générateurs historiques en une seule étape, `render_gam_report()` et `export_gam_docx()`, demeurent disponibles pour un usage programmatique.

### 13.3 Contenu du rapport

Le rapport généré comprend :

- **Dataset & Model** : fichier de données, taille de l'échantillon, famille, méthode et informations d'ajustement
- **Results Summary** :  $R^2$ ,  $R^2$  de validation croisée, déviance expliquée, AIC et métriques d'ajustement associées
- **Model Equation** : équation du modèle, avec la notation  $f_i$  pour les termes lisses et des coefficients explicites pour les termes linéaires, ainsi que les définitions des fonctions lisses. Les équations longues se répartissent sur plusieurs lignes dans le PDF.
- **Tableaux de synthèse** : termes lisses (EDF, F, valeur p) et termes paramétriques (estimation, erreur type, t, valeur p)
- **Interactions activées** : une matrice des termes d'interaction tensoriels et facteur  $\times$  terme lisse inclus dans le modèle (affichée uniquement en présence d'interactions)
- **Variable Importance** : diagramme en barres classé par statistique F (termes lisses) ou par  $|t|$  (termes paramétriques)
- **Smooth Effects** (tous les types de termes) :
  - graphiques de termes lisses univariés (versions statiques ggplot2 des graphiques interactifs)

- graphiques multicourbes des termes lisses par facteur
- cartes de chaleur des interactions (termes  $ti$  et  $te$ )
- graphiques des termes paramétriques (nuage de points pour les termes numériques, diagramme en barres pour les facteurs)
- carte de chaleur **Predictor Correlations** (corrélations entre prédicteurs)
- **Diagnostics** : diagnostics de résidus et nuage de points observé vs prédit
- **mgcv Summary** : la sortie texte complète du modèle ajusté
- **Annexe de notation** : une section de clôture non numérotée expliquant la notation des fonctions de contribution à la valeur employée dans le rapport

Une coche blanche apparaît sur le bouton Generate Quarto Report une fois la génération réussie.

## 14 Export des fonctions du modèle

mgcvUI peut exporter les fonctions lisses du GAM ajusté sous forme de code autonome dans plusieurs langages. Chaque terme lisse est évalué sur une grille de 200 points et exporté sous forme de table de correspondance avec interpolation linéaire, ce qui permet d'utiliser le modèle en dehors de R sans aucune dépendance à `mgcv`.

L'export de fonctions est programmatique dans la version actuelle. Les fonctions utilitaires exportées prennent un objet de résultat ajusté (produit par `fit_gam()`) et écrivent les fichiers décrits ci-dessous :

- `generate_r_code()`, `generate_python_code()` et `generate_cpp_code()` — fichiers source par langage
- `export_functions_sqlite()` — base de données SQLite de tables de correspondance
- `export_functions_zip()` — regroupe n'importe quelle combinaison des quatre formats dans une archive zip

### 14.1 Fonctions R

Fichier : `gam_functions.R`

- Commentaires d'en-tête indiquant la variable réponse, le  $R^2$ , le  $R^2$  de validation croisée et la formule.
- Une constante `INTERCEPT`.
- Un `approxfun()` par variable lisse, avec des vecteurs `x/y` précalculés.
- Les coefficients des termes linéaires sous forme de constantes nommées.
- Utilisation : additionner l'ordonnée à l'origine, chaque `g_var(value)` et chaque `coef_var * value`.

### 14.2 Fonctions Python

Fichier : `gam_functions.py`

- Import de `numpy`.
- Constante `INTERCEPT`.
- Par variable : tableaux `_x_var` et `_fx_var`, fonction `g_var(x)` utilisant `np.interp()`.
- Coefficients des termes linéaires sous forme de constantes `COEF_*`.

### 14.3 En-tête C++

Fichier : `gam_functions.hpp`

- En-tête seul, avec `#pragma once`.
- Espace de noms `gam_functions`.
- Fonction utilitaire `interp()` reposant sur `std::lower_bound`.
- Par variable : fonction `inline` avec des tableaux `static const`.
- Coefficients des termes linéaires en `constexpr double`.

### 14.4 Base de données SQLite

Fichier : `gam_functions.sqlite`

- Table `smooth_grids` : colonnes (variable, `x`, `fx`) — les tables de correspondance à 200 points.
- Table `model_info` : `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.
- Table `linear_terms` : variable, coefficient.

— Table **intercept** : valeur de l'ordonnée à l'origine.

## 15 Comparaison avec earthUI et glmnetUI

L'application earth et ses Post Processing Modules forment un unique ensemble d'outils de modélisation par régression. Ils partagent le même format de données, les mêmes types de colonnes spéciales, le même flux de travail RCA et les mêmes jeux de données de démonstration, mais reposent sur des moteurs de modélisation différents.

### 15.1 Principales différences

| Caractéristique        | mgcvUI                                     | earthUI                            | glmnetUI                      |
|------------------------|--------------------------------------------|------------------------------------|-------------------------------|
| Méthode                | GAM (mgcv)                                 | MARS/Earth (earth)                 | Elastic net (glmnet)          |
| Relations              | Non linéaires lisses                       | Linéaires par morceaux             | Linéaires                     |
| Sélection de variables | Rétrécissement<br><code>select=TRUE</code> | Élagage pas à pas                  | Pénalité Lasso                |
| Interactions           | Produits tensoriels (te/ti)                | Produits de charnières (degré 2–3) | Produits deux à deux          |
| Effets partiels        | Courbes lisses avec IC                     | g-fonctions (par morceaux)         | Pentes linéaires              |
| Chaîne earth           | Importe les nœuds d'earth                  | Source des nœuds                   | Importe la base de charnières |
| Export de code         | R, Python, C++, SQLite                     | R, Python, C++, SQLite             | —                             |
| Formats de rapport     | Word, PDF, HTML                            | Word, PDF, HTML                    | Word, PDF, HTML               |

Un élément différenciant propre à earthUI mérite d'être signalé : earthUI comprend une fonctionnalité expérimentale Prolog/DCG qui extrait des attributs structurés de biens à partir des commentaires MLS en texte libre.

### 15.2 Fonctionnalités communes

Les trois outils offrent :

- une arborescence de projet `regProj` partagée (par défaut `~/regProj`) — projets, fichiers d'entrée et sorties propres à chaque méthode sont partagés entre les applications sœurs
- l'importation CSV et Excel avec conversion des colonnes en `snake_case`
- la prise en charge des locales de plus de 30 pays, avec séparateur CSV, symbole décimal et format de papier
- les désignations de colonnes spéciales (`contract_date`, `living_area`, `area`, `latitude`, `longitude`, etc.)
- le calcul automatique de `sale_age` à partir de la date d'effet
- la bascule mode sombre/clair (thème Nord)
- la persistance des réglages (SQLite et/ou `localStorage`)
- la carte de chaleur des corrélations, l'importance des variables et les diagnostics
- le score CQA et le flux de travail d'ajustement RCA
- la grille de comparaison des ventes avec sélection des comparables et sortie Excel mise en forme
- la compatibilité avec les mêmes jeux de données de démonstration (`Appraisal_1.csv`)

### 15.3 Flux de travail recommandé

- **Commencez par earthUI** pour découvrir les variables importantes, les relations non linéaires et les interactions. La construction automatique de modèle d'earth offre le meilleur point de départ.
- **Affinez avec mgcvUI** si vous souhaitez des effets partiels plus lisses. Importez les nœuds d'earthUI pour une transition sans heurt.
- **Utilisez glmnetUI** lorsque vous avez besoin d'un modèle purement linéaire avec une sélection de variables agressive ou des contraintes de signe sur les coefficients.
- **Comparez les trois** sur les mêmes données afin d'évaluer si la complexité supplémentaire des modèles non linéaires est justifiée.

## 16 Jeu de données de démonstration : Appraisal\_1.csv

### 16.1 Description

Le jeu de données de démonstration est partagé avec earthUI et glmnetUI. Il se trouve dans le dossier `demo_mls/` des sources de mgcvUI ou, si earthUI est installé :

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

Le fichier contient environ 1 500 enregistrements de ventes résidentielles issus d'un export MLS simulé, avec le bien objet en ligne 1 et les ventes comparables sur les lignes suivantes. Les données représentent des ventes de maisons individuelles dans un marché à plusieurs zones, avec une gamme variée de tailles, d'âges et de localisations.

Il ne s'agit pas de données réelles, mais elles s'inspirent d'un quartier réaliste du nord de la Californie. Toute information d'identification a été modifiée ou supprimée.

### 16.2 Colonnes

| Colonne             | Type      | Type spécial  | Description                                                              |
|---------------------|-----------|---------------|--------------------------------------------------------------------------|
| sale_type           | character | sale_type     | Code de type d'enregistrement distinguant l'objet des ventes comparables |
| weight              | numeric   | weight        | Poids de l'observation (0 = exclue de l'ajustement)                      |
| id                  | numeric   | display_only  | Identifiant interne de l'enregistrement                                  |
| property_id         | numeric   | display_only  | Identifiant MLS du bien                                                  |
| listing_id          | character | display_only  | Numéro d'annonce MLS                                                     |
| parcel_number       | character | display_only  | Numéro de parcelle cadastrale (APN)                                      |
| street_address      | character | display_only  | Adresse du bien                                                          |
| city_name           | character | display_only  | Ville                                                                    |
| postal_code         | character | display_only  | Code postal                                                              |
| county_name         | character | display_only  | Comté                                                                    |
| contract_date       | Date      | contract_date | Date du contrat de vente (permet de calculer sale_age)                   |
| sale_age            | numeric   | —             | Jours entre la date de contrat et la date d'effet (précalculé)           |
| coe_date            | Date      | display_only  | Date de clôture de l'entiercement                                        |
| listing_status      | character | display_only  | Statut de l'annonce (par exemple « Sold »)                               |
| sale_price          | numeric   | (cible)       | Prix de vente — variable réponse                                         |
| rent                | numeric   | —             | Loyer mensuel (pour les modèles à cibles multiples)                      |
| list_price          | numeric   | display_only  | Prix affiché                                                             |
| original_list_price | numeric   | display_only  | Prix affiché d'origine                                                   |
| living_sqft         | numeric   | living_area   | Surface habitable brute en pieds carrés                                  |
| ppsf                | numeric   | display_only  | Prix au pied carré                                                       |
| beds_total          | integer   | —             | Nombre de chambres                                                       |
| baths_total         | numeric   | —             | Nombre total de salles de bains (par exemple 2,5 = 2 complètes + 1 demi) |
| lot_size            | numeric   | lot_size      | Superficie du terrain en pieds carrés                                    |
| area_id             | integer   | area          | Identifiant de zone MLS                                                  |
| area_text           | character | display_only  | Nom de la zone                                                           |
| age                 | numeric   | actual_age    | Âge du bien en années                                                    |
| year_built          | integer   | display_only  | Année de construction                                                    |
| latitude            | numeric   | latitude      | Latitude (arrondie à 3 décimales pour le modèle)                         |
| longitude           | numeric   | longitude     | Longitude (arrondie à 3 décimales pour le modèle)                        |
| latitude6           | numeric   | display_only  | Latitude en précision complète                                           |
| longitude6          | numeric   | display_only  | Longitude en précision complète                                          |
| garage_spaces       | integer   | —             | Nombre de places de garage                                               |
| fp_count            | integer   | —             | Nombre de cheminées                                                      |

| Colonne          | Type      | Type spécial | Description                                           |
|------------------|-----------|--------------|-------------------------------------------------------|
| no_of_stories    | numeric   | —            | Nombre d'étages                                       |
| style            | character | —            | Style architectural                                   |
| view             | character | —            | Type de vue (par exemple « Neighborhood », « Hills ») |
| days_on_market   | integer   | dom          | Jours sur le marché                                   |
| listing_date     | Date      | listing_date | Date de mise en vente                                 |
| sale_concessions | numeric   | concessions  | Concessions du vendeur                                |

### 16.3 Prise en main rapide suggérée

1. Lancez la routine mgcv : `earthUI::launch_mgcv()`
2. Créez un nouveau projet (section 1) ayant pour objet **For Appraisal**
3. Placez `Appraisal_1.csv` dans le dossier d'entrée du projet et sélectionnez-le à la section 2 (cliquez sur **Refresh** si nécessaire)
4. Sélectionnez `sale_price` comme cible
5. Éventuellement, réglez Response Transform sur **Log (natural)** pour obtenir des ajustements proportionnels
6. Attribuez les types spéciaux comme indiqué dans le tableau ci-dessus
7. Incluez les prédicteurs : `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (comme facteur), `age`, `latitude`, `longitude`, `garage_spaces`
8. Conservez les valeurs par défaut : REML, gamma = 1,2, base tp, k = auto, select = TRUE
9. Cliquez sur **Fit Mgcvm GAM Model**
10. Examinez les onglets Summary, Contribution (graphiques de termes lisses) et Diagnostics
11. Téléchargez les sorties (étape 7) et examinez le classement CQA
12. Calculez les ajustements RCA (étape 8) avec un score CQA d'environ 5,00
13. Générez une grille de comparaison des ventes (étape 9) avec les comparables recommandés

## 17 Configuration requise et dépannage

### 17.1 Plateformes prises en charge

mgcvUI fonctionne sous **macOS**, **Windows** et **Linux**. L'application est développée et principalement testée sous macOS avec RStudio. Les remarques propres à chaque plateforme figurent ci-dessous.

### 17.2 Version de R

$R \geq 4.1.0$  est requis. RStudio Desktop (2023.06 ou ultérieur) est vivement recommandé — il embarque **Quarto** et **pandoc** (nécessaires au rendu des rapports) et fournit un environnement pratique pour lancer l'application.

### 17.3 Packages R requis

Les packages suivants sont installés automatiquement en tant que dépendances :

| Package           | Rôle                                                                     |
|-------------------|--------------------------------------------------------------------------|
| shiny             | Cadre d'application web                                                  |
| mgcv              | Ajustement des modèles GAM                                               |
| valengrCore       | Noyau d'évaluation partagé (rôles spéciaux, report des réglages earthUI) |
| ggplot2, plotly   | Graphiques de contribution statiques et interactifs                      |
| gratia            | Extraction des termes lisses et diagnostics                              |
| DT                | Tableaux de données interactifs                                          |
| quarto, rmarkdown | Génération et rendu des rapports Quarto                                  |
| readxl            | Importation de fichiers Excel                                            |
| shinyFiles        | Explorateurs de fichiers et de dossiers                                  |
| RSQLite, DBI      | Bases de données des réglages et des projets (SQLite)                    |
| jsonlite          | Sérialisation de l'état des variables                                    |

## 17.4 Dépendances facultatives

| Composant                                | Quand il est nécessaire                                                                                                                                                                                       |
|------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>LaTeX</b> (TinyTeX, MiKTeX ou MacTeX) | Génération de rapports PDF uniquement. Si aucune installation LaTeX n'est détectée, l'option PDF est automatiquement masquée dans le menu des formats. Installation : <code>tinytex::install_tinytex()</code> |
| <b>sysfonts + showtext</b>               | Police Roboto Condensed pour les graphiques ggplot2. En cas d'indisponibilité ou de fonctionnement hors ligne, l'application bascule automatiquement sur la police sans empattement du système.               |
| <b>earth</b> ( $\geq 5.3.0$ )            | Nécessaire uniquement à la chaîne d'importation earthUI. Inutile pour un usage autonome.                                                                                                                      |
| <b>callr</b>                             | Ajustement du modèle en arrière-plan (l'application reste réactive). Sans lui, l'ajustement s'exécute de manière synchrone.                                                                                   |
| <b>writexl, openxlsx</b>                 | Téléchargements Excel (sorties du modèle, RCA, grille de ventes). L'application invite à les installer s'ils manquent.                                                                                        |
| <b>bslib</b>                             | Thématisation Bootstrap 5 (Nord Light/Dark).                                                                                                                                                                  |
| <b>officer</b>                           | Export Word programmatique historique ( <code>export_gam_docx()</code> ).                                                                                                                                     |

## 17.5 Remarques propres à chaque plateforme

### 17.5.1 macOS

C'est la plateforme qui pose le moins de problèmes. Homebrew est recommandé pour les bibliothèques système. Si vous avez besoin de rapports PDF :

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

### 17.5.2 Windows

Fonctionne bien avec RStudio Desktop. Points essentiels :

- **Machines d'entreprise verrouillées** : si le répertoire temporaire ou le répertoire de données utilisateur est restreint, la persistance des réglages et la génération des rapports peuvent échouer. L'application affiche un avertissement dans la console R mais continue de fonctionner sans persistance.
- **LaTeX** : installez TinyTeX depuis R (même commande que sous macOS ci-dessus) ou installez MiKTeX depuis [miktex.org](https://miktex.org).
- **Chemins de fichiers longs** : Windows impose une limite de 260 caractères pour les chemins. Gardez un chemin de répertoire de travail court.

### 17.5.3 Linux

C'est la plateforme la plus variable d'une distribution à l'autre. Les utilisateurs d'Ubuntu et de Debian peuvent avoir besoin de bibliothèques système avant que les packages R ne compilent :

```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

Pour les rapports PDF : `tinytex::install_tinytex()` ou installation de `texlive-full` depuis le gestionnaire de packages du système.

Sur les serveurs sans affichage, les graphiques interactifs plotly sont rendus sous forme d'images statiques de repli dans les rapports. L'application Shiny elle-même nécessite une connexion depuis un navigateur web.

## 17.6 Dégradation progressive

mgcvUI est conçu pour fonctionner même lorsque des composants facultatifs sont absents :

| Composant absent                                      | Comportement                                                                                                                                                                          |
|-------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pas de LaTeX                                          | L'option PDF est masquée dans le menu des formats de rapport. HTML et Word restent disponibles.                                                                                       |
| Pas d'accès Internet / échec des polices              | La police Roboto Condensed est remplacée par la police sans empattement du système. Un message est consigné dans la console.                                                          |
| Pas de RSQLite / système de fichiers en lecture seule | Les réglages ne persistent pas d'une session à l'autre. Un avertissement s'affiche dans la console. L'application fonctionne normalement par ailleurs.                                |
| Échec de <code>gratia : appraise</code>               | L'onglet Diagnostics et les rapports affichent un message de remplacement. Utilisez <code>mgcv::gam.check()</code> dans la console comme solution de rechange.                        |
| Répertoire temporaire non accessible en écriture      | La génération de rapports échoue avec un message d'erreur explicite. Corrigez en pointant la variable d'environnement <code>TMPDIR</code> vers un emplacement accessible en écriture. |

## 17.7 Dépannage

### 17.7.1 « PDF option is not available »

Aucune installation LaTeX n'a été détectée. Exécutez dans R :

```
install.packages("tinytex")
tinytex::install_tinytex()
```

Redémarrez ensuite l'application. L'option PDF apparaîtra.

### 17.7.2 « Settings will not persist »

Le répertoire de la base de données SQLite ne peut être ni créé ni écrit. Vérifiez les permissions sur :

- macOS/Linux : `~/Library/Application Support/R/mgcvUI/` ou `~/.local/share/R/mgcvUI/`
- Windows : `%APPDATA%/R/data/mgcvUI/`

### 17.7.3 « Diagnostics plot unavailable »

La fonction `gratia::appraise()` échoue occasionnellement sur des modèles complexes (nombreuses interactions tensorielles, familles non standard). Il s'agit d'un problème connu en amont. Contournement : exécutez `mgcv::gam.check(model)` dans la console R pour afficher directement les graphiques de diagnostic.

### 17.7.4 « Report error : number of items to replace... »

Il s'agissait d'un problème connu des versions antérieures, lié au package `gratia`. Assurez-vous d'avoir installé la dernière version de `mgcvUI`. La version actuelle intercepte cette erreur et produit le rapport avec un contenu de remplacement pour la section des diagnostics.

### 17.7.5 Le port 7880 est déjà utilisé

Si l'application ne démarre pas parce que le port 7880 est occupé (par une session précédente), exécutez :

```
# macOS/Linux:
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

### 17.7.6 L'installation des packages échoue sous Linux

Bibliothèques système manquantes. Installez les en-têtes de développement listés dans la section Linux ci-dessus, puis relancez `install.packages()`.

## Références

- WOOD, S. N. (2003). « Thin-plate regression splines ». In : *Journal of the Royal Statistical Society (B)* 65.1, p. 95-114. DOI : [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). « Stable and efficient multiple smoothing parameter estimation for generalized additive models ». In : *Journal of the American Statistical Association* 99.467, p. 673-686. DOI : [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). « Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models ». In : *Journal of the Royal Statistical Society (B)* 73.1, p. 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<sup>e</sup> éd. Chapman et Hall/CRC.
- WOOD, S. N., N. PYA et B. SÄFKEN (2016). « Smoothing parameter and model selection for general smooth models (with discussion) ». In : *Journal of the American Statistical Association* 111, p. 1548-1575. DOI : [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).

# Fondements

## Vue d'ensemble (version française)

Foundations • The Valuation Engineer

Bert Craytor, rédacteur en chef • 1<sup>er</sup> août 2026

Version 1.0.0

DOI: [10.67330/82vssv81](https://doi.org/10.67330/82vssv81)

---

*Traduction française de l'entrée « Overview » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

*Foundations* publie des traitements précis et citables de concepts qui sous-tendent la méthodologie moderne de l'évaluation. Les entrées prennent des termes que les évaluateurs praticiens rencontrent — parfois sous d'autres noms, parfois sans les reconnaître — et les replacent dans leur contexte formel : d'où ils viennent, ce qu'ils signifient techniquement, où ils apparaissent dans la pratique et pourquoi ils importent pour la défendabilité.

Le numéro inaugural établit les six concepts fondamentaux du cadre hédonique, agencés de sorte que chaque entrée s'appuie sur le vocabulaire des précédentes. La séquence va du *bien hétérogène* (la propriété de l'immobilier qui rend la méthodologie d'évaluation nécessaire tout court), en passant par l'*espace des caractéristiques* et la *fonction de prix hédoniques* (l'appareil formel qui organise le panier d'attributs qu'un bien immobilier représente), puis par le *prix implicite* et la *régression hédonique* (comment la contribution marginale de chaque attribut est extraite des données de marché), pour aboutir enfin à la *variable latente* (ce que représente réellement le signal résiduel non récupéré de tout modèle hédonique).

Un unique exemple fil rouge — huit comparables d'un quartier de Pacifica, en Californie — traverse les six entrées et constitue le fil pédagogique du numéro.

## Pourquoi ces six concepts, dans cet ordre

Le choix des six entrées n'est pas arbitraire. Chaque concept est porteur pour la méthodologie hédonique moderne, chacun est réellement rencontré dans la pratique quotidienne de l'évaluation immobilière, et chacun est assez souvent mal compris ou laissé sans nom pour que sa mise en forme citable ait une valeur pratique.

L'ordre est dicté par la dépendance conceptuelle. Le *bien hétérogène* est la racine : il nomme la propriété de l'immobilier (et des voitures d'occasion, des marchés du travail et de quelques autres biens) qui rend nécessaire l'ensemble du cadre hédonique. L'*espace des caractéristiques* est l'appareil formel permettant de représenter les biens hétérogènes comme des vecteurs ; rien de plus ne peut être dit avec précision sans lui. La *fonction de prix hédoniques* est l'objet d'équilibre qui associe des prix aux points de l'espace des caractéristiques. Les *prix implicites* en sont les dérivées partielles — les contributions marginales que chaque ajustement d'une grille de comparaison des ventes peut être lu comme affirmant, que l'évaluateur les nomme ainsi ou non. La *régression hédonique* est l'estimateur empirique qui extrait les prix implicites des transactions observées ; c'est l'entrée où la théorie devient analyse de données. Les *variables latentes* prennent le relais là où commencent les résidus de la régression : les attributs non observés dont les effets subsistent dans le résidu comme signal structuré plutôt que comme bruit à écarter.

Un lecteur qui procède dans cet ordre devrait en ressortir avec un fil continu allant de « pourquoi la méthodologie d'évaluation existe-t-elle tout court ? » à « que faut-il penser des résidus d'un modèle ajusté ? » — un parcours que, d'après l'expérience de l'auteur, la plupart des évaluateurs ne traversent que par fragments, au fil de nombreuses années de pratique.

## L'exemple fil rouge

Un unique jeu de données de huit comparables résidentiels de Pacifica, en Californie, traverse les six entrées :

| Comp | GLA (sf) | Terrain (sf) | Vue | État      | Prix de vente |
|------|----------|--------------|-----|-----------|---------------|
| A    | 1,650    | 9,000        | oui | bon       | \$1,425,000   |
| B    | 1,650    | 7,200        | non | bon       | \$1,280,000   |
| C    | 1,850    | 7,500        | non | bon       | \$1,407,500   |
| D    | 1,450    | 7,000        | non | bon       | \$1,155,000   |
| E    | 1,700    | 8,000        | oui | excellent | \$1,470,000   |
| F    | 1,550    | 7,100        | non | moyen     | \$1,177,500   |
| G    | 1,700    | 10,500       | non | bon       | \$1,392,500   |
| H    | 1,600    | 7,300        | non | bon       | \$1,312,500   |

L'exemple est délibérément assez petit pour être raisonné à la main et assez riche pour mettre chaque concept à l'épreuve. L'entrée 001 examine l'écart de \$145,000 entre les comps A et B comme la contribution conjointe de deux caractéristiques (vue et terrain) enchevêtrées dans une seule observation de ventes appariées. L'entrée 002 place les huit comps comme points d'un espace des caractéristiques à quatre dimensions et montre que des métriques de distance naïves les classent mal. L'entrée 003 propose une forme fonctionnelle pour la fonction de prix hédoniques de ce quartier. L'entrée 004 récupère analytiquement les prix implicites et décompose l'écart initial A-contre-B en une contribution de \$127,039 provenant de la vue et du terrain, avec \$18,000 non expliqués. L'entrée 005 ajuste le modèle formellement dans R et présente la sortie complète de `lm()`, y compris les erreurs types et les résidus. L'entrée 006 examine le seul résidu que le modèle ne peut expliquer — un écart de +\$46,455 sur le comp H — et le relie au cadrage par variables latentes qui sous-tend la méthodologie Residual Constraint Approach développée par ailleurs par l'auteur.

Le jeu de données est disponible dans le dépôt du numéro à `shared/data/pacifica_comps.csv` pour les lecteurs qui souhaitent reproduire l'un ou l'autre des calculs.

## Comment lire ce numéro

Un lecteur qui découvre la méthodologie hédonique devrait lire les six entrées dans l'ordre. Les dépendances sont réelles : chaque entrée suppose acquis le vocabulaire établi par les précédentes, et sauter en avant invite une confusion que les entrées elles-mêmes ne peuvent anticiper.

Un lecteur déjà exposé à la régression hédonique — ce qui inclut la plupart des évaluateurs qui ont suivi un cours de statistique avancée, travaillé avec un AVM ou lu la littérature de l'évaluation de masse — peut utiliser le numéro comme rappel et comme référence. Les entrées sont écrites pour être citables de manière atomique ; un article publié ailleurs dans la revue qui fait référence au *prix implicite* peut renvoyer à l'entrée 004 par son DOI plutôt que de redériver le concept.

Un lecteur intéressé spécifiquement par la méthodologie Residual Constraint Approach devrait lire d'abord l'entrée 006 pour comprendre le cadrage par variables latentes, puis revenir aux entrées 003–005 pour l'appareil hédonique que la RCA prolonge, puis aux entrées 001–002 pour le vocabulaire fondamental sur lequel tout repose.

Un lecteur intéressé par la question de *ce qui rend un ajustement d'évaluation défendable* constatera que la section « Pourquoi cela importe pour la défendabilité » de chaque entrée, lue en séquence, équivaut à un court essai sur le statut méthodologique de l'évaluation par ajustements. Cet essai

n'est assemblé explicitement nulle part dans le numéro, mais il est le sujet implicite de chaque entrée.

## Posture éditoriale

*Foundations* n'est pas un département de manuel scolaire. Chaque entrée est assez courte pour être lue en une seule séance, écrite pour un public déjà rompu à la pratique de l'évaluation, et conçue pour donner une formulation précise d'un concept plutôt qu'un exposé exhaustif. Le but est de fournir un référent stable que les autres articles de la revue peuvent citer.

Les entrées font l'objet d'une révision éditoriale et peuvent être accompagnées d'un commentaire invité d'un expert du domaine. Elles sont publiées sous licence Creative Commons Attribution et reçoivent des DOI à la fois via Crossref (comme citation canonique) et via Zenodo (comme archive versionnée). Chaque entrée peut être affinée au fil du temps — les révisions mineures mettent à jour la version Zenodo, les révisions majeures mettent à jour les deux DOI avec une note éditoriale décrivant le changement. Le DOI Crossref résout toujours vers la version la plus récente ; les versions historiques spécifiques restent adressables via Zenodo.

Les numéros à venir étendront le vocabulaire de *Foundations* à des domaines adjacents : dépendance spatiale, espérance conditionnelle, endogénéité, biais de variable omise, identification, et l'appareil formel qui distingue les variables proxy des variables latentes. Les suggestions de concepts méritant un traitement sont les bienvenues et peuvent être adressées au rédacteur en chef.

---

*Les six entrées de ce numéro :*

1. **Bien hétérogène** — pourquoi la méthodologie d'évaluation doit exister tout court.
2. **Espace des caractéristiques** — la représentation vectorielle des attributs du panier.
3. **Fonction de prix hédoniques** — l'application d'équilibre des caractéristiques vers les prix.
4. **Prix implicite** — le prix de marché marginal de chaque caractéristique.
5. **Régression hédonique** — estimer la fonction à partir des transactions observées, avec un exemple complet travaillé en R.
6. **Variable latente** — récupérer le signal structuré des résidus.

### Références.

- Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.
- Craytor, B. (2025). Residual Constraint Approach : Framework & Protocol. Zenodo. [doi:10.5281/zenodo.14787917](https://doi.org/10.5281/zenodo.14787917)

---

## Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (`000-overview.pl`) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```

%% 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 hétérogène (version française)

Foundations • The Valuation Engineer

Bert Craytor • 2 août 2026

Version 1.0.0

DOI: [10.67330/slscsv65](https://doi.org/10.67330/slscsv65)

---

*Traduction française de l'entrée « Heterogeneous Good » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

**Définition formelle.** Les biens hétérogènes sont des biens qui ne sont pas interchangeables parce que chacun est un panier distinct de caractéristiques que les acheteurs valorisent séparément, et dont les caractéristiques ne se négocient pas séparément.

**Cadrage intuitif.** Un bien homogène est un bien pour lequel n'importe quel exemplaire convient : un baril de brut WTI, un boisseau de maïs jaune n° 2, un kilowattheure du réseau, une action d'une société donnée. Les normes de classement ou les spécifications qui sous-tendent ces biens existent précisément pour rendre les exemplaires individuels interchangeables, de sorte que les prix s'appliquent à la catégorie plutôt qu'à l'exemplaire particulier. Un bien hétérogène est un bien pour lequel l'exemplaire spécifique compte — *cette* maison, pas *celle-là* ; *cette* voiture d'occasion, pas celle d'à côté.

Deux traits rendent un bien hétérogène au sens technique. Premièrement, les acheteurs se soucient de plusieurs attributs distinguables qui varient d'un exemplaire à l'autre sur le marché. Deuxièmement, ces attributs sont groupés en panier par la technologie de production du bien ; ils ne sont pas achetés séparément. On ne peut pas acheter 200 pieds carrés de surface habitable brute isolément. On ne peut pas acquérir un secteur scolaire indépendamment d'une parcelle de terrain. Les caractéristiques arrivent groupées, et le marché fixe un prix sur des paniers entiers plutôt que sur les caractéristiques elles-mêmes.

L'immobilier est le bien hétérogène canonique et présente cette propriété avec une intensité inhabituelle. Les caractéristiques de localisation sont parfaitement non répliquables : deux parcelles ne peuvent occuper le même emplacement, de sorte que chaque bien immobilier possède au moins un attribut (ses coordonnées) qui le distingue de tout autre bien sur le marché. Les attributs physiques, juridiques et environnementaux superposent des dimensions de variation supplémentaires sur un vecteur qui, pris dans son ensemble, est unique à chaque bien.

**Où les évaluateurs rencontrent ce concept.** Chaque ajustement dans la méthode de comparaison des ventes est une reconnaissance implicite que le bien à évaluer et les comparables sont des biens hétérogènes. Si les biens immobiliers étaient homogènes, les prix de vente seraient déjà directement comparables ; aucune grille ne serait nécessaire. La grille est une tentative de réconcilier les prix de paniers distincts en ajustant les différences de composition du panier. Il vaut la peine de noter ce que la grille ne fait *pas* : elle ne décompose jamais un prix de vente en contributions de valeur par caractéristique. La grille traditionnelle n'a pas de colonne pour les contributions de valeur — seulement des ajustements, estimés directement, caractéristique par caractéristique.

La méthode du coût traite l'hétérogénéité en reconstruisant le panier à partir de ses composants — fondations, ossature, finitions, dépréciation — et en tarifant chacun séparément. La méthode par le revenu la traite en abstrayant le panier en un profil unique de flux de trésorerie, les taux de capitalisation et de rendement absorbant l'hétérogénéité au niveau du panier dans un multiplicateur

dérivé du marché. Les trois méthodes d'évaluation sont des réponses méthodologiques au fait sous-jacent que l'immobilier est un bien hétérogène. Elles diffèrent dans la manière de composer avec ce fait, non dans le fait de l'affronter.

Les évaluateurs rencontrent aussi l'hétérogénéité de manière informelle chaque fois qu'ils raisonnent sur la sélection des comparables, l'analyse par ventes appariées ou les comparaisons « à conditions égales ». L'instinct de trouver des comparables semblables sur autant d'attributs que possible est la réponse du praticien à l'hétérogénéité : minimiser la distance de panier entre le bien à évaluer et le comparable, et la composante non observée du panier contribue moins à la différence de prix de vente.

**Pourquoi cela importe pour la défendabilité.** L'hétérogénéité de l'immobilier est la raison pour laquelle la méthodologie d'évaluation doit exister comme discipline. Si les biens immobiliers étaient interchangeables, la valeur de marché serait directement observable à partir des prix affichés ; l'évaluation se réduirait à une consultation. Parce qu'ils ne le sont pas, la valeur doit être inférée à partir des prix de biens apparentés mais non identiques, et chaque étape d'inférence est un choix méthodologique qui doit être justifié.

Reconnaître formellement l'hétérogénéité clarifie plusieurs questions de pratique dont dépendent l'USPAP et la défendabilité au titre des Standards Rules. La sélection des comparables est le choix d'exemplaires, dans l'espace des caractéristiques, assez proches du bien à évaluer pour que les prix implicites qui en sont extraits se généralisent. Le caractère raisonnable d'un ajustement dépend de la question de savoir si le prix implicite sous-jacent peut être étayé par les données de marché. La réconciliation entre les méthodes reflète le jugement de l'évaluateur quant à la méthode qui compose le plus crédiblement avec l'hétérogénéité de ce bien précis. Chacune de ces questions de défendabilité devient plus nette une fois que « bien hétérogène » est nommé comme la condition sous-jacente.

**Exemple d'évaluation travaillé.** Deux ventes récentes dans le même quartier de Pacifica, en Californie :

| Attribut                   | Comp A      | Comp B      |
|----------------------------|-------------|-------------|
| GLA (sf)                   | 1,650       | 1,650       |
| Chambres / Salles de bains | 3 / 2       | 3 / 2       |
| Terrain (sf)               | 9,000       | 7,200       |
| Vue                        | Oui         | Non         |
| Prix de vente              | \$1,425,000 | \$1,280,000 |

Sous un cadrage en biens homogènes, il s'agit du même bien : GLA identique, même nombre de pièces, même quartier. Sous le cadrage en biens hétérogènes, ce sont des paniers distincts, différant par deux caractéristiques (terrain et vue) que les acheteurs valorisent séparément. L'écart de prix de \$145,000 est la contribution *conjointe* de ces deux attributs — deux prix implicites enchevêtrés dans une seule paire.

Récupérer les prix implicites individuellement exige davantage de paires, idéalement en faisant varier un attribut à la fois. Trois ou quatre ventes supplémentaires couvrant les combinaisons vue/sans vue et grand terrain/petit terrain commencent à identifier les prix implicites séparément. Avec assez de variation sur assez de comparables, la régression hédonique opère formellement le démêlage ; avec une variation limitée, l'évaluateur est contraint à un raisonnement encadré par ventes appariées dont la fiabilité dépend explicitement de la proximité des paniers les uns par rapport aux autres dans l'espace des caractéristiques.

L'exemple illustre pourquoi l'hétérogénéité est la condition racine : même deux maisons « identiques » ne sont pas identiques, et l'appareil méthodologique de l'évaluation existe précisément pour traiter la distance entre elles.

### Références.

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.

---

*Renvois : espace des caractéristiques ; fonction de prix hédoniques ; prix implicite.*

---

## Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (001-heterogeneous-good.pl) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```
%% 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).
```

## Espace des caractéristiques (version française)

Foundations • The Valuation Engineer

Bert Craytor • 3 août 2026

Version 1.0.0

DOI: [10.67330/3yymh010](https://doi.org/10.67330/3yymh010)

---

*Traduction française de l'entrée « Characteristics Space » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

**Définition formelle.** L'espace des caractéristiques d'une classe de biens hétérogènes est l'espace vectoriel dont les coordonnées indexent les attributs que les acheteurs valorisent, et dans lequel chaque bien est représenté par le vecteur de ses niveaux d'attributs.

**Cadrage intuitif.** L'entrée précédente a établi qu'un bien hétérogène est un panier distinct de caractéristiques. La question suivante, naturelle, est : comment décrire et comparer formellement de tels paniers ? La réponse, due à Lancaster (1966), consiste à traiter chaque bien comme un point d'un espace vectoriel dont les axes sont les caractéristiques qui importent aux acheteurs.

Dans cette représentation, un bien résidentiel de Pacifica n'est plus « une maison à Park Pacifica » mais un vecteur avec des coordonnées pour la surface habitable brute, la taille du terrain, la vue, l'état, le secteur scolaire, la distance à l'océan, et ainsi de suite. Deux biens partageant toutes ces valeurs d'attributs occuperaient le même point de l'espace des caractéristiques ; des biens différant sur n'importe quel attribut occupent des points différents. La géométrie de l'espace encode la proximité ou la différence entre deux biens, d'une manière qui capture toutes les dimensions sur lesquelles les acheteurs opèrent des distinctions.

La dimensionnalité de l'espace dépend des attributs pertinents pour le marché étudié. Pour une évaluation résidentielle type, l'espace pertinent pourrait avoir une douzaine de dimensions : physiques (GLA, terrain, chambres, salles de bains, âge, qualité, état), de localisation (secteur scolaire, vue, proximité d'aménités ou de nuisances) et juridico-économiques (zonage, servitudes). Pour une évaluation commerciale, les dimensions pertinentes se déplacent vers les attributs de flux de trésorerie (structure des baux, qualité des locataires, taux d'occupation), mais l'idée géométrique sous-jacente est la même.

Ce qui rend l'espace des caractéristiques utile, c'est qu'il convertit une intuition qualitative de praticien (« ces comparables sont semblables au bien à évaluer ») en une affirmation quantitative (« ces comparables sont proches du bien à évaluer dans l'espace des caractéristiques »). Une fois la distance définie, chaque décision de comparaison des ventes peut être examinée comme un énoncé sur la position dans l'espace.

**Où les évaluateurs rencontrent ce concept.** Les évaluateurs travaillent constamment dans l'espace des caractéristiques, généralement sans le nommer. Chaque ligne d'une grille de comparaison des ventes est un axe de coordonnées ; chaque comparable est un vecteur ligne de valeurs de coordonnées ; chaque ajustement est une tentative de se déplacer le long d'un axe en maintenant les autres fixes.

Les huit comparables de Pacifica que nous utiliserons tout au long de ce numéro illustrent directement la construction. Chaque comp est un point d'un espace des caractéristiques à quatre dimensions engendré par la GLA, la taille du terrain, la vue et l'état :

| Comp | GLA (sf) | Terrain (sf) | Vue | État      | Prix de vente |
|------|----------|--------------|-----|-----------|---------------|
| A    | 1,650    | 9,000        | oui | bon       | \$1,425,000   |
| B    | 1,650    | 7,200        | non | bon       | \$1,280,000   |
| C    | 1,850    | 7,500        | non | bon       | \$1,407,500   |
| D    | 1,450    | 7,000        | non | bon       | \$1,155,000   |
| E    | 1,700    | 8,000        | oui | excellent | \$1,470,000   |
| F    | 1,550    | 7,100        | non | moyen     | \$1,177,500   |
| G    | 1,700    | 10,500       | non | bon       | \$1,392,500   |
| H    | 1,600    | 7,300        | non | bon       | \$1,312,500   |

Les comps A et B de l'entrée précédente sont les deux premières lignes ; les six autres élargissent l'exemple de manières que nous exploiterons dans le reste du numéro.

Chaque comp est un point de  $\mathbb{R}^4$  :

- Le comp A est en (1650, 9000, 1, 2).
- Le comp B est en (1650, 7200, 0, 2).
- Le comp G est en (1700, 10500, 0, 2).

Les coordonnées de vue et d'état sont encodées numériquement (la vue dans  $\{0, 1\}$ , l'état dans  $\{1, 2, 3\}$ ) ; le prix n'est *pas* une coordonnée de l'espace des caractéristiques mais une fonction de la position en son sein.

**Pourquoi cela importe pour la défendabilité.** Le cadre de l'espace des caractéristiques précise ce qu'est réellement la sélection des comparables. Un comparable n'est pas simplement « un bien semblable » mais un point de la région pertinente de l'espace des caractéristiques, proche du bien à évaluer. Plus un comparable est proche du bien à évaluer dans l'espace, plus les ajustements requis sont petits et plus la valeur indiquée devient fiable.

Ce recadrage expose plusieurs questions de pratique que la grille informelle masque :

- *Quelles dimensions comptent ?* Le choix des axes est une décision de modélisation. Inclure la « vue » comme axe n'a de sens que si le marché tarife la vue distinctement ; inclure l'« orientation Feng Shui » n'a de sens que si un segment de marché la tarife. Une évaluation défendable exige de défendre les axes, pas seulement les valeurs le long de ceux-ci.
- *Quelle métrique de distance s'applique ?* Deux comparables peuvent différer de \$200,000 en prix parce qu'ils sont éloignés dans l'espace des caractéristiques, mais le calcul de distance exige une mise à l'échelle pour chaque axe. La distance euclidienne non transformée traite un pied carré de GLA comme équivalent à un pied carré de terrain, ce que le marché ne fait pas.
- *Les comparables se trouvent-ils dans une région où la fonction de prix se comporte bien ?* Si le bien à évaluer se situe dans un coin peu échantillonné de l'espace des caractéristiques, aucun comparable n'en est proche et l'indication de valeur est nécessairement extrapolative.

**Exemple d'évaluation travaillé.** Les huit comps de Pacifica occupent une région relativement compacte de l'espace des caractéristiques à quatre dimensions. Considérons la question de savoir quel comp est le « plus proche » d'un bien à évaluer hypothétique situé en (1,700, 7,500, 1, 2) — 1,700 sf, terrain de 7,500 sf, vue, bon état.

Si nous utilisons naïvement la distance euclidienne avec des axes non mis à l'échelle, la dimension du terrain domine parce que ses valeurs (7,000–10,500) sont numériquement bien plus grandes que la vue (0–1) ou l'état (1–3). Distances au bien à évaluer :

| Comp | Distance (euclidienne non mise à l'échelle) | Notes                                    |
|------|---------------------------------------------|------------------------------------------|
| A    | 1,501                                       | vue concordante ; le terrain domine      |
| B    | 308                                         | vue discordante ; terrain proche         |
| C    | 230                                         | vue discordante ; terrain proche         |
| D    | 559                                         | vue discordante ; GLA à l'écart          |
| E    | 502                                         | vue concordante ; état différent         |
| F    | 422                                         | vue et état différents                   |
| G    | 3,000                                       | vue discordante ; terrain très éloigné   |
| H    | 224                                         | vue discordante ; le plus proche de tous |

À la distance brute, le comp H est le plus proche voisin. Mais le comp H n'a pas de vue et le bien à évaluer en a une. La métrique naïve nous a induits en erreur : l'échelle numérique de la dimension du terrain a évincé la dimension catégorielle de la vue, qui compte en réalité davantage pour le prix.

Une métrique de distance défendable mettrait chaque axe à l'échelle de son prix implicite — en gros, du nombre de dollars que vaut une unité de variation sur cet axe. Une fois les axes mis à l'échelle par le prix implicite, la vue (valant ~\$100,000) n'est plus dominée par le terrain (~\$25 par pied carré). Le comp A, qui concorde avec le bien à évaluer sur la vue, émerge alors comme le comparable principal approprié.

C'est précisément ce que fournit l'appareil des prix implicites développé dans les entrées suivantes : une mise à l'échelle raisonnée de l'espace des caractéristiques qui transforme la « proximité » d'une conjecture géométrique en un calcul dérivé du marché.

### Références.

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.

*Renvois : bien hétérogène ; fonction de prix hédoniques ; prix implicite.*

## Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (002-characteristics-space.pl) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```
%% 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).

```

## Fonction de prix hédoniques (version française)

Foundations • The Valuation Engineer

Bert Craytor • 2 août 2026

Version 1.0.0

DOI: [10.67330/9x1a0489](https://doi.org/10.67330/9x1a0489)

---

*Traduction française de l'entrée « Hedonic Price Function » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

**Définition formelle.** La fonction de prix hédoniques est l'application d'équilibre allant des points de l'espace des caractéristiques d'un bien hétérogène vers les prix de transaction, formée comme l'enveloppe des fonctions d'enchère des acheteurs et des fonctions d'offre des vendeurs sur un marché concurrentiel.

**Cadrage intuitif.** Nous disposons maintenant de deux ingrédients : un bien hétérogène (un panier de caractéristiques) et un espace des caractéristiques (un espace vectoriel dans lequel chaque panier est un point). La fonction de prix hédoniques les combine. C'est la fonction  $p(\mathbf{z})$  qui assigne un prix à chaque point  $\mathbf{z}$  de l'espace des caractéristiques, nous indiquant ce que le marché paie pour un panier présentant cette combinaison particulière d'attributs.

L'idée clé de Rosen (1974) est que  $p(\mathbf{z})$  n'est choisie par aucun acheteur ni vendeur individuel. Elle émerge de l'équilibre du marché. Chaque acheteur possède une *fonction d'enchère* exprimant le maximum qu'il paierait pour un panier quelconque, compte tenu de ses préférences et de son revenu. Chaque vendeur possède une *fonction d'offre* exprimant le minimum qu'il accepterait, compte tenu de ses coûts et de ses alternatives. La fonction de prix hédoniques est l'enveloppe de ces enchères et de ces offres — la frontière le long de laquelle acheteurs et vendeurs transigent effectivement.

Pour l'immobilier,  $p(\mathbf{z})$  est la réponse à la question : si vous spécifiez un panier complet de caractéristiques, quel prix le marché produirait-il pour lui ? La réponse n'est pas arbitraire. Elle reflète les préférences agrégées de tous les acheteurs du marché et les coûts et alternatives agrégés de tous les vendeurs, médiés par l'appariement d'équilibre des acheteurs aux paniers.

La fonction de prix hédoniques est l'objet théorique central de l'analyse hédonique. Les prix implicites en sont les dérivées partielles. La régression hédonique en est une estimation empirique. La défendabilité de tout ajustement de comparaison des ventes se réduit en dernière analyse à une affirmation au sujet de cette fonction.

**Où les évaluateurs rencontrent ce concept.** Les évaluateurs raisonnent continuellement sur  $p(\mathbf{z})$ , là encore généralement sans la nommer. Plusieurs activités de la pratique sont des manipulations implicites de la fonction de prix hédoniques :

**Les ajustements dans la méthode de comparaison des ventes.** Lorsqu'un évaluateur ajuste le comp B à la hausse de \$100,000 pour tenir compte du fait que le bien à évaluer a une vue et que le comp B n'en a pas, l'évaluateur affirme une proposition au sujet de  $p(\mathbf{z})$  : à savoir que, toutes les autres caractéristiques étant maintenues fixes, la fonction augmente de \$100,000 lorsque la caractéristique de vue passe de 0 à 1. C'est un énoncé sur la pente locale de  $p$  dans la dimension de la vue.

**La réconciliation entre les méthodes.** La méthode du coût construit une indication de valeur par le bas, en additionnant les coûts des composants d'un panier répliqué. La méthode par le revenu dérive la valeur des flux de trésorerie, qui à l'équilibre reflètent le prix des paniers produisant des flux semblables. La méthode de comparaison des ventes lit le prix directement

dans les transactions observées de paniers semblables. Sur un marché qui fonctionne bien, les trois méthodes devraient converger vers le même point de  $p(\mathbf{z})$ . Lorsqu'elles divergent, l'évaluateur doit juger laquelle a mesuré la fonction le plus fidèlement pour la position du bien à évaluer dans l'espace des caractéristiques.

**L'analyse de marché et les ajustements temporels.** La fonction de prix hédoniques se déplace dans le temps à mesure que l'offre, la demande, le financement et les anticipations changent. Un taux d'appréciation annuel de 6 pour cent est une affirmation sur la façon dont  $p(\mathbf{z})$  s'est déplacée dans son ensemble ; constater que les biens à prime de vue se sont appréciés plus vite que les biens sans vue est une affirmation selon laquelle la fonction s'est déplacée de manière inégale à travers l'espace des caractéristiques.

**Pourquoi cela importe pour la défendabilité.** Nommer explicitement la fonction de prix hédoniques clarifie plusieurs questions fondamentales de défendabilité :

**La fonction existe-t-elle ?** Le cadre de Rosen suppose un marché concurrentiel avec suffisamment de transactions pour qu'une fonction d'équilibre ait un sens. Sur des marchés étroits, sur des marchés en évolution rapide, ou dans des sous-marchés aux bassins d'acheteurs idiosyncrasiques, la fonction peut être mal définie ou instable. Un évaluateur travaillant sur un tel marché devrait reconnaître que la machinerie inférentielle opère sur un fondement théorique fragile.

**La fonction est-elle la même partout ?** Le cadre de Rosen produit une fonction par segment de marché, non une fonction globale. Pacifica et San Francisco sont des marchés distincts avec des fonctions hédoniques distinctes ; les maisons individuelles et les appartements en copropriété sont des sous-marchés distincts même au sein d'une seule ville. Un évaluateur qui confond les segments suppose implicitement qu'une seule fonction couvre les deux, ce qui est généralement faux.

**Quelle forme fonctionnelle  $p$  prend-elle ?** La fonction n'a pas à être linéaire, log-linéaire ou de toute autre forme spécifique. Rosen a seulement montré que  $p$  est l'enveloppe des enchères et des offres ; il n'a pas spécifié sa forme. Empiriquement,  $p$  présente des non-linéarités (la valeur marginale d'une chambre supplémentaire est souvent non constante en la GLA), des interactions (la prime de vue dépend de l'orientation du terrain) et des seuils (une cuisine en dessous d'un certain niveau de qualité peut attirer une décote disproportionnée par rapport à son score d'état). Une estimation défendable exige soit une forme fonctionnelle étayable par le marché, soit un estimateur flexible capable de récupérer la forme à partir des données.

**Exemple d'évaluation travaillé.** Pour le quartier de Pacifica étudié, nous travaillerons avec une forme fonctionnelle additive simple que l'entrée suivante exploitera plus à fond :

$$p(\mathbf{z}) = \beta_0 + \beta_{\text{GLA}} z_{\text{GLA}} + \beta_{\text{lot}} z_{\text{lot}} + \beta_{\text{view}} z_{\text{view}} + \beta_{\text{cond}} z_{\text{cond}}.$$

Cette forme incorpore deux hypothèses fortes : que la contribution de chaque caractéristique est linéaire en son niveau, et que les caractéristiques n'interagissent pas. Les deux hypothèses sont testables et fréquemment violées ; nous les adoptons ici comme point de départ.

Pour les huit comps de l'exemple fil rouge de ce numéro, notre analyse dans les entrées suivantes estimera des valeurs de  $\beta$  qui produisent une fonction ressemblant à ceci :

| Paramètre                          | Valeur estimée |
|------------------------------------|----------------|
| $\beta_0$ (constante)              | \$80,777       |
| $\beta_{\text{GLA}}$ (par sf)      | \$575          |
| $\beta_{\text{lot}}$ (par sf)      | \$22.70        |
| $\beta_{\text{view}}$ (oui/non)    | \$86,179       |
| $\beta_{\text{cond}}$ (par palier) | \$49,824       |

Injectée dans la forme additive, cette fonction prédit un prix pour tout point de l'espace des caractéristiques à quatre dimensions. En substituant les coordonnées du comp A, on obtient un prix prédit proche de son prix observé de \$1,425,000 ; en substituant le bien à évaluer hypothétique en (1700, 7500, 1, 2), on obtient une indication de valeur pour ce bien.

Le propos de cette entrée n'est pas ces chiffres précis mais le geste conceptuel : *la fonction existe, en principe, avant que nous ne l'estimions*. La fonction est une propriété du marché. Notre estimation est une tentative de la récupérer. La récupération peut être partielle, biaisée ou bruitée, et les trois entrées suivantes portent sur la façon de mener cette récupération de manière responsable.

### Références.

- 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.

---

*Renvois : bien hétérogène ; espace des caractéristiques ; prix implicite ; régression hédonique.*

---

## Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (003-hedonic-price-function.pl) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```
%% 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).

```

## Prix implicite (version française)

Foundations • The Valuation Engineer

Bert Craytor • 2 août 2026

Version 1.0.0

DOI: [10.67330/md6wke17](https://doi.org/10.67330/md6wke17)

*Traduction française de l'entrée « Implicit Price » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

**Définition formelle.** Le prix implicite d'une caractéristique est la dérivée partielle de la fonction de prix hédoniques par rapport à cette caractéristique, évaluée en un point de l'espace des caractéristiques ; il représente le prix de marché marginal d'une unité supplémentaire de la caractéristique, toutes les autres caractéristiques étant maintenues fixes.

**Cadrage intuitif.** La fonction de prix hédoniques  $p(\mathbf{z})$  assigne un prix à chaque panier. La plupart des questions pratiques de l'évaluation, cependant, ne portent pas sur le prix d'un panier complet mais sur le prix d'un *composant* spécifique d'un panier. Combien vaut la vue ? Combien ajoutent cent pieds carrés supplémentaires ? Combien contribue la différence entre un état « bon » et « excellent » ?

Le prix implicite répond à ces questions. Le terme provient de Rosen (1974), qui a montré que les prix observés de produits différenciés, joints à leurs caractéristiques, « définissent un ensemble de prix implicites ou "hédoniques" » ; Lancaster (1966) a fourni la vision sous-jacente des biens comme paniers de caractéristiques. Pour chaque caractéristique  $z_i$ , le prix implicite est le taux auquel  $p(\mathbf{z})$  varie lorsque  $z_i$  varie tandis que toutes les autres caractéristiques sont maintenues fixes :

$$\pi_i(\mathbf{z}) = \frac{\partial p(\mathbf{z})}{\partial z_i}.$$

Le prix implicite a pour unités des dollars par unité de la caractéristique — dollars par pied carré de GLA, dollars par pied carré de terrain, dollars par palier d'état. Pour une caractéristique binaire comme la vue, le prix implicite est la prime totale payée pour passer de sans-vue (0) à vue (1).

Le mot « implicite » capture le fait central : ces prix ne sont *pas* affichés. Personne ne propose de vendre « un pied carré de GLA » pour \$575. Le prix est implicite dans les prix d'équilibre des paniers qui, eux, *sont* affichés, et l'appareil hédonique existe pour l'en extraire.

Les prix implicites sont le produit du cadre hédonique pertinent pour l'évaluation. L'ajustement dans une grille de comparaison des ventes est, en principe, une estimation d'un prix implicite multipliée par une différence de caractéristique. Chaque ajustement est un pari sur la valeur de  $\pi_i$  au point pertinent de l'espace des caractéristiques.

**Où les évaluateurs rencontrent ce concept.** La grille d'ajustement de la comparaison des ventes est une table d'applications de prix implicites, même lorsque les évaluateurs ne les nomment pas ainsi. Une ligne de grille qui ajuste un comparable de « +\$10,000 pour 200 sf de GLA de moins que le bien à évaluer » affirme que le prix implicite de la GLA, à l'endroit pertinent de l'espace des caractéristiques, est de \$50 par pied carré. Une ligne de grille qui ajuste « \$100,000 pour vue inférieure » affirme un prix implicite de \$100,000 pour la caractéristique binaire de vue. Toute la défendabilité de la méthode de comparaison des ventes repose sur la question de savoir si ces prix implicites peuvent être étayés par le marché. En pratique, toutefois, les évaluateurs estiment les ajustements directement — à partir de paires appariées, de l'expérience ou de la pratique des pairs — sans jamais calculer les contributions de valeur sous-jacentes ; la grille traditionnelle n'enregistre

que les ajustements eux-mêmes. La lecture en termes de prix implicites est une interprétation de ce que ces ajustements affirment, non une description de la façon dont ils sont dérivés.

Cette dénomination clarifie aussi pourquoi l'analyse par ventes appariées est la technique fondamentale de dérivation des ajustements. Deux comparables qui diffèrent par exactement une caractéristique donnent une estimation directe du prix implicite de cette caractéristique : la différence de prix divisée par la différence de caractéristique. La fiabilité de l'analyse par ventes appariées dépend de la question de savoir si les deux comparables sont vraiment assez proches dans l'espace des caractéristiques pour que le prix implicite qui en est extrait se généralise au bien à évaluer.

Trois propriétés des prix implicites qui comptent pour la pratique :

**Ils sont locaux.** Le prix implicite de la GLA au point de 1,650 sf n'est pas nécessairement le même qu'au point de 2,500 sf. La fonction de prix hédoniques est généralement non linéaire, de sorte que ses dérivées partielles varient avec la position dans l'espace des caractéristiques. Un ajustement dérivé d'une analyse de ventes appariées à 1,500 sf ne s'applique pas nécessairement à un bien à évaluer de 3,000 sf.

**Ils sont conditionnels.** Le prix implicite de la vue dans un sous-marché de front de mer n'est pas le même que dans un sous-marché de l'intérieur. La dérivée partielle est prise en maintenant les autres caractéristiques fixes, mais la valeur de cette dérivée dépend de *l'endroit* où les autres caractéristiques sont maintenues fixes.

**Ils dépendent du marché.** Les prix implicites sont des prix d'équilibre, ils reflètent donc les préférences du bassin d'acheteurs au moment des transactions. Un déplacement du marché peut modifier les prix implicites même si les caractéristiques elles-mêmes sont inchangées. Les ajustements temporels sont implicitement des aveux que les prix implicites d'hier peuvent ne pas être ceux d'aujourd'hui.

**Pourquoi cela importe pour la défendabilité.** Le cadrage en prix implicites expose des questions de défendabilité des ajustements que la grille informelle peut masquer :

**Où vous trouvez-vous dans l'espace des caractéristiques ?** Un prix implicite n'est défendable que s'il est estimé près de la position du bien à évaluer dans l'espace des caractéristiques. Une prime de vue dérivée de comparables de l'intérieur et appliquée à un bien de front de mer extrapole à travers une région de l'espace des caractéristiques où le comportement de la fonction est inconnu.

**Quelle est la géométrie locale ?** Si deux caractéristiques interagissent — par exemple, si la vue vaut davantage sur un plus grand terrain — alors les traiter comme additives séparément est une mauvaise spécification. Le prix implicite de la vue dépend de la taille du terrain, et un ajustement qui ignore l'interaction sous-estime la structure réelle de la fonction.

**Comment le prix implicite a-t-il été obtenu ?** Une analyse par ventes appariées, un coefficient de régression, une règle empirique de la profession et un « jugement professionnel » pointent tous vers la même quantité — le prix implicite — mais avec des statuts épistémiques très différents. La défendabilité de l'évaluateur dépend de sa capacité à dire quelle source a produit le prix implicite utilisé dans chaque ajustement.

**Exemple d'évaluation travaillé.** Pour l'exemple fil rouge de Pacifica avec la spécification additive adoptée dans l'entrée précédente, les prix implicites sont simplement les coefficients de la fonction linéaire :

$$\pi_{\text{GLA}} = \beta_{\text{GLA}}, \quad \pi_{\text{lot}} = \beta_{\text{lot}}, \quad \pi_{\text{view}} = \beta_{\text{view}}, \quad \pi_{\text{cond}} = \beta_{\text{cond}}.$$

Avec les coefficients estimés de l'entrée 003 :

| Caractéristique | Prix implicite | Unités      |
|-----------------|----------------|-------------|
| GLA             | \$575          | par sf      |
| Terrain         | \$22.70        | par sf      |
| Vue             | \$86,179       | par oui/non |
| État            | \$49,824       | par palier  |

Ces prix implicites impliquent une analyse par ventes appariées. Revenons à la différence entre le comp A et le comp B de l'entrée 001 :

- Le comp A a une vue ( $\Delta z_{\text{view}} = +1$ ) ; contribution +\$86,179.
- Le comp A a un terrain plus grand de 1,800 sf ( $\Delta z_{\text{lot}} = +1,800$ ) ; contribution  $1,800 \times \$22.70 = +\$40,860$ .
- Toutes les autres caractéristiques identiques.
- Différence de prix prédite :  $\$86,179 + \$40,860 = \$127,039$ .

La différence de prix observée était de \$145,000. La décomposition en prix implicites en explique environ \$127,000 ; les \$18,000 restants reflètent la variabilité d'échantillonnage dans le petit jeu de données (et, plus généralement, les types de facteurs non observés que l'entrée sur les variables latentes reprendra).

Deux points opérationnels méritent d'être notés. Premièrement, l'hypothèse de linéarité fait un vrai travail : sous une  $p(\mathbf{z})$  non linéaire, la même différence physique de terrain pourrait valoir davantage sur un petit terrain que sur un grand, et la décomposition additive ci-dessus serait biaisée. Deuxièmement, le prix implicite de \$22.70 par pied carré de terrain est la valeur marginale d'un pied carré supplémentaire, non la valeur totale du terrain. La contribution totale d'un terrain de 9,000 sf au prix du panier est  $9,000 \times \$22.70 = \$204,300$ , mais l'ajout d'un pied carré de plus n'augmente le prix du panier que de \$22.70.

Cette distinction — marginal contre total — est la source d'une confusion considérable dans le langage courant de l'évaluation et constitue l'un des gains conceptuels les plus importants de la dénomination explicite du prix implicite.

### Références.

- 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.

---

*Renvois : bien hétérogène ; espace des caractéristiques ; fonction de prix hédoniques ; régression hédonique.*

---

### Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (004-implicit-price.pl) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```
%% 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).

```

## Régression hédonique (version française)

Foundations • The Valuation Engineer

Bert Craytor • 2 août 2026

Version 1.0.0

DOI: [10.67330/ams1t139](https://doi.org/10.67330/ams1t139)

*Traduction française de l'entrée « Hedonic Regression » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

**Définition formelle.** La régression hédonique est l'estimation de la fonction de prix hédoniques à partir des transactions observées sur un marché de biens hétérogènes, produisant des estimations empiriques des prix implicites des caractéristiques qui entrent dans le modèle.

**Cadrage intuitif.** Les entrées précédentes ont décrit la fonction de prix hédoniques et ses prix implicites comme des objets théoriques. La fonction existe en principe, mais en pratique nous n'observons qu'un ensemble fini de transactions : des paniers qui se sont échangés, et les prix auxquels ils se sont échangés. La régression hédonique est le pont empirique entre les deux.

Le dispositif est simple. Chaque transaction fournit une paire  $(\mathbf{z}_i, p_i)$  : le vecteur de caractéristiques du panier échangé et son prix de vente. La collection de ces paires est un échantillon de la fonction de prix hédoniques (non observée). La régression ajuste une forme fonctionnelle à cet échantillon, en estimant des paramètres qui indexent la fonction. Une fois les paramètres estimés, les dérivées partielles de la fonction ajustée produisent des estimations de prix implicites.

Trois choix de modélisation doivent être faits avant que la régression puisse tourner :

1. *Quelles caractéristiques entrent dans le modèle ?* Le choix détermine les dimensions de la fonction estimée. Omettre une caractéristique pertinente fait fuir son effet dans les coefficients des caractéristiques incluses corrélées (biais de variable omise).
2. *Quelle forme fonctionnelle ?* Linéaire, log-linéaire, log-log, polynomiale, à expansion de base, semi-paramétrique, apprentissage automatique. Le choix contraint la forme que la fonction estimée peut prendre et les motifs de prix implicites qu'elle peut récupérer.
3. *Quel estimateur ?* MCO, moindres carrés généralisés, M-estimateurs robustes, régression quantile, estimateurs d'économétrie spatiale, régression pénalisée. Le choix détermine ce que l'estimateur optimise et quelles hypothèses soutiennent son inférence.

Chaque choix est méthodologiquement substantiel et chacun emporte des conséquences de défendabilité.

**Où les évaluateurs rencontrent ce concept.** La régression hédonique est le fondement de l'évaluation de masse et le moteur de la plupart des modèles d'évaluation automatisés (AVM). Lorsqu'un assesseur de comté réévalue 100,000 parcelles chaque année, c'est une régression hédonique sur les ventes récentes qui fait le travail. Lorsque Zillow ou Redfin produit une estimation automatisée, c'est un modèle de type hédonique (souvent un ensemble d'arbres plutôt que des MCO, mais conceptuellement toujours un estimateur hédonique) qui la produit. Lorsque l'AVM de marché secondaire d'un prêteur accepte ou rejette un bien pour un refinancement, la même machinerie est en jeu.

L'évaluation d'un bien individuel s'appuie elle aussi de plus en plus sur la régression hédonique, même lorsque l'indication finale de valeur provient de la comparaison des ventes. Des prix implicites estimés à partir d'une régression sur un échantillon de quartier peuvent étayer les amplitudes d'ajustement de manière bien plus défendable que l'intuition de l'évaluateur ou des pourcentages

issus de règles empiriques. Les coefficients de régression deviennent des preuves, et non des opinions, pour les ajustements appliqués dans la grille.

**Pourquoi cela importe pour la défendabilité.** Les estimations fondées sur la régression présentent un profil de défendabilité distinct de celui de l'analyse par ventes appariées. Les avantages : une régression utilise toutes les ventes disponibles plutôt que deux à la fois ; elle peut isoler l'effet d'une caractéristique tout en contrôlant simultanément beaucoup d'autres ; et elle produit des erreurs types qui quantifient l'incertitude de chaque prix implicite. Les contreparties : une régression impose une forme fonctionnelle qui peut ne pas correspondre à la vraie fonction de prix ; elle exige plus de données qu'une analyse par ventes appariées ; et ses résultats peuvent être sensibles au choix de l'échantillon et de la spécification.

Plusieurs questions de défendabilité reviennent dans l'évaluation fondée sur la régression :

**L'échantillon est-il approprié ?** La régression estime la fonction hédonique telle qu'elle opère dans le segment de marché et la période de l'échantillon. Une régression sur des ventes de l'intérieur appliquée à un bien de front de mer est un échec de défendabilité, aussi propres que paraissent les statistiques.

**La spécification est-elle appropriée ?** Une spécification linéaire est une hypothèse forte. Si la vraie fonction présente des rendements décroissants de la GLA au-delà d'un certain seuil, l'ajustement linéaire estimera mal les prix implicites dans cette plage. Les diagnostics, les graphiques de résidus et les spécifications alternatives font partie du flux de travail défendable.

**Que signifie l'erreur type ?** Une erreur type capture la variabilité d'échantillonnage sous la spécification supposée. Elle ne capture ni la mauvaise spécification du modèle, ni le biais de variable omise, ni la sélection d'échantillon. Une petite erreur type sur une spécification fausse est plus dangereuse qu'une grande erreur type sur une spécification correcte.

**Le résultat se recoupe-t-il avec les autres méthodes ?** Un prix implicite issu d'une régression et défendable devrait être cohérent avec ce que suggère l'analyse par ventes appariées, avec ce qu'impliquent les coûts des composants de la méthode du coût, et avec ce que croient les évaluateurs qui connaissent le marché. Un coefficient de régression en désaccord avec les trois est une invitation à investiguer, pas à publier.

**Exemple d'évaluation travaillé.** Nous pouvons maintenant mettre en œuvre l'ensemble de la machinerie de la régression hédonique sur les huit comps de Pacifica. Le jeu de données est disponible à `shared/data/pacifica_comps.csv`.

**Code R :**

```
comps <- read.csv("shared/data/pacifica_comps.csv")
fit <- lm(price ~ gla + lot + view + cond, data = comps)
summary(fit)
```

**Sortie :**

```
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
```

**Interprétation.** Le modèle ajusté estime la fonction de prix hédoniques comme :

$$\hat{p}(\mathbf{z}) = 80,777 + 575 z_{\text{GLA}} + 22.70 z_{\text{lot}} + 86,179 z_{\text{view}} + 49,824 z_{\text{cond}}.$$

Chaque coefficient est le prix implicite estimé de la caractéristique correspondante. La GLA contribue \$575 par pied carré (avec l'erreur type la plus étroite et la significativité la plus forte) ; la vue contribue environ \$86,000 à un bien de Pacifica ; une unité supplémentaire d'état (par exemple, passer de « bon » à « excellent ») vaut à peu près \$50,000.

**Plusieurs mises en garde sont immédiatement visibles :**

*L'échantillon est minuscule.* Huit observations et quatre prédicteurs laissent trois degrés de liberté résiduels. Les erreurs types sont en conséquence larges, et les valeurs  $p$  pour le terrain, la vue et l'état n'atteignent pas le seuil conventionnel de 0.05 malgré leur plausibilité économique. C'est un artefact de petit échantillon, non une preuve que ces caractéristiques ne comptent pas. Avec 80 comps au lieu de 8, les valeurs  $p$  rétréciraient spectaculairement.

*L'ajustement global est fort.* Un  $R^2$  ajusté de 0.93 et une valeur  $p$  de modèle de 0.012 indiquent que les caractéristiques incluses expliquent conjointement l'essentiel de la variation des prix dans l'échantillon. La combinaison d'un ajustement conjoint fort et d'une significativité individuelle faible est un signal classique de multicolinéarité, qui est ici structurelle : la GLA, le terrain et l'état covarient dans l'échantillon.

*Un résidu se détache.* Le résidu maximal est de +\$46,455, correspondant au comp H. Les sept autres résidus se situent tous dans une bande de  $\pm \$18,000$ . Quelque chose du comp H n'est pas capturé par la fonction ajustée — le modèle sous-prédit systématiquement son prix d'un montant près de trois fois supérieur au résidu suivant.

Ce résidu aberrant n'est pas du bruit. C'est le signal qu'une caractéristique non observée du comp H contribue à son prix d'une manière que les variables incluses ne peuvent récupérer. L'entrée sur la variable latente reprend exactement cette question.

**Validation par le raisonnement des ventes appariées.** L'analyse par ventes appariées de l'entrée 004 prédisait une différence de \$127,039 entre les comps A et B à partir de la vue et du terrain seuls. Les prix implicites de la régression sont identiques (puisque'ils sont la source de ce calcul). La régression généralise la logique des ventes appariées aux huit comps simultanément, mais elle ne produit rien de fondamentalement nouveau à cette taille d'échantillon ; elle confirme l'analyse appariée tout en rapportant des erreurs types que le raisonnement par ventes appariées ne peut fournir.

*Références.*

- 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.

*Renvois : bien hétérogène ; espace des caractéristiques ; fonction de prix hédoniques ; prix implicite ; variable latente.*

## Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (005-hedonic-regression.pl) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```
%% 005-hedonic-regression.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 005
%% Title: Hedonic Regression
%% Author: Bert Craytor
%% Concept: hedonic_regression
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(hedonic_regression, "Hedonic Regression", econometric_method,
% "Hedonic regression is the estimation of the hedonic price
% function from observed transactions in a heterogeneous-goods
% market, yielding empirical estimates of the implicit prices of
% the characteristics that enter the model.",
% published, 'dictionary-2026.1').
%
% term(mass_appraisal, ...).
% term(automated_valuation_model, ...). % synonym: avm
% term(omitted_variable_bias, ...).
% term(multicollinearity, ...).
% term(residual, ...).
% term(ordinary_least_squares, ...).
%
% introduced_by(hedonic_regression, court_1939).
% introduced_by(hedonic_regression, griliches_1961).
% formalized_by(hedonic_regression, rosen_1974).
% depends_on(hedonic_regression, hedonic_price_function).
% depends_on(hedonic_regression, implicit_price).
% -----

:- use_terms([ hedonic_regression, hedonic_price_function, implicit_price,
characteristics_space, heterogeneous_good, latent_variable,
mass_appraisal, automated_valuation_model,
omitted_variable_bias, multicollinearity, residual,
ordinary_least_squares,
sales_comparison_approach, paired_sales_analysis, adjustment ]).
:- use_relations([related/2, depends_on/2,
introduced_by/2, formalized_by/2]).
```

```

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_regression).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.005').

% =====
% LOCAL VOCABULARY
% =====

local_term(specification).
local_term(coefficient).
local_term(standard_error).
local_term(p_value).
local_term(r_squared).
local_term(adjusted_r_squared).
local_term(residual_standard_error).
local_term(degrees_of_freedom).
local_term(f_statistic).
local_term(significance_level).
local_term(diagnostic).

% =====
% THE EMPIRICAL BRIDGE
% =====
%
% Hedonic regression is the bridge from theoretical  $p(z)$  to a sample of
% observed  $(z_i, p_i)$  pairs.

setup_of_regression(
    each_transaction_supplies_pair("(z_i, p_i): characteristics vector and sale price"),
    sample_from("the unobserved hedonic price function"),
    fits("a functional form to that sample, estimating parameters")).

% =====
% THREE MODELING CHOICES (per the article)
% =====

modeling_choice(which_characteristics,
    "Determines the dimensions of the estimated function. Omission causes omitted_variable_bias.").
modeling_choice(functional_form,
    "Constrains what shape the estimated function can take. Linear, log-linear, polynomial, basis-
    ↪ expanded, semiparametric, ML.").
modeling_choice(estimator,
    "OLS, GLS, robust, quantile, spatial, penalized. Determines optimization criterion and
    ↪ inference assumptions.").

% =====
% PRACTICE CONTEXTS WHERE REGRESSION IS THE ENGINE
% =====

regression_engine_context(mass_appraisal,
    "Assessor revaluation of 100,000 parcels annually.").
regression_engine_context(automated_valuation_model,
    "Zillow/Redfin/lender AVMs; tree ensembles are conceptually hedonic estimators.").
regression_engine_context(single_property_appraisal,
    "Regression coefficients support adjustments more defensibly than appraiser intuition.").

```

```
% =====
% DEFENSIBILITY PROFILE: REGRESSION vs PAIRED-SALES
% =====

advantage_over_paired_sales(uses_all_data,
  "Regression uses all available sales rather than two at a time.").
advantage_over_paired_sales(simultaneous_control,
  "Isolates one characteristic while controlling for many others.").
advantage_over_paired_sales(quantified_uncertainty,
  "Produces standard errors quantifying uncertainty in each implicit price.").

tradeoff(functional_form_constraint,
  "Regression imposes a form that may not match the true price function.").
tradeoff(data_requirement,
  "Requires more observations than paired-sales analysis.").
tradeoff(sample_specification_sensitivity,
  "Results can be sensitive to the choice of sample and specification.").

% =====
% RECURRING DEFENSIBILITY QUESTIONS
% =====

defensibility_question(sample_appropriateness,
  "The regression estimates the function for the sample's market segment and time period;
  ↪ mismatch is a defensibility failure regardless of statistical fit.").
defensibility_question(specification_appropriateness,
  "Linear is a strong assumption; diagnostics, residual plots, and alternative specifications
  ↪ belong in the defensible workflow.").
defensibility_question(standard_error_interpretation,
  "Standard errors capture sampling variability under the assumed specification, NOT
  ↪ misspecification, omitted-variable bias, or sample selection.").
defensibility_question(triangulation,
  "Regression coefficients should be consistent with paired-sales, cost components, and local
  ↪ appraiser opinion; isolated coefficients are an invitation to investigate.").

% =====
% WORKED EXAMPLE: lm(price ~ gla + lot + view + cond)
% =====

:- consult('../..kb/pacifica_comps').

% The fitted model.
fit_call("lm(formula = price ~ gla + lot + view + cond, data = comps)").
fit_estimator(ordinary_least_squares).
fit_software(r_lm).

% Fitted coefficients, with standard errors and significance from
% the R output reproduced in the article.
fitted_coefficient(intercept, 80777.07, 164509.35, 0.491, 0.6571, ' ').
fitted_coefficient(gla, 574.93, 109.76, 5.238, 0.0135, '*').
fitted_coefficient(lot, 22.70, 10.38, 2.187, 0.1165, ' ').
fitted_coefficient(view, 86179.03, 31342.97, 2.750, 0.0708, '.').
fitted_coefficient(cond, 49824.33, 27761.27, 1.795, 0.1706, ' ').

% Diagnostics from the R output.
residual_summary(min, -17995).
residual_summary(q1, -8243).
residual_summary(median, -5433).
```

```

residual_summary(q3, -1409).
residual_summary(max, 46455).

model_summary(residual_standard_error, 30481).
model_summary(degrees_of_freedom, 3).
model_summary(r_squared, 0.9707).
model_summary(adjusted_r_squared, 0.9317).
model_summary(f_statistic, 24.86).
model_summary(model_p_value, 0.01231).

% =====
% INTERPRETIVE CAVEATS FROM THE ARTICLE
% =====

caveat(tiny_sample,
    "n=8, four predictors, three residual df. Standard errors are correspondingly wide; lot/view/
    ↪ cond fail to reach the 0.05 threshold despite economic plausibility. Small-sample artifact.").
    ↪

caveat(structural_multicollinearity,
    "Adj R^2 = 0.93 with model p = 0.012 indicates strong joint fit; weak individual significance
    ↪ is a textbook multicollinearity signature. Here GLA, lot, and condition covary in the sample
    ↪ .").

caveat(comp_h_residual,
    "Comp H's residual is +$46,455; the other seven fall within +/- $18,000. The model
    ↪ systematically under-predicts Comp H by ~3x the next-largest residual. NOT noise. Signal of
    ↪ an unobserved characteristic.").

% =====
% RELATING TO PAIRED-SALES (entry 004)
% =====
%
% The regression's implicit prices ARE the values used in entry 004's
% paired-sales decomposition of A vs B. The regression generalizes that
% logic to all 8 comps simultaneously while reporting standard errors
% that paired-sales reasoning cannot produce.

confirms_paired_sales_from(entry_004,
    "A vs B differential of $127,039 from view and lot alone - coefficients are the same;
    ↪ regression confirms the paired analysis while adding standard errors.").

% =====
% RESIDUAL OF COMP H POINTS FORWARD TO ENTRY 006
% =====

points_forward_to(latent_variable,
    "Comp H's anomalous +$46,455 residual is the signal that an unobserved characteristic is
    ↪ contributing to price. Entry 006 takes up the question.").

% =====
% PREDICTED PRICES FROM THE FITTED MODEL
% =====
%
% Computes predicted price from the fitted coefficients. The residual
% for each comp is sale_price - predicted_price.

predicted_price(P, Predicted) :-
    coord(P, gla_sqft, GLA),

```

```

    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond),
    fitted_coefficient(intercept, B0, _, _, _, _),
    fitted_coefficient(gla,      Bg, _, _, _, _),
    fitted_coefficient(lot,     Bl, _, _, _, _),
    fitted_coefficient(view,    Bv, _, _, _, _),
    fitted_coefficient(cond,    Bc, _, _, _, _),
    Predicted is B0 + Bg * GLA + Bl * Lot + Bv * View + Bc * Cond.

residual_of(P, R) :-
    sale_price(P, Observed),
    predicted_price(P, Predicted),
    R is Observed - Predicted.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V
    ; A == view, V == yes -> N = 1
    ; A == view, V == no  -> N = 0
    ; V = N
    ).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).
cross_reference(latent_variable).

```

## Variable latente (version française)

Foundations • The Valuation Engineer

Bert Craytor • 2 août 2026

Version 1.0.0

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

---

*Traduction française de l'entrée « Latent Variable » du département Foundations. Traduction préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

**Définition formelle.** Une variable latente est un attribut non observé qui influence des résultats observés, et dont la valeur doit être inférée de ses effets sur des variables observables plutôt que mesurée directement.

**Cadrage intuitif.** Les cinq premières entrées ont traité le vecteur de caractéristiques  $\mathbf{z}$  comme s'il pouvait être entièrement mesuré : GLA, terrain, vue, état, et ainsi de suite, tous disponibles comme colonnes de données. En pratique, aucun vecteur de caractéristiques n'est jamais complet. Certains attributs ne figurent pas dans les données parce qu'ils sont difficiles à mesurer (qualité des finitions intérieures, fluidité du plan, micro-orientation micro-localisationnelle), parce qu'ils sont absents des flux de données standard (les fiches MLS ne capturent pas la qualité de l'éclairage, l'insonorisation ou les améliorations récentes réalisées sans permis), ou parce qu'ils n'ont pas encore été reconnus comme pertinents (des attributs à l'échelle du quartier qui ne se révèlent prédictifs du prix qu'après que l'analyse les a mis au jour).

Ces attributs non observés mais pertinents sont des *variables latentes*. Ils existent, ils influencent le prix, mais ils ne sont pas dans les colonnes de notre régression. Leurs effets ne disparaissent pas parce qu'ils ne sont pas mesurés ; ils s'accumulent dans les résidus du modèle, où ils apparaissent comme du « bruit » du point de vue des variables incluses mais sont en réalité un signal structuré du point de vue du processus générateur de données sous-jacent.

Les variables latentes sont le pont entre le cadre hédonique idéalisé et la réalité désordonnée des données d'évaluation. Les reconnaître explicitement est une condition préalable à un raisonnement honnête sur l'incertitude du modèle, la structure des résidus et les limites de ce que les caractéristiques observables peuvent expliquer.

**Où les évaluateurs rencontrent ce concept.** Les évaluateurs rencontrent les variables latentes dans trois contextes récurrents.

**Les ajustements de qualité et d'état.** La « qualité » et l'« état » comme catégories d'évaluation sont elles-mêmes des tentatives de condenser un construit latent multidimensionnel en une seule échelle ordinale. Une note d'état C3 dans l'UAD 3.6 est censée résumer l'entretien différé d'un bien, l'usure des finitions, l'âge des systèmes mécaniques et la présentation — quatre à dix attributs sous-jacents condensés en un seul nombre. Le résumé en un seul nombre est un proxy d'un construit latent ; le métier de l'évaluateur inclut le jugement sur la fidélité avec laquelle le proxy capture la vérité sous-jacente dans chaque cas.

**Le résidu inexpliqué dans la grille.** Lorsque trois comparables ajustés indiquent une valeur mais divergent plus qu'attendu, le désaccord est informatif. Il est rarement tout entier du bruit d'échantillonnage ; le plus souvent il reflète un attribut que l'évaluateur n'a pas encore identifié ni ajusté. Enquêter sur le désaccord résiduel — « pourquoi ce comparable indique-t-il \$50,000 au-dessus des autres ? » — est la version informelle, chez l'évaluateur, de l'inférence sur variables latentes.

**L'amélioration récente que personne ne peut voir.** Un bien peut avoir fait l'objet d'une rénovation de cuisine, d'un remplacement de toiture ou de la construction d'un ADU avec permis qui n'est pas encore reflété dans les données MLS ou de l'assesseur. Du point de vue de l'évaluateur, l'amélioration est observable sur place mais pas dans les données. Du point de vue de la régression, l'amélioration est latente. L'évaluateur doit réconcilier l'observation sur place avec l'indication issue des données, et les deux devraient converger vers la contribution réelle de l'attribut latent.

**Pourquoi cela importe pour la défendabilité.** Les variables latentes introduisent plusieurs considérations de défendabilité que le modèle linéaire additif de l'entrée 005 ne traite pas directement.

**Biais de variable omise.** Si une variable latente est corrélée à une variable observée, son effet fuit dans le coefficient de la variable observée. Si les maisons de Pacifica récemment vendues ont tendance à avoir des cuisines rénovées (un attribut latent) et que la date du contrat de vente figure dans les données, alors le coefficient de l'ancienneté de la vente (alias `sale_age` ou « jours depuis la vente ») absorbera une partie de l'effet de rénovation de cuisine. La contribution estimée du prix implicite de `sale_age` devient biaisée — elle reflète `sale_age` *plus* une portion du statut non mesuré de rénovation de cuisine. Une évaluation défendable reconnaît que les coefficients des variables observées sont conditionnels à ce qui manque au modèle.

**Interprétation des résidus.** La théorie standard de la régression traite les résidus comme un bruit gaussien de moyenne nulle. La vision par variables latentes traite les résidus comme un mélange de bruit et de signal : une partie est de la variation d'échantillonnage, une partie est l'effet cumulé des attributs non observés. Identifier laquelle est laquelle est méthodologiquement substantiel. Un résidu qui corrèle avec la localisation, avec la date de vente, avec l'identité de l'évaluateur ou avec toute autre variable observable suggère que la variable latente correspondante peut être partiellement récupérée.

**Sélection de modèle.** Ajouter davantage de variables observées est le remède évident au biais de variable omise, mais les variables latentes qui comptent le plus sont souvent celles qui sont les plus difficiles à ajouter au modèle (le grain fin de l'état, la qualité des améliorations récentes, les micro-effets de quartier). La réponse défendable consiste parfois à reconnaître que le modèle ne les capturera jamais et à allouer la variance inexpliquée au résidu plutôt que de forcer l'ajustement de variables supplémentaires.

**Exemple d'évaluation travaillé.** La régression de l'entrée 005 a produit un résidu inhabituel pour le comp H : le modèle a sous-prédit le prix du comp H de \$46,455 alors que les sept autres résidus se situaient dans une bande de  $\pm \$18,000$ . Les caractéristiques observées du comp H — GLA 1,600, terrain 7,300, sans vue, état « bon » — le placent fermement au milieu de l'échantillon ; rien dans son profil observable ne prédit un prix inhabituellement élevé. Et pourtant il s'est vendu \$46,455 de plus que ce que le modèle dit qu'il aurait dû.

L'hypothèse naturelle est que le comp H possède un attribut latent qui contribue à son prix mais n'apparaît pas dans les données. Dans la construction du jeu de données de ce numéro, c'est exactement le cas : le comp H a fait l'objet d'une rénovation récente de cuisine qui ajoute environ \$60,000 à la valeur de son panier. La rénovation ne figurait pas dans la liste de variables de la régression, de sorte que la contribution s'est accumulée dans le résidu plutôt que dans un coefficient. Le résidu estimé (\$46,455) est légèrement inférieur à la vraie contribution latente (\$60,000) parce que la régression répartit de petites portions de l'effet latent sur les autres coefficients, les biaisant à la hausse dans les dimensions corrélées.

Trois conséquences opérationnelles s'ensuivent :

**Traiter le résidu aberrant comme une preuve, non comme une erreur.** La première réponse appropriée au résidu du comp H n'est pas de l'écarter comme aberrant mais d'enquêter

sur ce qui, dans le bien, explique la sous-prédiction. Dans ce cas, une inspection sur place ou une recherche dans les registres de permis révélerait la rénovation. L'échec du modèle a une valeur diagnostique.

**Reconnaître que les autres coefficients sont légèrement biaisés.** Les coefficients de vue, d'état et de terrain ont absorbé de petites fractions de l'effet de rénovation via les corrélations avec les caractéristiques observées du comp H. Le biais est faible à  $n = 8$  et rétrécirait encore avec plus de comparables et plus de covariables, mais il n'est pas nul.

**Les résidus dans leur ensemble portent une structure récupérable.** C'est l'affirmation substantielle de la méthodologie Residual Constraint Approach (RCA) développée par l'auteur (Craytor, 2025) : dans les données d'évaluation résidentielle, les résidus ne sont systématiquement pas un bruit gaussien iid. Ils reflètent la contribution structurée d'attributs non observés — le grain fin de l'état, le grain fin de la qualité, l'utilité fonctionnelle et d'autres construits latents — qui peuvent être partiellement récupérés en traitant le résidu lui-même comme un signal de variable latente à modéliser plutôt qu'à écarter.

Le traitement RCA de l'état et de la qualité rend cette structure à deux niveaux opérationnelle. Les notes ordinales de l'évaluateur — les proxys à gros grain discutés ci-dessus, que l'on peut étiqueter état-1 (état-gros-grain) et qualité-1 (qualité-gros-grain) — sont assez grossières pour qu'on puisse aussi bien les traiter elles-mêmes comme des variables latentes, mais elles peuvent tout de même être saisies comme régresseurs pour obtenir des contributions approchées de prix implicites. Le protocole déplace ensuite ces contributions estimées hors des attributs mesurés de la section régression de la grille des ventes et les renvoie dans le résidu, d'où elles sont réattribuées aux variables latentes à grain fin état-2 (état-grain-fin) et qualité-2 (qualité-grain-fin) et affinées au regard de preuves supplémentaires telles que les remarques publiques de la MLS. La note à gros grain fournit une attribution de valeur initiale ; la variable latente à grain fin porte la structure récupérable. À la fin, la contribution de valeur à grain fin remplace la contribution de valeur à gros grain.

L'illustration à un seul comparable présentée ici est le cas le plus simple de l'intuition RCA. Avec plus de données et une modélisation plus riche, la structure du résidu peut être partitionnée en facteurs latents reconnaissables, chacun contribuant au prix de marché du panier de manières que la spécification hédonique standard ne peut atteindre. Les variables latentes ne sont pas une nuisance à neutraliser ; elles sont la prochaine frontière de la modélisation défendable en évaluation.

### Références.

- 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)

---

*Renvois : espace des caractéristiques ; fonction de prix hédoniques ; prix implicite ; régression hédonique.*

---

### Addendum : formalisation en Prolog

Le compagnon lisible par machine ci-dessous (006-latent-variable.pl) formalise cette entrée dans la base de connaissances Prolog de la revue. Il est également publié sous forme de galée consultable aux côtés de cet article.

```

%% 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).
```

## Décisions de justice récentes

# Décisions de justice récentes — États-Unis

*Décisions de justice récentes • The Valuation Engineer*

William Bert Craytor • Vol. 1, Numéro 1 • Été 2026

DOI : 10.67330/fz34f711

*Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

---

## Résumés de décisions par ressort

### Californie

**The Retail Property Trust v. Orange County Assessment Appeals Board No. 1.** California Court of Appeal (cour d'appel de Californie), décision d'avril 2026.

La cour d'appel s'est prononcée sur la possibilité d'obtenir un dégrèvement de la valeur d'imposition foncière au titre de l'article 170 du California Revenue and Taxation Code (code californien des recettes et de la fiscalité) à raison des restrictions d'accès imposées par des arrêtés gouvernementaux pris pendant la pandémie de COVID-19. Le centre commercial demandeur soutenait que l'accès restreint durant la pandémie constituait un « dommage » au bien au sens de l'article 170, lui ouvrant droit à un dégrèvement.

*Conséquence pour les experts.* L'interprétation que donne l'arrêt de la notion de « dommage » au sens de l'article 170 orientera les futures demandes de dégrèvement fondées sur des diminutions d'utilité du bien d'origine externe, qu'elles résultent d'arrêtés de santé publique, de projets d'infrastructure ou d'autres actions publiques affectant l'accès. Les experts appelés à se prononcer sur l'incidence de restrictions d'accès sur la valeur d'un bien commercial devraient étudier le raisonnement de la cour sur le point de savoir si une utilité restreinte — par opposition à une atteinte matérielle — constitue un dommage indemnisable aux fins d'un dégrèvement d'assiette.

### Colorado

**Cheroutes v. Rocket Mortgage & Solidifi.** U.S. District Court, D. Colorado (tribunal fédéral de première instance du district du Colorado) ; instance en cours, mises à jour en 2026.

La demande initiale allègue un biais d'évaluation contraire au Fair Housing Act (loi fédérale sur l'égalité d'accès au logement) à propos d'une évaluation réalisée en 2021 sur un duplex de Denver estimé à environ \$640,000 — soit plus de \$220,000 en dessous de la valeur retenue l'année précédente. Un juge fédéral avait auparavant rejeté la requête en irrecevabilité des défendeurs, jugeant que Rocket Mortgage et la société de gestion d'expertises (*appraisal management company*, AMC) Solidifi devaient se défendre au fond plutôt que de décliner toute responsabilité au motif que les prêteurs et les AMC ne sauraient répondre du comportement d'experts indépendants rémunérés à l'acte. L'affaire est au stade de la communication des pièces (*discovery*) en 2026.

*Conséquence pour les experts.* L'affaire est le principal vecteur actif permettant d'éprouver la responsabilité des AMC et des prêteurs, au titre du Fair Housing Act, du fait du comportement des experts rémunérés à l'acte. Si la juridiction retient en définitive la responsabilité de l'AMC et du prêteur au fond, l'ensemble des parties exposées aux actions en discrimination au logement nées d'une

évaluation s'en trouverait sensiblement élargi, avec des répercussions probables sur les contrats de mission entre AMC et experts.

## Floride

**Recours collectif floridien contestant les honoraires d'AMC facturés aux emprunteurs résidentiels.** Introduit le 2 janvier 2026.

La demande conteste les honoraires facturés par les sociétés de gestion d'expertises aux emprunteurs hypothécaires résidentiels, en soutenant que la marge entre les honoraires versés à l'expert et ceux facturés à l'emprunteur à la signature excède les limites fixées par le droit fédéral et fait l'objet d'une information insuffisante. L'affaire se situe au croisement de l'exigence d'honoraires « habituels et raisonnables » issue du Dodd-Frank Act (12 U.S.C. § 1639e) et des régimes d'information du consommateur prévus par le Truth in Lending Act (loi sur la transparence du crédit) et le Real Estate Settlement Procedures Act (loi sur les procédures de règlement immobilier).

*Conséquence pour les experts.* Le litige met à l'épreuve la structure de rémunération des AMC et, s'il est admis comme recours collectif et tranché au fond, pourrait remodeler les pratiques d'information de tout le secteur. L'issue pourrait affecter indirectement l'environnement de fixation des honoraires des experts travaillant par l'intermédiaire d'AMC, en soumettant à l'examen du juge l'écart entre les honoraires payés par l'AMC et ceux facturés à l'emprunteur.

## Géorgie

**Affaires Green Rock relatives aux servitudes de conservation.** United States Tax Court (cour fiscale fédérale des États-Unis) ; arrêts commentés dans *Tax Notes* le 2 mars 2026.

La Tax Court a repris un à un les volets du critère de l'usage optimal (*highest and best use*) pour apprécier des servitudes de conservation (*conservation easements*) consenties sur des parcelles contiguës (la Green Rock Property et Green Rock South) dans la province du Piedmont, en Géorgie. Les experts des requérants affirmaient que l'usage optimal des biens avant l'octroi de la servitude était l'exploitation d'une carrière de granite et se sont appuyés sur une analyse par flux de trésorerie actualisés pour estimer une valeur de marché avant servitude d'environ \$13.75 millions, contre une valeur après servitude de \$180,000.

La juridiction a exprimé des réserves sur chacun des volets du critère. Sur la licéité juridique, elle n'a pas été convaincue qu'une modification du zonage autorisant l'exploitation d'une carrière fût raisonnablement probable, en relevant notamment qu'un changement de classement de zonage pour la Green Rock Property avait été obtenu peu après la commande d'une nouvelle évaluation de servitude de conservation — enchaînement auquel la juridiction a eu du mal à ajouter foi. Sur la possibilité matérielle, elle a estimé que les éléments géologiques produits par les requérants ne suffisaient pas à établir qu'une carrière viable pouvait être aménagée sur la parcelle de Harman Road, nonobstant la présence généralement connue de granite exploitable dans la province du Piedmont. La juridiction a en outre jugé que, même à supposer acquises la licéité juridique et la possibilité matérielle, cela ne serait pas déterminant sans satisfaction des volets restants de l'analyse.

*Conséquence pour les experts.* L'arrêt constitue un exemple traité d'application séquentielle rigoureuse du critère de l'usage optimal à quatre volets. Il conforte la ligne établie par *Ranch Springs, LLC v. Commissioner*, 164 T.C. No. 6 (2025), et *Beaverdam Creek Holdings, LLC v. Commissioner*, T.C. Memo. 2025-53, selon laquelle la Tax Court n'admettra pas de prétentions spéculatives sur l'usage optimal, même lorsque la faisabilité générale de l'usage proposé est documentée dans la

région environnante. Les experts qui préparent des évaluations de servitudes devraient réunir, pour chaque volet, des éléments propres à la parcelle plutôt que de s'appuyer sur une faisabilité régionale ou établie par analogie, et se montrer particulièrement attentifs aux séquences temporelles — tel un changement de zonage obtenu au moment même de la commande de l'évaluation — susceptibles d'appeler le scepticisme du juge.

## Maryland

**Acquisition du Pimlico Race Course par la ville de Baltimore.** Procédure d'expropriation pour cause d'utilité publique (*eminent domain*), en cours en 2025–2026.

L'acquisition du Pimlico Race Course par la Ville de Baltimore au moyen du pouvoir d'expropriation soulève, au titre de la juste indemnité, des questions tenant au rapport entre valeur d'exploitation (*going-concern value*) et valeur du bien immobilier pour les installations à usage spécialisé. Comme l'activité d'exploitation de Pimlico est fonctionnellement indissociable de ses installations matérielles, l'analyse classique par ventes comparables immobilières est d'une utilité limitée, et les évaluations par la méthode du revenu dépendent lourdement de l'hypothèse d'une poursuite de l'activité de courses.

*Conséquence pour les experts.* La procédure mérite d'être suivie par tout expert appelé à évaluer une installation monofonctionnelle dans laquelle l'activité d'exploitation et l'immeuble sont fonctionnellement indissociables : hippodromes, studios de cinéma, parcs à thème, marinas, et certains actifs de santé ou d'hôtellerie. La méthodologie qui se dégagera de ces instances influencera vraisemblablement la manière dont le calcul de la juste indemnité traite les suppléments de valeur d'exploitation dans les futures affaires d'expropriation.

## Actions fédérales fondées sur le Fair Housing Act contre des experts

**Braxton v. Bank of America.** Introduite en avril 2026.

Une propriétaire a assigné Bank of America ainsi qu'un expert établi en Virginie, en alléguant une discrimination raciale contraire au Fair Housing Act. La demanderesse soutient qu'elle-même et ses locataires étaient présents lors de l'expertise et que le comportement de l'expert révélait une animosité raciale. L'affaire prolonge le mouvement amorcé par les précédents contentieux relatifs aux biais d'évaluation — notamment l'affaire *Connolly/Mott v. loanDepot* devant le tribunal fédéral du district du Maryland et l'affaire *Austin v. Miller* dans le comté de Marin, en Californie — fait d'actions individuelles fondées sur le Fair Housing Act contre des experts, des AMC et des prêteurs.

*Conséquence pour les experts.* Le rythme soutenu des actions pour biais d'évaluation dirigées contre des experts pris individuellement, conjugué au repli fédéral en matière de répression des discriminations indirectes (*disparate impact*), indique un déplacement du champ d'exposition : les demandeurs privés et les autorités d'État chargées de l'égalité d'accès au logement sont désormais les principales sources de responsabilité pour biais d'évaluation, plutôt que les agences fédérales. Les experts devraient réexaminer leurs pratiques de documentation afin que la sélection des comparables, l'établissement des ajustements et la réconciliation de la valeur finale soient défendables de façon autonome face à un examen a posteriori, et rester conscients que les propos ou comportements tenus lors de la visite elle-même constituent fréquemment le socle factuel des demandes.

# Décisions de justice récentes — Union européenne

*Décisions de justice récentes • The Valuation Engineer*

William Bert Craytor • Vol. 1, Numéro 1 • Été 2026

DOI : 10.67330/r1yn1y02

*Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

---

## Champ et sources

Ce panorama est établi à partir de sources primaires accessibles au public lorsqu'elles sont disponibles — au premier chef la base officielle de jurisprudence de la Cour de justice de l'Union européenne (curia.europa.eu), HUDOC pour la Cour européenne des droits de l'homme, les décisions publiées du Bundesfinanzhof (bundesfinanzhof.de), la Cour constitutionnelle fédérale allemande (bverfg.de), Légifrance pour la jurisprudence française et l'Upper Tribunal (Lands Chamber) pour le Royaume-Uni — ainsi que des commentaires secondaires parus dans des publications professionnelles et universitaires. Les décisions rendues en langues nationales ont été lues dans leur version originale et sont décrites en anglais dans la version de référence ; les intitulés d'affaires conservent leur orthographe et leurs signes diacritiques d'origine.

Les bases de recherche juridique sur abonnement (juris.de, beck-online, Westlaw UK, Lamy, Aranzadi) ne sont pas utilisées. Les décisions des juridictions nationales inférieures aux cours suprêmes sont recensées de façon sélective, la priorité étant donnée aux décisions d'une portée manifeste sur le plan du précédent ou de la méthode. Les références d'affaires citées ici devraient être vérifiées auprès du recueil officiel de la juridiction avant toute utilisation en pratique ou toute republication.

Cette livraison de référence traite l'Allemagne, la France et le Royaume-Uni comme les principaux ordres juridiques nationaux. L'Italie, l'Espagne, les Pays-Bas et les pays nordiques ne font pas l'objet d'un examen distinct dans la présente livraison, mais seront ajoutés dans les livraisons trimestrielles suivantes selon l'importance de la matière. La Suisse, bien qu'extérieure à l'Union européenne, sera traitée dans de futures livraisons en raison de la forte influence méthodologique des normes suisses d'évaluation sur la pratique européenne en général.

## Contexte réglementaire européen

Deux évolutions réglementaires au niveau de l'Union européenne affectent substantiellement l'environnement de travail des experts, bien qu'aucune ne soit elle-même une décision de justice :

**Règlement européen sur l'IA et modèles d'évaluation automatisée.** Le règlement (UE) 2024/1689 (*EU AI Act*) est entré en vigueur en 2024, ses dispositions applicables aux systèmes d'IA à haut risque prenant effet de manière échelonnée jusqu'en août 2026. Le 19 mai 2026, la Commission européenne a publié un projet de lignes directrices sur la classification des systèmes d'IA à haut risque, soumis à consultation des parties prenantes. L'annexe III du règlement range l'évaluation de la solvabilité des personnes physiques parmi les huit catégories d'usages présumés à haut risque, et le projet de lignes directrices de la Commission développe cette catégorie en détail.

Les modèles d'évaluation automatisée (*automated valuation models*, AVM) utilisés pour l'octroi de crédits hypothécaires, le suivi de portefeuilles hypothécaires et la souscription d'assurance étant fonctionnellement indissociables de l'évaluation de la solvabilité relative aux prêts sous-jacents, les AVM exploités par des établissements de crédit ou des assureurs au sein de l'Union européenne seront, dans la plupart des configurations, qualifiés de systèmes d'IA à haut risque. Les conséquences en matière de conformité comprennent des systèmes obligatoires de gestion des risques, des contrôles de gouvernance des données, une journalisation automatique assortie d'une conservation des journaux d'au moins six mois, des mécanismes de supervision humaine permettant aux opérateurs de comprendre et de passer outre les décisions du système, ainsi que des obligations de surveillance après commercialisation.

*Conséquence pour les experts.* Les experts européens doivent s'attendre à ce que les flux de travail assistés par AVM mis en œuvre par leurs clients prêteurs soient soumis à des exigences de documentation et d'explicabilité sensiblement renforcées. Lorsque les experts fournissent des données d'entrée du modèle, des données de calibrage ou un contrôle humain des sorties d'AVM, ces activités deviennent elles-mêmes des éléments du dispositif de conformité « IA à haut risque » du prêteur.

**Directive sur la performance énergétique des bâtiments (refonte).** La directive (UE) 2024/1275, refonte de la directive sur la performance énergétique des bâtiments (*Energy Performance of Buildings Directive*, EPBD), a été publiée au Journal officiel le 8 mai 2024 et est entrée en vigueur le 28 mai 2024. Le délai de transposition pour les États membres est fixé au 29 mai 2026. Parmi d'autres dispositions, l'EPBD refondue impose aux États membres de donner priorité à la rénovation des seize pour cent les moins performants du parc bâti (non résidentiel) et des quarante-trois pour cent les moins performants du parc résidentiel ; fixe des normes minimales de performance énergétique (*minimum energy performance standards*, MEPS) ; exige des bâtiments à émissions nulles pour les constructions publiques neuves à partir de 2028 et pour toutes les constructions neuves à partir de 2030 ; et rend obligatoire l'évaluation du carbone sur l'ensemble du cycle de vie des bâtiments neufs à partir de 2030.

*Conséquence pour les experts.* L'EPBD refondue affectera les évaluations immobilières par plusieurs canaux : les seuils minimaux de performance énergétique créent une nouvelle catégorie d'obsolescence induite par la conformité ; les notes des diplômes de performance énergétique (*energy performance certificate*, EPC) deviennent un déterminant substantiel du prix, tant pour le parc résidentiel que commercial ; et les coûts de rénovation nécessaires à la mise en conformité doivent être reflétés dans les approches par les coûts et dans la déduction au titre de la durée de vie économique résiduelle. Les experts doivent aussi anticiper que les prêteurs soumis à la directive sur le crédit hypothécaire exigeront de plus en plus un traitement explicite des notes EPC et des besoins d'investissement liés aux MEPS dans les évaluations de garanties.

## Notes pour les livraisons suivantes

Les livraisons trimestrielles qui suivront ce panorama de référence porteront sur les évolutions : nouvelles décisions, issue des affaires pendantes identifiées ici (notamment le traitement, par la Cour constitutionnelle fédérale allemande, de l'affaire *1 BvR 472/26*) et législation de transposition, par les États membres, de l'EPBD refondue en droit national. Le cadre juridique suisse, le traitement des décisions des cours suprêmes italienne et espagnole ainsi que les développements devant le *Conseil constitutionnel* français seront ajoutés selon l'importance de la matière.

## Résumés de décisions par ressort

### Cour de justice de l'Union européenne

**GR Real (affaire C-351/23).** Cour de justice de l'Union européenne, arrêt du 24 juin 2025.

La Cour de justice s'est prononcée sur un renvoi préjudiciel d'une juridiction slovaque né d'une procédure d'exécution hypothécaire extrajudiciaire dans laquelle des consommateurs avaient perdu leur résidence familiale par voie d'adjudication alors qu'une instance contestant des clauses contractuelles potentiellement abusives était encore pendante. La Cour a jugé que les garanties de protection des consommateurs issues de la directive 93/13/CEE exigent que les règles procédurales nationales permettent un contrôle juridictionnel effectif du caractère abusif des clauses avant qu'une résidence hypothéquée puisse être vendue par voie d'exécution extrajudiciaire, et que les procédures autorisant la vente à devenir irréversible avant l'achèvement de ce contrôle sont incompatibles avec la directive.

*Conséquence pour les experts.* La décision n'est pas directement une affaire d'évaluation, mais elle remodèle l'environnement procédural dans lequel les évaluations de garanties hypothécaires sont commandées et utilisées. Dans plusieurs États membres dotés de procédures d'exécution extrajudiciaire (notamment la Slovaquie, la République tchèque, la Hongrie et les États baltes), l'effet pratique est de ralentir le processus d'adjudication et d'allonger la période durant laquelle il est fait fond sur l'évaluation de la garantie du prêteur. Les experts qui préparent des évaluations de garanties destinées à l'exécution dans ces ressorts doivent anticiper que les rapports d'évaluation seront utilisés plus longtemps qu'auparavant et pourront être confrontés aux mouvements de marché intervenus entre la date d'évaluation initiale et la vente finale.

### Cour européenne des droits de l'homme

**Démolition d'un immeuble à Bakou (Azerbaïdjan), arrêt de juillet 2025.** Cour européenne des droits de l'homme, article 1 du Protocole n° 1.

La Cour européenne des droits de l'homme a constaté une violation du droit au respect des biens à propos de la démolition et de l'expropriation d'un immeuble à Bakou, ordonnées à la suite de l'invocation par les autorités d'un risque sismique. La juridiction de Strasbourg a estimé que les autorités internes, en fixant l'indemnisation, n'avaient pas suffisamment motivé le rejet du rapport d'évaluation produit par les requérants et avaient retenu de façon injustifiée le seul montant issu de l'expertise ordonnée par la juridiction elle-même. La Cour a alloué au total €75,000 au titre du préjudice matériel, €3,000 au titre du préjudice moral et €1,055 au titre des frais de justice.

*Conséquence pour les experts.* La décision conforte une ligne jurisprudentielle relative à l'article 1 du Protocole 1 imposant aux juridictions internes d'examiner de façon motivée les éléments d'évaluation concurrents dans les procédures d'expropriation, et exigeant que le rejet du rapport d'évaluation d'un requérant au profit d'une expertise ordonnée par la juridiction soit lui-même motivé. Pour les experts européens exerçant en matière d'expropriation et de juste indemnité, l'affaire souligne que la défendabilité documentaire d'un rapport d'évaluation — le raisonnement explicite étayant la sélection des comparables, l'établissement des ajustements et la réconciliation de la valeur finale — relève elle-même des droits fondamentaux lorsque le rapport est produit à l'appui d'une demande dirigée contre l'État.

## France

**Cour de cassation, Chambre commerciale, 19 juin 2024, n° 22-24.169.** (Commentée dans la littérature professionnelle jusqu'en 2025.)

La Cour de cassation a jugé, dans le contentieux de la *valeur vénale* en matière de droits de succession, qu'une vente du bien hérité intervenant après le fait générateur de l'impôt (le décès du *de cujus*) ne peut être invoquée par les héritiers pour réviser à la baisse la *valeur vénale* initialement déclarée, quand bien même le prix de vente postérieur serait sensiblement inférieur et la vente interviendrait dans les mois suivant le décès. La décision applique le principe selon lequel la *valeur vénale* en matière de droits de succession s'apprécie à la date du fait générateur et n'est pas corrigée au vu de transactions ultérieures portant sur le même bien.

*Conséquence pour les experts.* La décision confirme une discipline stricte de la date d'évaluation dans la pratique française des droits de succession et limite la pertinence des ventes postérieures comme éléments rétrospectifs de valeur. Les experts qui préparent des évaluations successorales françaises devraient réunir des éléments de comparaison complets fondés sur des transactions intervenues *à la date d'évaluation ou avant celle-ci*, et ne devraient pas produire les ventes postérieures du bien lui-même comme éléments confirmatifs. Le même raisonnement a vocation à s'étendre aux *droits de donation* et au contexte de l'*IFI*.

**Cour de cassation, 3<sup>e</sup> Chambre civile, 8 janvier 2026, n° 23-22.803.** Rapports d'expertise amiable (non judiciaire).

La troisième chambre civile de la Cour de cassation a rendu une décision redéfinissant la valeur juridique des rapports d'expertise amiable (non judiciaire) dans la procédure civile française. La décision fait suite à la réforme procédurale plus large opérée par le *décret* du 18 juillet 2025, qui a érigé la médiation et le règlement amiable des différends en règle par défaut de la procédure civile. La Cour a précisé que les rapports d'expertise amiable établis à la demande contractuelle d'une seule partie peuvent néanmoins être reçus comme éléments de preuve et se voir accorder une force probante, sous réserve du principe du contradictoire et de l'appréciation, par la juridiction, de l'indépendance et de la méthodologie de l'expert.

*Conséquence pour les experts.* La décision renforce la position probatoire des *expertises immobilières* établies hors du cadre des *experts judiciaires* désignés par le juge, à condition que le rapport respecte les standards méthodologiques de la profession et que l'indépendance de l'expert soit documentée. Pour les experts français intervenant sur des évaluations contestées en matière successorale, de divorce ou de litiges commerciaux, la décision crée une base plus solide permettant de citer et d'invoquer des rapports commandés par le client dans un contentieux ultérieur. Une attention particulière à la documentation de l'indépendance (absence de lien de subordination, d'intérêt économique ou de lien familial avec le client) et à l'application des méthodes prescrites par les normes professionnelles devient d'autant plus importante.

## Allemagne

**Bundesfinanzhof, arrêts du 12 novembre 2025 (II R 25/24, II R 31/24, II R 3/25).** La trilogie du *Grundsteuer Bundesmodell* — constitutionnalité du système réformé d'évaluation de la taxe foncière. Le deuxième sénat de la Cour fédérale des finances allemande (*Bundesfinanzhof*, BFH) a jugé, dans trois pourvois en révision parallèles, que les dispositions de la méthode d'évaluation par capitalisation des revenus (*Ertragswertverfahren*) appliquée dans le cadre

du *Bundesmodell* pour l'évaluation des biens résidentiels dans onze Länder allemands aux fins de la taxe foncière réformée (*Grundsteuer*) applicable à compter du 1<sup>er</sup> janvier 2025 sont conformes à la Constitution. Les demandeurs étaient des propriétaires d'appartements en Rhénanie-du-Nord-Westphalie, en Saxe et à Berlin, qui contestaient leurs impositions au motif que les règles d'évaluation uniformisées violeraient le principe général d'égalité (article 3, paragraphe 1, de la Loi fondamentale) et produiraient des valeurs systématiquement inexactes.

Le BFH a jugé que le législateur avait créé un système d'évaluation *conceptuellement orienté vers la valeur de marché* (*konzeptionell einer Verkehrswertorientierung folgt*) et conçu pour approcher, en moyenne sur l'ensemble des biens évalués, la « valeur objectivement réaliste du bien » à l'intérieur d'un corridor de la valeur vénale (*gemeiner Wert*). La juridiction a estimé que le législateur n'avait pas excédé la marge d'appréciation qui lui est reconnue pour concilier les objectifs du dispositif d'évaluation et les inégalités qui s'y attachent nécessairement, et que le poids important accordé à la volonté d'éviter un nouveau « retard d'évaluation » par une actualisation périodique automatisée était constitutionnellement admissible. Le BFH a également souligné que les contribuables conservent la faculté, consacrée par ses propres *Beschlüsse* du 27 mai 2024 (II B 78/23 et II B 79/23), de démontrer une valeur vénale (*gemeiner Wert*) inférieure au moyen d'une expertise qualifiée fondée sur le § 198 BewG, sous la réserve que la valeur inférieure soit étayée par un rapport d'expertise établi par un expert immobilier publiquement agréé appliquant des méthodes d'évaluation reconnues.

Un recours constitutionnel (*Verfassungsbeschwerde*), soutenu par le *Bund der Steuerzahler Deutschland* et *Haus & Grund Deutschland*, est désormais pendant devant la Cour constitutionnelle fédérale (*Bundesverfassungsgericht*) sous le numéro 1 BvR 472/26.

*Conséquence pour les experts.* Cette trilogie constitue la jurisprudence allemande la plus significative de la période en matière d'évaluation immobilière. Pour les experts, la principale conséquence pratique est la demande accrue d'expertises « de valeur inférieure » fondées sur le § 198 BewG (*Verkehrswertgutachten zum Nachweis des niedrigeren gemeinen Werts*) dans les cas où le *Grundsteuerwert* déterminé selon le *Bundesmodell* paraît excéder la valeur de marché réelle du bien concerné. La consécration explicite de cette voie par le BFH, conjuguée à sa déférence générale envers le *Bundesmodell*, signifie que le rapport d'expertise est désormais le mécanisme opérant d'allègement individuel du contribuable face à une surévaluation systématique. Les experts doivent s'attendre à une hausse durable des missions de ce type et veiller au respect de la méthodologie ImmoWertV 2021 ainsi que des exigences du § 198 BewG. Le recours constitutionnel pendant devant le BVerfG (1 BvR 472/26) devra être suivi de près pour toute évolution ultérieure de ce cadre, et devrait constituer un sujet principal de la prochaine livraison trimestrielle.

**Bundesfinanzhof, arrêt du 7 octobre 2025, IX R 26/24. Ventilation du prix d'acquisition entre terrain et construction (*Kaufpreisaufteilung*).** Dans une affaire d'impôt sur le revenu portant sur l'assiette de l'amortissement (*Absetzung für Abnutzung, AfA*) d'un immeuble locatif résidentiel, le neuvième sénat de la Cour fédérale des finances s'est prononcé sur la méthode de ventilation du prix d'acquisition contractuel entre la composante foncière non amortissable (*Grund und Boden*) et la composante bâtie amortissable (*Gebäude*). Les demandeurs avaient contesté le recours, par la juridiction du fond, aux *Bodenrichtwerte* (valeurs foncières de référence) au motif que la base de données sous-jacente n'était pas documentée de façon fiable pour la localisation précise du bien.

*Conséquence pour les experts.* La décision prolonge une ligne jurisprudentielle du BFH (notamment

BFH II R 25/24 du 12 novembre 2025, cité ci-dessus, et, antérieurement, BFH IX R 12/21 du 20 septembre 2022) confirmant que l’ImmoWertV est la source des principes d’évaluation reconnus pour la ventilation du prix d’acquisition. L’affaire illustre aussi l’importance de la base de données qui sous-tend les *Bodenrichtwerte* : lorsque les *Gutachterausschüsse* (comités d’experts en évaluation) ne disposent pas de données de ventes comparables suffisantes pour la localisation précise, le *Bodenrichtwert* peut être contesté sur le terrain de la qualité des données. Pour les experts qui préparent des ventilations de prix d’acquisition ou des avis sur l’assiette d’amortissement, cela signifie que la documentation de la base de données du *Bodenrichtwert* — y compris la *Kaufpreissammlung* sous-jacente — fait partie intégrante d’une analyse défendable.

## Royaume-Uni

*Mehrunnisa Alikhan Ahmed Khan v Wolverhampton City Council*. Upper Tribunal (Lands Chamber), 2025.

L’Upper Tribunal (Lands Chamber) a confirmé que la date d’évaluation pertinente pour la fixation de l’indemnité d’expropriation (*compulsory purchase*) est celle à laquelle le bien est transféré à l’autorité expropriante en vertu d’une déclaration générale de transfert de propriété (*general vesting declaration*, GVD). Le tribunal s’est également prononcé sur le rapport entre évaluations contemporaines et rétrospectives : en l’espèce, l’évaluation initiale de la collectivité locale, réalisée immédiatement après la GVD, s’élevait à £150,000, tandis qu’une seconde évaluation, préparée en 2025 en cours d’instance et se référant à la date de transfert présumée trois mois après la notification du propriétaire, s’établissait à £120,000. Nonobstant la confirmation par le tribunal de la date d’évaluation juridiquement pertinente, celui-ci a préféré le chiffre calculé le plus près dans le temps de la date à laquelle la GVD avait été prise, d’où un paiement plus élevé au profit du demandeur.

*Conséquence pour les experts*. La décision est instructive sur le plan méthodologique pour les experts britanniques intervenant en matière d’expropriation. Si la date d’évaluation juridiquement pertinente est celle du transfert, la préférence du tribunal pour l’évaluation contemporaine plutôt que pour une évaluation rétrospective préparée des années plus tard, à la même date d’évaluation juridique, montre que les éléments de marché contemporains conservent un poids pratique même lorsqu’ils ne seraient pas formellement recevables à la date de transfert elle-même. Les experts devraient veiller à ce que les rapports d’évaluation préparés à une date proche d’une GVD soient étayés par des éléments suffisants pour servir de trace principale, en anticipant que des instances ultérieures pourront s’appuyer sur le rapport contemporain de préférence à une reconstitution rétrospective.

# Décisions de justice récentes — Amérique latine

*Décisions de justice récentes • The Valuation Engineer*

William Bert Craytor • Vol. 1, Numéro 1 • Été 2026

DOI : 10.67330/ywgzhg47

*Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

---

## Champ et sources

Ce panorama est établi à partir de sources primaires accessibles au public lorsqu'elles sont disponibles — au premier chef le répertoire de jurisprudence de la Cour interaméricaine des droits de l'homme (corteidh.or.cr), les portails du Tribunal fédéral suprême brésilien (STF) et de la Cour supérieure de justice (STJ) (portal.stf.jus.br, stj.jus.br), le *Semanario Judicial de la Federación* mexicain (sjfsemanal.scjn.gob.mx), la *Corte Suprema de Justicia de la Nación* argentine (csjn.gov.ar), le portail du *Poder Judicial* chilien (poderjudicial.cl) et le *Diario Constitucional*, ainsi que la *Corte Constitucional* colombienne (corteconstitucional.gov.co) — et de commentaires secondaires parus dans des publications professionnelles et universitaires. Les décisions en espagnol et en portugais ont été lues dans leur version originale et sont décrites en anglais dans la version de référence ; les intitulés d'affaires conservent leur orthographe et leurs signes diacritiques d'origine.

Les bases de recherche juridique sur abonnement (vLex, La Ley, Thomson Reuters Proview Latin America) ne sont pas utilisées. Les décisions des juridictions nationales inférieures aux cours suprêmes sont recensées de façon sélective, la priorité étant donnée aux décisions d'une portée manifeste sur le plan du précédent ou de la méthode. Les références d'affaires citées ici devraient être vérifiées auprès du recueil officiel de la juridiction avant toute utilisation en pratique ou toute republication.

Cette livraison de référence traite l'Argentine, le Brésil, le Chili, la Colombie et le Mexique comme les principaux ordres juridiques nationaux. Le Pérou, le Venezuela, l'Équateur, la Bolivie, le Paraguay, l'Uruguay ainsi que les ordres juridiques d'Amérique centrale et des Caraïbes ne font pas l'objet d'un examen distinct dans la présente livraison, mais seront ajoutés dans les livraisons trimestrielles suivantes selon l'importance de la matière. Les travaux d'évaluation liés à la restitution de biens au Venezuela et aux sanctions visant Cuba seront traités dans une livraison future, en raison de leur singularité méthodologique et de leur pertinence pour la pratique transfrontalière.

## Contexte régional

À la différence du panorama consacré à l'Union européenne, ce panorama latino-américain n'a pas d'équivalent direct de la couche réglementaire supranationale que représentent le règlement européen sur l'IA et la refonte de la directive sur la performance énergétique des bâtiments. La réglementation de l'évaluation immobilière en Amérique latine est massivement nationale ; les principales influences supranationales sont la jurisprudence de la Cour interaméricaine des droits de l'homme sur les droits fondamentaux, au titre de l'article 21 de la Convention américaine (droit de propriété), et l'adoption, par la plupart des ordres juridiques de la région, des normes internationales d'information financière

(en particulier IFRS 13, *Fair Value Measurement*) aux fins de l'information financière. Deux thèmes méthodologiques traversent néanmoins la jurisprudence régionale et méritent d'être mentionnés ici :

**Évaluation de la propriété communautaire autochtone.** Le défi méthodologique le plus caractéristique de la pratique latino-américaine de l'évaluation est l'évaluation des propriétés communautaires autochtones et afro-descendantes dans les contextes de restitution, d'expropriation et d'indemnisation. La jurisprudence de la Cour interaméricaine a constamment jugé que la propriété communautaire autochtone est protégée constitutionnellement au même titre que la propriété privée individuelle, mais les méthodes d'évaluation de tels biens — lorsqu'il n'existe typiquement aucune transaction comparable conclue dans des conditions de concurrence normale, que le bien est détenu de façon communautaire et inaliénable, et que la signification culturelle et spirituelle du territoire fait partie intégrante de sa valeur — demeurent peu développées. Fin 2025, vingt-quatre arrêts de la Cour interaméricaine rendus en faveur de peuples autochtones restaient inexécutés ou seulement partiellement exécutés, plusieurs comportant des calculs d'indemnisation en suspens qui dépendent de questions de méthode d'évaluation non résolues.

**Inflation, monnaie et indexation dans l'indemnisation d'expropriation.** Plusieurs ordres juridiques de la région, au premier chef l'Argentine, imposent que les méthodes d'évaluation et les calculs d'indemnisation composent avec une inflation élevée, des *tasaciones* libellées en devises et des régimes d'indexation complexes. La Cour suprême argentine a jugé qu'une indemnisation d'expropriation initialement exprimée en devise étrangère ne peut être convertie en monnaie locale à un taux de parité artificiel, et que l'évaluation doit refléter la valeur économique réelle du bien à la date de l'indemnisation. Des préoccupations analogues valent au Venezuela (en lien avec les évaluations de restitution et celles liées aux sanctions) et, périodiquement, au Brésil en périodes d'inflation élevée. Pour les experts exerçant dans plusieurs ordres juridiques de la région, ces considérations d'inflation et de monnaie font partie intégrante de la méthode d'évaluation d'une manière qui n'a aucun équivalent européen ou américain.

## Notes pour les livraisons suivantes

Les livraisons trimestrielles qui suivront ce panorama de référence porteront sur les évolutions : nouvelles décisions de la Cour interaméricaine, issue des affaires pendantes identifiées ici, avancées dans la mise en place de la *Jurisdicción Agraria y Rural* colombienne en vue de son entrée en fonction en 2027, et développements nationaux relatifs à la pratique de l'évaluation à la juste valeur selon IFRS 13. Le Pérou, le Venezuela, l'Équateur et les ordres juridiques d'Amérique centrale seront ajoutés selon l'importance de la matière. Une attention particulière sera accordée au développement méthodologique de l'évaluation des propriétés communautaires autochtones, domaine dans lequel les solutions de fond retenues par la Cour interaméricaine ont devancé la méthodologie d'évaluation pratiquement disponible pour les experts et les autorités chargées de l'indemnisation.

## Résumés de décisions par ressort

### Cour interaméricaine des droits de l'homme

*Maya Q'eqchi' Agua Caliente vs. Guatemala et Comunidad Garífuna de San Juan y sus Miembros vs. Honduras.* Cour interaméricaine des droits de l'homme, arrêts de 2024, commentés dans la jurisprudence régionale jusqu'en 2025.

Dans deux arrêts contemporains, la Cour interaméricaine s'est prononcée sur les obligations du Guatemala et du Honduras de délimiter, de titrer et de prévenir la dépossession des territoires ancestraux appartenant à des communautés autochtones et afro-descendantes. La communauté Maya Q'eqchi' Agua Caliente n'avait obtenu un titre de propriété définitif sur son territoire qu'après quarante-cinq années de démarches administratives et judiciaires ; la communauté Garífuna de San Juan, au Honduras, faisait face à une dépossession continue de terres côtières ancestrales. Les deux arrêts traitent des obligations de l'État, au titre de l'article 21 (droit de propriété) de la Convention américaine, de reconnaître, délimiter et protéger la propriété communautaire autochtone. Des commentateurs régionaux ont observé que le cadre de restitution retenu par la Cour interaméricaine dans ces décisions est quelque peu plus restrictif que dans ses arrêts antérieurs, en particulier dans son traitement des « tiers de bonne foi » (*terceros inocentes*) ayant acquis des droits concurrents sans connaissance de la revendication autochtone.

*Conséquence pour les experts.* Les décisions confirment que la propriété communautaire autochtone et afro-descendante est protégée constitutionnellement dans l'ensemble du système interaméricain et que l'évaluation de tels biens, en contexte de restitution ou d'indemnisation, ne peut reposer sur des ventes comparables classiques conclues dans des conditions de concurrence normale (lesquelles n'existent généralement pas) sans complément méthodologique. Les experts mandatés dans des contextes de restitution autochtone doivent anticiper que l'évaluation devra intégrer trois composantes distinctes : la valeur de marché sous-jacente de terres présentant des caractéristiques physiques comparables dans la région ; les coûts d'opportunité et les améliorations imputables à l'usage autochtone ; et la signification culturelle et spirituelle non économique du territoire, que la jurisprudence de la Cour interaméricaine traite comme une dimension indemnisable du droit de propriété. La méthodologie relative à la troisième composante demeure peu développée et fait l'objet d'un raisonnement continu de la Cour interaméricaine.

**Mapuche / Consejo de Todas las Tierras vs. Chile.** Cour interaméricaine des droits de l'homme, arrêt de 2024 (commenté dans la jurisprudence de 2025).

La Cour interaméricaine a condamné l'État chilien pour des violations des droits de l'homme commises à l'encontre de 135 membres du peuple mapuche à propos des actions de revendication territoriale du Consejo de Todas las Tierras (1989–1992) et de l'application d'une législation antiterroriste à des revendications territoriales autochtones. L'arrêt prolonge la doctrine de *Norín Catrimán et autres vs. Chile* (2014) et reconnaît explicitement le droit des peuples autochtones à l'autodétermination en lien avec les revendications territoriales.

*Conséquence pour les experts.* Si les principales solutions retenues concernent des garanties de procédure pénale plutôt que directement la méthodologie d'évaluation, la reconnaissance de l'autodétermination autochtone en matière territoriale crée une strate procédurale qui affecte les travaux d'évaluation liés à la restitution territoriale autochtone au Chili et aux contextes voisins. Les experts mandatés dans de tels contextes devraient veiller à ce que la méthodologie intègre des processus de consultation appropriés et reconnaisse l'autodétermination de la communauté quant à la manière dont ses valeurs territoriales sont caractérisées.

## Argentine

**Doctrine argentine relative au Tribunal de Tasaciones de la Nación.** Consolidée par la jurisprudence jusqu'en 2025.

La Cour suprême de justice argentine (CSJN) et les tribunaux fédéraux inférieurs ont continué de

développer la doctrine entourant les évaluations réalisées par le Tribunal de Tasaciones de la Nación (TTN) en matière d'expropriation et d'*expropiación inversa* (expropriation de fait). La doctrine établie de la CSJN (Fallos 313 :82, appliquée à maintes reprises jusqu'en 2024–2025) veut que les évaluations du TTN bénéficient d'une force présomptive en raison de la capacité technique de ses membres, de la base probatoire sur laquelle elles reposent et de l'uniformité de la pratique ; s'écarter d'une évaluation du TTN suppose la démonstration d'une erreur ou d'une omission manifeste. Les tribunaux fédéraux appliquant cette doctrine en 2024–2025 — notamment le Juzgado Federal de Oberá dans *Sosa v. Entidad Binacional Yacyretá* (juin 2024, *expropiación inversa* née du réservoir hydroélectrique de Yacyretá) — continuent de s'appuyer sur les évaluations du TTN comme ancrage opérant de l'indemnisation.

*Conséquence pour les experts.* Pour les experts argentins, la conséquence pratique est que les rapports d'expertise privés produits en contradiction d'une évaluation du TTN supportent une charge probatoire substantielle. Pour écarter ou modifier une évaluation du TTN, le rapport privé doit démontrer une erreur ou une omission méthodologique dans le travail du TTN, et non simplement proposer une autre conclusion de valeur. Les experts qui préparent de tels rapports devraient concentrer leur critique méthodologique sur la sélection des ventes comparables par le TTN, sur l'établissement des ajustements et sur le traitement des conditions de marché locales, plutôt que sur l'ampleur de l'écart de valeur à elle seule.

**Tasaciones en devises et parité de l'indemnisation.** La doctrine de la CSJN selon laquelle une indemnisation d'expropriation initialement exprimée en devise étrangère ne peut être convertie en pesos argentins à une parité artificielle de un pour un demeure applicable jusqu'en 2025. Cette doctrine protège les propriétaires expropriés contre des indemnisations qui, en raison de l'inflation ou de la politique monétaire, ne refléteraient pas la valeur économique réelle du bien à la date de l'indemnisation. Le principe a été périodiquement réaffirmé en 2024–2025 dans des procédures d'expropriation de longue durée.

*Conséquence pour les experts.* Les experts argentins qui préparent des évaluations dans des affaires d'expropriation s'étendant sur des périodes d'inflation ou de changement de politique monétaire significatifs devraient documenter explicitement la base monétaire de l'évaluation et examiner la méthode d'indexation appropriée entre la date d'évaluation et la date prévisible de paiement. Les rapports qui n'abordent pas le risque d'inflation et de change ont peu de chances d'emporter la conviction des tribunaux appliquant la doctrine d'indemnisation de la CSJN.

## Brésil

**Supremo Tribunal Federal, RE 922.144 (Tema 865 de repercussão geral).** Tribunal fédéral suprême brésilien, arrêt de 2024, d'application continue en 2025.

Le STF a jugé que l'indemnisation dans les procédures de *desapropriação* (expropriation) est constitutionnellement compatible avec le régime fédéral du *precatório* institué par l'article 100 de la Constitution fédérale. Plus précisément, la différence entre la somme initialement consignée par la puissance publique (aux fins de l'*imissão provisória na posse*, prise de possession provisoire) et la valeur finalement fixée par l'expertise judiciaire doit être payée via le système du *precatório* plutôt que sous forme de consignation complémentaire immédiate. La décision met en balance l'exigence constitutionnelle d'une indemnisation *justa, prévia e em dinheiro* (article 5, XXIV) et les impératifs d'ordonnancement des paiements du système du *precatório*.

*Conséquence pour les experts.* La décision a des conséquences pratiques directes sur l'évaluation

servant à la consignation initiale, préparée par les experts agissant pour les autorités expropriantes. Toute sous-estimation de la consignation initiale étant payée ultérieurement via le système du *precatório* (lequel peut entraîner des délais considérables), il existe une incitation structurelle pour les demandeurs à contester les évaluations initiales comme sous-évaluées. Les experts qui préparent les évaluations de consignation initiale devraient documenter la méthodologie avec un soin particulier afin de la défendre face à l'expertise judiciaire ultérieure, en anticipant que l'écart entre la consignation initiale et la détermination judiciaire finale sera soumis à des intérêts compensatoires (*juros compensatórios*) selon la jurisprudence du STF et le cadre du Decreto-Lei 3.365/1941.

**Superior Tribunal de Justiça, REsp 2.129.162 (*Tema repetitivo*).** Cour supérieure de justice brésilienne, arrêt de mai 2025.

Le STJ a tranché une controverse répétitive portant sur les honoraires d'avocat dans les affaires où l'autorité expropriante se désiste de l'action en *desapropriação*. La Cour a établi que les honoraires, en cas de désistement, relèvent du cadre de l'article 27, § 1, du Decreto-Lei 3.365/1941 plutôt que des règles procédurales générales du *Código de Processo Civil*. La décision est née d'une action de Cemig (Companhia Energética de Minas Gerais) tendant à l'obtention d'une servitude administrative pour une ligne de distribution électrique.

*Conséquence pour les experts.* La décision est procédurale plutôt que méthodologique, mais elle emporte des conséquences sur la structure des missions d'expertise en matière de servitudes administratives et de *desapropriação*. Les experts désignés comme experts judiciaires (*peritos judiciais*) dans de telles procédures devraient veiller à ce que les conventions d'honoraires soient documentées par référence au cadre du décret-loi applicable plutôt qu'aux règles procédurales générales.

**Superior Tribunal de Justiça, doctrine sur les *juros compensatórios*.** D'application continue en 2025.

Le STJ a maintenu sa doctrine selon laquelle les intérêts compensatoires (*juros compensatórios*) en matière de *desapropriação* ne courent qu'après la décision judiciaire sur le titre du bien exproprié. Les intérêts s'appliquent à la différence entre la consignation initiale et la valeur finale d'indemnisation, au taux historiquement établi de douze pour cent par an (Súmula 618 du STF), et ils visent à indemniser le propriétaire exproprié de la perte anticipée de possession avant le paiement de l'indemnisation définitive.

*Conséquence pour les experts.* Les experts brésiliens doivent anticiper que l'écart d'évaluation qu'ils mettent en évidence dans l'expertise judiciaire — entre la consignation initiale de la puissance publique et la valeur de marché estimée — devient l'assiette du calcul des intérêts compensatoires. La portée économique de cet écart est donc plus grande que ne le laisse penser la différence nominale, en particulier dans les procédures longues.

## Chili

**Corte Suprema, Rol N° 31.423-2025.** Cour suprême du Chili, arrêt de septembre 2025.

La Cour suprême du Chili a rejeté un pourvoi en cassation formé par un promoteur immobilier (*inmobiliaria*) contre le Trésor public (*Fisco de Chile*) dans une procédure spéciale d'expropriation fondée sur l'article 12 du Decreto-Ley N° 2.186. Le demandeur soutenait que l'évaluation fixée par la Commission d'experts (*Comisión de Peritos*) ne reflétait pas l'usage optimal du bien (*mayor y mejor uso*) ni la valeur des constructions existantes et des ouvrages complémentaires, et réclamait en

outre une indemnisation pour manque à gagner (*lucro cesante*) au titre d'un bail existant qui aurait généré des revenus locatifs futurs. La Cour a jugé que les arguments du demandeur revenaient à un désaccord sur l'appréciation des preuves par la juridiction du fond, et non à une véritable violation des normes juridiques gouvernant l'évaluation, et que le raisonnement de la juridiction du fond — lequel s'appuyait sur les éléments de la Commission d'experts, conjugués à des preuves documentaires et testimoniales — était suffisamment motivé.

*Conséquence pour les experts.* La décision illustre la déférence substantielle que la doctrine de la Cour suprême chilienne accorde à l'évaluation de la Commission d'experts dans les procédures d'expropriation, ainsi que la portée limitée du contrôle de cassation sur les conclusions d'évaluation lorsqu'il n'est invoqué qu'un désaccord méthodologique. Les experts qui préparent des évaluations contradictoires dans les affaires chiliennes d'expropriation devraient se concentrer sur une erreur méthodologique démontrable dans le travail de la Commission plutôt que sur un raisonnement d'évaluation alternatif, et rester conscients que les demandes de *lucro cesante* fondées sur des baux existants supposent des exigences probatoires substantielles.

**Corte Suprema (expropriation de Coquimbo).** Cour suprême du Chili, arrêt du 17 juillet 2025.

La Cour suprême du Chili a ratifié une décision d'appel portant l'indemnisation, dans une affaire d'expropriation à Coquimbo, à UF 2.969,75 (sur la base du rapport d'un expert privé) contre le montant de UF 2.349,07 soutenu par l'État (fondé sur des ventes inscrites au *Conservador de Bienes Raíces*). La Cour a souligné que le recours, par la cour d'appel, au rapport de l'expert privé, examiné conjointement avec des témoignages et un rapport d'expertise complémentaire portant sur un bien voisin, relevait d'un exercice admissible de la *sana crítica* (appréciation raisonnée des preuves), et que le contrôle de cassation ne s'étend pas à la remise en cause d'une telle appréciation.

*Conséquence pour les experts.* La décision de Coquimbo est le pendant de l'arrêt Rol N° 31.423-2025 : de même que l'évaluation de la Commission l'emporte lorsqu'elle est contestée par un promoteur, l'évaluation plus élevée d'un expert privé peut également l'emporter lorsqu'elle est soutenue par un raisonnement méthodologique adéquat et des éléments corroborants. La leçon pratique pour les experts chiliens est que la méthodologie et le support documentaire d'une expertise privée sont décisifs, et que les ventes inscrites ne suffisent pas, à elles seules, à écarter une expertise privée bien étayée s'appuyant sur des comparables supplémentaires et des éléments contextuels.

**Tribunal Constitucional, octobre 2025.** *Tasación des bienes raíces* au regard du Code des impôts.

Le Tribunal Constitucional chilien a rejeté une contestation constitutionnelle dirigée contre les dispositions du *Código Tributario* régissant la *tasación* des biens immobiliers (*bienes raíces*) aux fins de l'*impuesto territorial* (taxe foncière). Les ministres dissidents ont soutenu que les normes contestées offraient des mécanismes insuffisants permettant à la partie exécutée de contester la *tasación*, compromettant ainsi la garantie constitutionnelle d'une procédure rationnelle et juste.

*Conséquence pour les experts.* La décision préserve le cadre administratif existant de *tasación* pour la fiscalité immobilière chilienne, mais le raisonnement dissident identifie une fragilité du cadre actuel susceptible d'être reprise dans de futures contestations constitutionnelles ou de droit administratif. Les experts chiliens qui préparent des évaluations dans le cadre de contentieux de fiscalité immobilière devraient suivre l'environnement politique postérieur à la décision, des réformes administratives visant à améliorer les mécanismes de contestation par le contribuable apparaissant plausibles.

## Colombie

**Corte Constitucional, approbation de la *Jurisdicción Agraria y Rural*.** Cour constitutionnelle colombienne, décision de 2025.

La Cour constitutionnelle colombienne a validé, au terme du contrôle de constitutionnalité, la législation créant la nouvelle *Jurisdicción Agraria y Rural*, ordre juridictionnel spécialisé en matière foncière rurale au sein du pouvoir judiciaire colombien. La Cour a approuvé les articles clés définissant l'objet de la réforme, son intégration au pouvoir judiciaire, la création de tribunaux et de juridictions spécialisés, ainsi que les aspects budgétaires. Ce nouvel ordre juridictionnel doit entrer en fonction en 2027 et est conçu pour trancher les litiges de restitution de terres, les litiges de propriété agraire et les questions nées de la mise en œuvre de l'accord de paix de 2016. À la date de l'approbation, des litiges agraires portant sur plus de 61 000 hectares étaient pendants devant le système administratif et judiciaire existant.

*Conséquence pour les experts.* Le nouvel ordre juridictionnel deviendra le forum principal des litiges d'évaluation nés des procédures de restitution de terres et des affaires de propriété rurale en Colombie. Les experts colombiens exerçant en matière de restitution (y compris ceux travaillant pour l'*Unidad de Restitución de Tierras* et l'*Instituto Geográfico Agustín Codazzi*) doivent anticiper que les nouvelles juridictions développeront leur propre jurisprudence d'évaluation et que les standards méthodologiques pourront évoluer par rapport au cadre actuel, piloté par l'administration. Les praticiens devraient suivre l'entrée en fonction en 2027 et la constitution du premier stock d'affaires de ces juridictions.

**Cadre d'évaluation existant fondé sur l'IGAC pour la restitution de terres.** D'application continue en 2025.

La pratique colombienne de la restitution de terres continue de reposer sur l'*avalúo comercial* établi par l'*Instituto Geográfico Agustín Codazzi* (IGAC) comme base principale de l'indemnisation en matière de restitution et de substitution, complété par des *avalúos* établis par les *Lonjas de Propiedad Raíz* lorsque cela est requis. Le cadre est exposé dans la décision SU-034 de 2018 de la Cour constitutionnelle (sur les exigences procédurales d'exécution) et dans des décisions ultérieures, dont T-120 de 2024 (sur le rapport entre l'évaluation de l'IGAC et les préférences individuelles des bénéficiaires dans les situations d'indemnisation en nature).

*Conséquence pour les experts.* Pour les experts intervenant dans la restitution de terres en Colombie, le cadre institutionnel place les évaluations de l'IGAC au centre opérant du calcul de l'indemnisation, les *avalúos* des *Lonjas de Propiedad Raíz* servant de mécanisme complémentaire ou de résolution des controverses. La future *Jurisdicción Agraria y Rural* pourrait modifier cette architecture institutionnelle, et le poids relatif des évaluations de l'IGAC par rapport à celles des *Lonjas* est l'une des questions que les nouvelles juridictions auront vraisemblablement à développer.

## Mexique

**Suprema Corte de Justicia de la Nación, assujettissement à l'ISR de l'écart entre *avalúo* et prix.** Cour suprême du Mexique, décision annoncée début 2026.

Le *Pleno* de la Cour suprême du Mexique (SCJN) a déclaré conformes à la Constitution les articles 125, 130(IV) et 132 de la *Ley del Impuesto sobre la Renta* (ISR, impôt sur le revenu), qui prévoient que lorsque l'*avalúo* d'une acquisition immobilière excède de plus de dix pour cent le prix d'achat contractuel, la différence est traitée comme un revenu accumulable pour l'acquéreur et est donc

soumise à l'ISR. La Cour a jugé que cette règle ne viole pas les principes constitutionnels de proportionnalité et d'équité de l'impôt, au motif que l'acquisition d'un bien à un prix substantiellement inférieur à la valeur de marché représente un accroissement objectif du patrimoine de l'acquéreur et constitue donc à bon droit un avantage économique imposable.

*Conséquence pour les experts.* La décision emporte des conséquences pratiques substantielles pour les transactions immobilières mexicaines et pour la pratique d'expertise qui les accompagne. Les *avalúos* réalisés par des *corredores públicos* ou d'autres professionnels habilités ont désormais des conséquences fiscales directes pour les acquéreurs lorsque l'*avalúo* excède le prix contractuel de plus de dix pour cent. Les experts doivent s'attendre à un examen accru de la méthodologie des *avalúos* transactionnels et à une demande accrue d'*avalúos* documentant avec une rigueur particulière la base des ventes comparables. La décision élève aussi implicitement la portée du choix méthodologique entre approches d'évaluation concurrentes, puisque ce choix peut désormais affecter la dette fiscale de l'acquéreur.

**Suprema Corte de Justicia de la Nación, *Amparo en revisión* 391/2024.** Deuxième chambre de la Cour suprême du Mexique, arrêt du 25 juin 2025.

La deuxième chambre de la Cour suprême du Mexique a tranché une contestation constitutionnelle formée par une société minière contre un décret affectant le cadre juridique de l'expropriation des lots miniers et des droits accessoires connexes (occupation temporaire, servitudes, préférences en matière de concession d'eau et dispositions de renouvellement des baux). La Cour a jugé que le décret ne violait pas le principe de non-rétroactivité, les dispositions affectées portant sur de simples espérances de droits futurs plutôt que sur des droits acquis.

*Conséquence pour les experts.* La décision est importante pour l'évaluation des droits miniers et des biens miniers au Mexique. La qualification, par la Cour, des droits de prorogation des lots miniers et des dispositions connexes comme de « simples espérances » plutôt que comme des droits acquis affecte l'évaluation des concessions minières et l'appréciation du risque réglementaire dans cette évaluation. Les experts qui préparent des évaluations de biens miniers ou de droits connexes au Mexique devraient traiter explicitement la dimension de risque réglementaire de l'évaluation et éviter de traiter des valeurs conditionnelles ou dépendantes d'une prorogation comme si elles étaient pleinement acquises.