

Valuation Engineer Journal

Wertermittlung als Ingenieurdisziplin: gemessen, diszipliniert, reproduzierbar.

Band 1 · Heft 1 · Sommer 2026



IN DIESEM HEFT

Die Lücke zwischen Lehre und Praxis in der Immobilienbewertung

Residual Constraint Approach (RCA): Rahmen und Protokoll

earthUI, glmnetUI & mgcvUI: Offene Modellierungsoberflächen

GRUNDLAGEN (F0–F6) Heterogenes Gut · Merkmalsraum ·

Hedonische Preisfunktion · Impliziter Preis · Hedonische Regression · Latente Variable

AKTUELLE GERICHTSENTSCHEIDUNGEN Vereinigte Staaten · Europäische Union · Lateinamerika

Valuation Engineer Journal

Band 1 · Heft 1
Sommer 2026
ISSN 3143-5033

Chefredakteur William Bert Craytor, SRA

Inhaltsverzeichnis

Auftrag	2
Editorial	3
Zur Eröffnungsausgabe	4
Leserbriefe	6
Eine Einladung zum Briefwechsel	7
Beiträge	9
Die Lücke zwischen Lehre und Praxis in der Immobilienbewertung	10
Residual Constraint Approach (RCA): Rahmen und Protokoll . . .	42
earthUI: Eine interaktive Oberfläche für das Paket earth (MARS)	103
glmnetUI: Eine interaktive Oberfläche für das Paket glmnet (Lasso und Elastic Net)	152
mgcvUI: Eine interaktive Oberfläche für das Paket mgcv (Verall- gemeinerte additive Modelle)	190
Grundlagen	233
Überblick (F0)	234
Heterogenes Gut (F1)	240
Merkmalsraum (F2)	248
Hedonische Preisfunktion (F3)	255
Impliziter Preis (F4)	262
Hedonische Regression (F5)	269
Latente Variable (F6)	277
Aktuelle Gerichtsentscheidungen	284
Aktuelle Gerichtsentscheidungen – USA (R1)	285
Aktuelle Gerichtsentscheidungen – EU (R2)	289
Aktuelle Gerichtsentscheidungen – Lateinamerika (R3)	295

Auftrag

Das *Valuation Engineer Journal* veröffentlicht strenge, praxisorientierte Forschung und Fallstudien zur Ingenieurarbeit am Wert: Methoden, Modelle und Messungen, die technische Entscheidungen mit wirtschaftlichen Ergebnissen verbinden. Wir legen besonderen Wert auf Reproduzierbarkeit, offengelegte Annahmen und klare Darstellung.

Editorial

RUBRIK: EDITORIAL

Willkommen zur Erstausgabe

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft;

die englische Fassung ist maßgeblich.

William Bert Craytor, SRA — Chefredakteur

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

DOI: 10.67330/0tymfa32

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

Dieser Open-Access-Artikel wird unter den Bedingungen der
Creative-Commons-Namensnennung-4.0-Lizenz verbreitet.

Zusammenfassung. Die Botschaft des Herausgebers zu Jahrgang 1, Ausgabe 1: warum es diese Zeitschrift gibt, was die fünf Artikel und drei Rubriken beitragen, und eine Einladung zur Mitwirkung.

Schlüsselwörter: Editorial, Valuation Engineering, Erstausgabe

Willkommen zur Erstausgabe des *Valuation Engineer Journal*.

Diese Zeitschrift verdankt ihre Existenz einer Lücke. Die Bewertungslehre beschreibt die Wertermittlung seit Langem als die disziplinierte Messung des Marktverhaltens, doch die alltägliche Praxis stützt sich nach wie vor weitgehend auf die manuelle Auswahl von Vergleichsobjekten und auf Anpassungen, die allein nach fachlichem Ermessen geschätzt werden. Zugleich sind die statistischen Verfahren, mit denen sich messen lässt, was Märkte tatsächlich für einzelne Objektmerkmale zahlen — hedonische Regression, moderne Spline-Modelle, kreuzvalidiertes maschinelles Lernen —, über Jahrzehnte hinweg gereift, ohne in der praktischen Wertermittlung nennenswerte Verbreitung zu finden. *Valuation Engineer* (Bewertungsingenieur) ist unsere Bezeichnung für den Praktiker, der diese Lücke schließt: staatlich zugelassene gutachterliche Kompetenz, verbunden mit Data Engineering, Statistik und reproduzierbarem Protokoll.

Die fünf Artikel dieser Ausgabe legen das Arbeitsfundament. Der Eröffnungsartikel untersucht die Lücke zwischen Lehre und Praxis selbst und legt dar, dass die berufseigenen Standards bereits mehr Messung verlangen, als die alltäglichen Methoden liefern. An zweiter Stelle — und im Zentrum all dessen, was diese Zeitschrift beabsichtigt — wird der Residual Constraint Approach vollständig dargestellt: ein Protokoll, das den Regressionskanal messen lässt, was messbar ist, das Nichtmessbare zu einer expliziten Residualaufgliederung diszipliniert und die Gesamtsumme so beschränkt, dass das fachliche Urteil den Wert verfeinern, aber niemals erfinden kann. Es folgen drei Software-Artikel, die earthUI und seine Begleitoberflächen für penalisierte Regression

und generalisierte additive Modelle dokumentieren — das offene Werkzeug, auf dem diese Zeitschrift aufbaut —, kurz vor Drucklegung zu einer einzigen Anwendung zusammengeführt, wie die Anmerkungen des Verlags in den jeweiligen Artikeln erläutern.

Drei Rubriken begleiten die Artikel. *Foundations* eröffnet mit sieben kurzen Einträgen (F0–F6), die die begriffliche Kette von heterogenen Gütern über den Merkmalsraum, die hedonische Preisfunktion, implizite Preise und die hedonische Regression bis zu den latenten Variablen nachzeichnen, die in den Residuen stecken. Jeder Eintrag weist seine wissenschaftliche Herkunft nach, trägt eine eigene DOI und schließt mit einer maschinenlesbaren Prolog-Formalisierung — dem Anfang einer Wissensbasis, die wir Ausgabe für Ausgabe zu einem präzisen gemeinsamen Fachvokabular ausbauen wollen. *Recent Court Decisions* sichtet Entscheidungen aus den Vereinigten Staaten, der Europäischen Union und Lateinamerika, die für die Bewertungspraxis von Bedeutung sind, denn das rechtliche Umfeld ist Teil der ingenieurmäßigen Aufgabe. *Letters to the Editor* (Leserbriefe) eröffnet die Rubrik für Zuschriften: Methoden, die zur Anwendung gedacht sind, werden auf Weisen erprobt, die ihre Autoren nie vorhergesehen haben — und genau hierher gehören diese Befunde und die Meinungsverschiedenheiten, die sie auslösen.

Jeder Beitrag dieser Ausgabe erscheint im Open Access unter CC BY 4.0, als PDF und als HTML, mit einer Crossref-DOI und unabhängig davon auf Zenodo archiviert. Wir laden zu Beiträgen ein — Artikel, Foundations-Einträge, Zuschriften und Widerspruch — von Gutachtern, Ökonomen und Ingenieuren gleichermaßen.

William Bert Craytor, SRA
Chefredakteur

Leserbriefe

RUBRIK: LESERBRIEFE

Eine Einladung zum Briefwechsel

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft;

die englische Fassung ist maßgeblich.

William Bert Craytor, SRA — Chefredakteur

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

DOI: 10.67330/r55mw896

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

Dieser Open-Access-Artikel wird unter den Bedingungen der
Creative-Commons-Namensnennung-4.0-Lizenz verbreitet.

Zusammenfassung. Der Herausgeber eröffnet die Rubrik Leserbriefe: was diese Zeitschrift von ihren Leserinnen und Lesern zu erhalten hofft — und auf welchem Weg.

Schlüsselwörter: Leserbriefe, Korrespondenz, fachliche Kommentierung

Eine Zeitschrift, die nur spricht, ist eine halbe Zeitschrift. Diese Rubrik ist für die andere Hälfte da.

The Valuation Engineer Journal veröffentlicht Methoden, die angewendet werden sollen — und angewendete Methoden werden auf Weisen erprobt, die ihre Verfasser nie vorhergesehen haben. Wenn Sie den Residual Constraint Approach angewandt und dabei einen Schritt gefunden haben, der in Ihrem Markt versagt, möchten wir diesen Brief lesen. Wenn ein Foundations-Eintrag einen Begriff enger fasst, als es Ihre Praxis verlangt, möchten wir diesen Brief lesen. Wenn eine von uns zusammengefasste Gerichtsentscheidung sich für einen Praktiker in der betreffenden Jurisdiktion anders liest, möchten wir auch diesen Brief lesen. Widerspruch, Replikation, Verfeinerung und schlichte Richtigstellung gehören sämtlich hierher; ebenso kurze Berichte aus der Praxis — eine Technik, die sich bewährt hat, ein gelöstes Datenproblem, ein Fehlschlag, vor dem andere gewarnt sein sollten.

Der Regeln sind wenige. Briefe sollen unterzeichnet, knapp (eine Seite ist ideal) und konkret sein — nennen Sie den Artikel, den Abschnitt oder den Eintrag, auf den Sie sich beziehen, und belegen Sie Ihre Aussagen so, wie Sie es in einer Arbeitsakte („workfile“) täten. Wir behalten uns Kürzungen und redaktionelle Bearbeitung vor, stets mit Zustimmung der Verfasserin oder des Verfassers. Leserbriefe erscheinen im Open Access unter CC BY 4.0, wie alles andere auf diesen Seiten, und Autorinnen und Autoren, deren Arbeit erörtert wird, erhalten die Gelegenheit, in derselben Ausgabe zu erwidern.

Senden Sie Ihre Briefe über die Website der Zeitschrift (journal.valuation-engineer.com) oder unmittelbar an die Redaktion. Die ersten echten Briefe,

die diese Zeitschrift erreichen, werden den Augenblick markieren, in dem sie von einer Publikation zu einem Gespräch wird — wir sehen ihnen mit Freude entgegen.

William Bert Craytor, SRA
Chefredakteur

Beiträge

Die Kluft zwischen Doktrin und Praxis in der Immobilienbewertung: Eine strukturierte Darstellung von Funktionen, Grenzen und Spannungen

Deutsche Übersetzung des englischen Originalbeitrags im Valuation Engineer Journal.

Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

William Bert Craytor

Jahrgang 1, Ausgabe 1 · 21. Mai 2026

Veröffentlicht am 1. Juli 2026

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

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

Dieser Open-Access-Artikel wird unter den Bedingungen der [Creative-Commons-Namensnennung-4.0-Lizenz](#) verbreitet.

Zusammenfassung. Die Immobilienbewertung nimmt eine Stellung von erheblichem öffentlichem Interesse ein, doch ihre offizielle Doktrin — ausgerichtet auf Marktwertgutachten unter vergangenen und gegenwärtigen Bedingungen — steht in wachsender Spannung dazu, wie Bewertungsprodukte in modernen Finanzprozessen tatsächlich verwendet werden. Dieser Beitrag entwickelt diese These in fünf Stufen, jeweils gekoppelt an eine explizite, abfragbare Prolog-Wissensbasis: Die ersten beiden legen die gesellschaftlichen und wirtschaftlichen Funktionen dar, denen die Bewertung dient; die dritte benennt ihr primäres Endprodukt, eine Marktwerteinschätzung (opinion of market value); und die vierte untersucht die zeitliche und methodische Orientierung der Bewertung im Verhältnis zur quantitativen und finanzwirtschaftlichen Prognostik samt der regulatorischen Grenze, die beide trennt. Diese Stufen laufen auf die zentrale These der fünften zu: In der Hypothekenkreditvergabe fungiert der mit einem Beleihungsauslauf (Loan-to-Value, LTV) skalierte Marktwert als implizite Näherungsgröße für den künftigen Zwangsversteigerungswert — eine zukunftsgerichtete Verwendung, die die Gegenwartswert-Doktrin nicht anerkennt. Viele Gutachter, denen diese Rolle in der Sicherheitenbewertung nicht bewusst ist, behandeln den Marktwert nicht mit der Vorsicht gegenüber künftigen Werttrends, die diese Rolle verlangt — ein Problem, das diese Darstellung offenlegen und zu weiterer Forschung anregen soll.

Schlüsselwörter: Immobilienbewertung, Marktwert, Bewertungsdoktrin, Wissensrepräsentation, Bewertungsstandards, Zwangsversteigerungswert

Inhaltsverzeichnis

1 Gesellschaftliche Bedeutung	2
2 Wirtschaftliche Bedeutung	6
3 Die Bedeutung von Marktwertschätzungen als Endprodukt der Wertermittlung	10
4 Die zeitliche Orientierung der Gutachtertätigkeit	14
5 Marktwert × Beleihungsauslauf als Näherungsgröße für den Zwangsversteigerungswert	21

6 Schlussfolgerung

28

1 Gesellschaftliche Bedeutung

Bewertungen von Immobilien und anderen Vermögenswerten erfüllen eine wesentliche gesellschaftliche und wirtschaftliche Funktion, indem sie Käufern, Verkäufern, Kreditgebern, Versicherern, Investoren, Gerichten, Steuerbehörden und der Allgemeinheit helfen, informierte und gerechte Entscheidungen über die Übertragung, Finanzierung, Besteuerung und Verwaltung von Immobilien und anderen Vermögensformen zu treffen. Hochwertige Gutachten tragen zu Markttransparenz, öffentlichem Vertrauen, effizienter Kapitalallokation und zur Stabilität der Finanzinstitute bei, indem sie unabhängige und gut begründete Werteinschätzungen liefern, die auf Evidenz, Analyse und fachlichem Urteil beruhen. (LaCour-Little und Malpezzi 2003; K’Akumu und Larsen 2024; Hendershott und Kane 1995; Lee, Peng und Liao 2020)

Mangelhafte Gutachten können dagegen auf mehreren Zeithorizonten schwerwiegende Folgen haben. Kurzfristig können ungenaue Gutachten einzelnen Käufern oder Verkäufern unmittelbaren finanziellen Schaden zufügen — durch Überzahlung, Verkauf unter Wert, überhöhte Besteuerung oder unangemessene Kreditentscheidungen. Mittelfristig können systematische Bewertungsmängel Marktsignale verzerren, die Kreditprüfung beeinträchtigen, spekulatives Verhalten begünstigen, zu ungerechter Besteuerung oder unangemessenem Versicherungsschutz beitragen und Rechtsstreitigkeiten wie transaktionales Misstrauen vermehren. Langfristig können anhaltende Qualitätsmängel in der Bewertung das Vertrauen in Märkte und Institutionen untergraben und zu finanzieller Instabilität, Fehlallokation von Ressourcen, dem Niedergang oder der Überinvestition ganzer Wohnviertel sowie zu breiteren sozialen Ungleichheiten beitragen, die Haushalte, Unternehmen und Gemeinwesen treffen. (LaCour-Little und Malpezzi 2003; Mayer und Nothaft 2022)

Da Immobilien und andere Vermögenswerte häufig einen erheblichen Teil des privaten, unternehmerischen und öffentlichen Vermögens ausmachen, ist Bewertungsqualität kein rein privates Fachanliegen, sondern eine Frage von erheblichem öffentlichem Interesse, verknüpft mit wirtschaftlicher Widerstandsfähigkeit, Fairness und dem langfristigen Gemeinwohl. (Mayer und Nothaft 2022)

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

:- module(appraisal_social,
    [ social_horizon/1
    , social_party/1
```

```

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

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

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

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

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

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

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

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

social_involves(property_transfer, buyers).
social_involves(property_transfer, sellers).
social_involves(financing, lenders).

```

```

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

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

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

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

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

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

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

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

%% ---- MEDIUM TERM: low quality ----
social_effect(low, medium_term, markets, distorted_market_signals).
social_effect(low, medium_term, financing,
↳ impaired_credit_underwriting).

```

```

social_effect(low, medium_term, markets,
↳ encouragement_of_speculation).
social_effect(low, medium_term, taxation, inequitable_taxation).
social_effect(low, medium_term, insurance_coverage,
↳ inequitable_insurance_coverage).
social_effect(low, medium_term, adjudication, increased_litigation).
social_effect(low, medium_term, markets, transactional_distrust).

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

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

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

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

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

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

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

social_positive_effect(Horizon, Domain, Effect) :-

```



```

social_effect(high, Horizon, Domain, Effect).

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

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

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

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

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

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

```

2 Wirtschaftliche Bedeutung

Bewertungen von Immobilien und anderen Vermögenswerten spielen eine grundlegende wirtschaftliche Rolle, indem sie das effiziente Funktionieren von Kreditmärkten, Investitionsentscheidungen, Steuersystemen, Finanzberichterstattung, Versicherungszeichnung, Nachlassverwaltung, Unternehmenstransaktionen und öffentlicher Politikplanung unterstützen. Verlässliche Gutachten helfen, glaubwürdige Maßstäbe für Sicherheiten- und Marktwert zu etablieren, die es ermöglichen, dass Kapital effizienter zwischen Kreditnehmern, Kreditgebern, Investoren, Regierungen und Unternehmen fließt. (LaCour-Little und Malpezzi 2003; Ebrahim und Hussain 2010)

Kurzfristig verringern genaue Gutachten transaktionale Unsicherheit, verbessern Kreditentscheidungen, erleichtern faire Verhandlungen und helfen, unmittelbare finanzielle Verluste aus Über- oder Unterbewertung zu vermeiden. Mittelfristig beeinflusst die Bewertungsqualität die Stabilität und Effizienz der Märkte insgesamt, indem sie sich auf Hypothekenrisiko, Investitionsallokation, Steuergerechtigkeit, Angemessenheit des Versicherungsschutzes und das Preisverhalten der Marktteilnehmer auswirkt. Schwache oder systematisch verzerrte Gutachten können in diesem Zeitraum zu aufgeblähten Vermögensblasen, beeinträchtigten Kreditportfolios, verzerrten Entwicklungsanreizen und erhöhtem Ausfall- oder Prozessrisiko beitragen. (LaCour-Little und Malpezzi 2003; Hendershott und Kane 1992; Griffin und Maturana 2016)

Langfristig werden die kumulativen wirtschaftlichen Folgen der Bewertungsqualität noch bedeutsamer: Sie beeinflussen Muster der Kapitalbildung, Infrastrukturinvestitionen, den Vermögensübergang zwischen Generationen, öffentliche Einnahmen, die Erschwinglichkeit von Wohnraum und die Widerstandsfähigkeit des gesamten Finanzsystems. Anhaltende Mängel in der Bewertungspraxis können zu chronischer Fehlallokation von Kapital, Instabilität des Bankensystems, sinkendem Anlegervertrauen und ineffizienter Nutzung von Boden und produk-

tiven Vermögenswerten beitragen, während starke Bewertungsstandards stabile Märkte, solide Wirtschaftsplanung und langfristiges Wirtschaftswachstum stützen. (Ghent, Torous und Valkanov 2019; Hendershott und Kane 1992; Ebrahim und Hussain 2010)

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

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

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

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

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

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

```

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

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

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

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

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

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

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

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

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

%% ---- SHORT TERM: high quality ----

```

```

economic_effect(high, short_term, credit_markets,
↳ reduced_transactional_uncertainty).
economic_effect(high, short_term, credit_markets,
↳ improved_lending_decisions).
economic_effect(high, short_term, corporate_transactions, fair_negotiations).
economic_effect(high, short_term, investment_decisions,
↳ prevention_of_immediate_losses).

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

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

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

%% ---- LONG TERM: high quality ----
economic_effect(high, long_term, investment_decisions, sound_capital_formation).
economic_effect(high, long_term, public_policy_planning,
↳ sound_infrastructure_investment).
economic_effect(high, long_term, estate_administration,
↳ orderly_intergenerational_wealth_transfer).
economic_effect(high, long_term, taxation_systems, stable_public_revenues).
economic_effect(high, long_term, public_policy_planning,
↳ supported_housing_affordability).
economic_effect(high, long_term, credit_markets,
↳ financial_system_resilience).
economic_effect(high, long_term, investment_decisions,
↳ long_term_economic_growth).

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

```

```

economic_effect(low, long_term, public_policy_planning,
↳ inefficient_productive_asset_use).

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

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

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

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

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

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

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

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

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

```

3 Die Bedeutung von Marktwertschätzungen als Endprodukt der Wertermittlung

Marktwertgutachten spielen eine wichtige Rolle in der Sicherheitenrisikoanalyse, weil sie eine fachlich entwickelte Schätzung des wahrscheinlichen Preises liefern, den eine Immobilie oder ein Vermögenswert unter normalen Marktbedingungen erzielen könnte — der Marktwert entspricht dabei in etwa dem deutschen Begriff des Verkehrswerts — und damit als grundlegender Maßstab für Kreditgeber, Investoren und Sicherheitenanalysten dienen, die Kreditexposition und mögliche Zwangsversteigerungsergebnisse beurteilen. (Agarwal, Ben-David und Yao 2015) Obwohl der Zwangsversteigerungs- oder Liquidationswert aufgrund von Notverkaufsbedingungen, verkürzten Vermarktungszeiträumen, Rechtskosten, baulichem Verfall oder negativer Marktstimmung erheblich vom Marktwert abweichen kann, stützen sich Sicherheitenanalysten häufig auf Marktwertgutachten als Ausgangspunkt für die Schätzung des Verwertungspotenzials unter Stressszenarien. (Campbell, Giglio und Pathak 2011; Pennington-Cross 2006; Clauretie und

Daneshvary 2009)

Durch die Analyse des Zusammenhangs zwischen Marktwert, Darlehenssalden, Beleihungsausläufen (Loan-to-Value, LTV), Markttrends, Objektzustand, Liquidität und lokalen Angebots-Nachfrage-Bedingungen beurteilen Sicherheitenanalysten Wahrscheinlichkeit und Schwere von Verlusten im Falle eines Kreditnehmersausfalls. (Qi und Yang 2009)

Kurzfristig helfen genaue Marktwertgutachten den Kreditgebern, umsichtige Kreditvergabeentscheidungen zu treffen und angemessene Sicherheitenmargen festzulegen. Mittelfristig unterstützen sie Portfoliorisikomanagement, Kreditverbriefung, Rückstellungsbildung und die Einhaltung aufsichtsrechtlicher Vorgaben, indem sie den Instituten helfen, die sich wandelnde Angemessenheit der Sicherheiten über Konjunkturzyklen hinweg zu bewerten. Langfristig trägt eine dauerhaft verlässliche Bewertungspraxis zur Stabilität des Finanzsystems bei, indem sie das Risiko übermäßiger Verschuldung, schlecht besicherter Kreditvergabe und systemischer Fehlbewertung von Sicherheiten verringert, die Zwangsversteigerungsverluste und breitere wirtschaftliche Verwerfungen in Phasen von Marktstress verstärken können. (Eriksen u. a. 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 Die zeitliche Orientierung der Gutachtertätigkeit

Gutachter entwickeln in der Regel stichtagsbezogene Marktwerteinschätzungen zu einem bestimmten Bewertungsstichtag — retrospektiv (Vergangenheit), aktuell oder, mit definierten Einschränkungen, prospektiv (Zukunft) — und keine probabilistischen Prognosen über längere Zeiträume. Diese zeitliche und methodische Orientierung stellt eine der zentralen Unterscheidungen zwischen der regulierten Gutachterpraxis und der Arbeit von quantitativen Analysten, Finanzanalysten, Investmentanalysten und Ökonomen dar. (Quan und Quigley 1991)

In den meisten Bewertungsaufträgen, insbesondere solchen mit Marktwertbezug nach Standards wie den Uniform Standards of Professional Appraisal Practice (USPAP, den US-amerikanischen Berufsstandards der Wertermittlung) oder den International Valuation Standards (IVS), erstellt der Gutachter eine abgestimmte Punktschätzung (oder, in zulässigen Fällen, eine enge Spanne), die an beobachtbarer Marktevidenz, den herrschenden Bedingungen und anerkannten Methoden zu einem einzigen Bewertungsstichtag verankert ist. USPAP lässt Wertschlussfolgerungen ausdrücklich als konkreten Betrag, als Zahlenspanne oder als Relation zu einem Referenzwert zu, doch die abgestimmte Einschätzung bleibt eine belastbare Punktindikation, die an diesen Stichtag gebunden ist. (The Appraisal Foundation 2024; International Valuation Standards Council 2025)

Zwar beziehen Gutachter notwendigerweise die vernünftigen Erwartungen der Marktteilnehmer zu Trends, Angebot und Nachfrage, Abschreibung, Kapitalisierungs- oder Diskontierungssätzen und Absorption ein, doch tun sie dies, um einen marktgestützten Wert zum Bewertungsstichtag zu begründen, nicht um spekulative Bewegungen zu prognostizieren. Prospektive Werteinschätzungen (z. B. der stabilisierte Wert nach Projektfertigstellung) erfordern explizite außergewöhnliche Annahmen oder hypothetische Bedingungen und bleiben auf einen

definierten künftigen Bewertungsstichtag mit marktabgeleiteter Stützung beschränkt.

Im Gegensatz dazu begegnen quantitative, Finanz- und Investmentanalysten der Unsicherheit häufig mit Schätzungen, die eine Bandbreite möglicher Ergebnisse über Zeiträume hinweg abdecken, und verwenden dazu Wahrscheinlichkeitsverteilungen, Monte-Carlo-Simulationen, Szenarioanalysen, Sensitivitätstests oder Value-at-Risk-Kennzahlen. Ihr Fokus liegt oft auf der erwarteten künftigen Wertentwicklung, risikoadjustierten Renditen oder dem Investitionswert für eine bestimmte Partei über unterschiedliche Horizonte — und nicht auf einem einzelnen Marktwertpunkt zu einem Stichtag. (Hull 2018)

Solche Arbeit setzt zudem technische Kompetenzen voraus, die eine Gutachterlizenz weder verlangt noch bescheinigt. Das Anpassen eines Modells mit Multivariate Adaptive Regression Splines (MARS) beispielsweise erfordert das Urteilsvermögen, die Modellkomplexität so zu steuern, dass eine starke In-Sample-Anpassung (ein hohes R^2) nicht um den Preis der Überanpassung erkauft wird, sondern sich auf zurückgehaltene Daten überträgt (ein hohes kreuzvalidiertes R^2), was wiederum ein funktionales Verständnis der zugrunde liegenden Algorithmen voraussetzt. (Friedman 1991; Steurer, Hill und Pfeifer 2021) Neben solcher quantitativen Fertigkeit belohnt die Arbeit die konzeptuelle Fähigkeit, mit hochstrukturierten Begriffen umzugehen, feine Bedeutungsunterschiede zu erkennen und sich unvertraute Methoden anzueignen. Dies sind Fähigkeiten, keine Berechtigungsnachweise: Ein lizenzierter Gutachter kann sie besitzen und ein nicht lizenzierter Analyst kann sie entbehren — genau deshalb entscheiden sie für sich genommen nicht darüber, auf welcher Seite der regulatorischen Linie eine gegebene Aufgabe liegt.

Gutachter können sich auch mit begrenzten zukunftsgerichteten Elementen befassen, etwa der Projektion von Einnahmeströmen, Leerstand, Kosten und Restwerten in Discounted-Cashflow-Analysen oder der Schätzung prospektiver Werte in Entwicklungs- oder Prozesskontexten. Diese Komponenten bleiben innerhalb der Grenzen der Wertermittlung, wenn sie in aktueller Marktevidenz und in Berufsstandards gegründet sind, die Belegbarkeit und die Vermeidung unbelegter Spekulation betonen. (Diaz und Wolverton 1998; Gallimore 1996; Crosby, Lizieri und McAllister 2010) Die moderne Bewertungspraxis integriert zunehmend statistische Werkzeuge, Szenarioüberlegungen und Risikoanalytik — insbesondere in der Massenbewertung, bei automatisierten Bewertungsmodellen (AVMs) oder in der Portfolioarbeit —, doch das Kernprodukt bleibt typischerweise eine stichtagsbezogene Marktwerteinschätzung. (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 und Neal 2023)

Man könnte die Gutachterpraxis als primär evidenzbasiert und auf Punktschätzungen von der Vergangenheit bis zur Gegenwart (oder begrenzt prospektiv) ausgerichtet charakterisieren, während die finanz- und quantitativ-analytische Arbeit stärker probabilistisch, verteilungsorientiert und ausdrücklich zukunftsgerichtet über Zeithorizonte hinweg ist. (Geltner 1991) Gutachter fragen: „Welche Marktwertindikation ist zum festgelegten Bewertungsstichtag angesichts der aktuellen Evidenz und der definierten Annahmen am besten belegbar?“ Finanz- und quantitative Analysten fragen häufiger: „Welche Bandbreite von Ergebnissen ist über einen künftigen Zeitraum wahrscheinlich, mit welchen Wahrscheinlichkeiten, und wie sollten Ressourcen alloziert werden?“

Die Grenze ist nicht absolut und entwickelt sich mit der Komplexität der Märkte weiter. Gleichwohl bestehen grundlegende Unterschiede in Zweck (Marktwert vs. Investitionswert), Anforderungen an die Evidenz und regulatorischen Rahmenwerken fort.

Die maßgebliche Linie wird jedoch nicht durch Können, sondern durch das Recht gezogen: Die entscheidende Frage ist schlicht, ob eine gegebene Aufgabe eine solche ist, für die eine Gutachterlizenz gesetzlich vorgeschrieben ist. Quantitative Analysten, Finanzanalysten, Ökonomen, Investmentanalysten, Portfoliomanager und verwandte Berufsgruppen dürfen künftige Vermögenswerte legal schätzen, prognostizieren, modellieren oder analysieren — einschließ-

lich verteilungs- oder spannenbasierter Projektionen — ohne Lizenz als Immobiliengutachter, sofern die Arbeit keine regulierte Wertermittlung darstellt. (Murphy 2012) Die wesentlichen Unterscheidungen beruhen auf:

1. der Art des Werts (z. B. eine Marktwert-Punktschätzung vs. Investitionswert oder eine probabilistische Prognose);
2. dem zeitlichen Fokus (ein bestimmter Bewertungsstichtag vs. ein Zeitraum oder Horizont);
3. der Form des Ergebnisses (ein Punkt oder eine enge Spanne vs. eine vollständige Wahrscheinlichkeitsverteilung);
4. dem beabsichtigten Zweck und der Frage, ob die Arbeit als Wertermittlung (appraisal) ausgegeben wird; und
5. der Beteiligung regulierter Kreditvergabe- oder Rechtskontexte, die die Einhaltung von Standards wie USPAP erfordern.

Dieses Rahmenwerk entstand größtenteils aus Bankenregulierungen wie FIRREA (1989, dem US-Gesetz zur Reform der Finanzinstitute) in den Vereinigten Staaten, die glaubwürdige, unabhängige Marktwerteinschätzungen für Transaktionen mit Bundesbezug sicherstellen sollten. (United States Congress 1989; K'Akumu und Larsen 2024) Viele anspruchsvolle Bewertungstätigkeiten — Kursziele der Aktienanalyse, Machbarkeitsstudien, Portfoliorisikomodellierung, Optionsbepreisung oder prädiktive Analytik — fallen nicht unter die Lizenzpflicht, weil sie Investitions-, Prognose- oder Beratungszwecken dienen und keine regulierten Bewertungsfunktionen erfüllen.

```

%% 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_stateLaw).

regulation_legal_exposure_condition(holding_self_out_as_licensed_appraiser).
regulation_legal_exposure_condition(preparing_appraisal_as_defined_by_stateLaw).
regulation_legal_exposure_condition(involved_in_transaction_requiring_licensed_appraiser).

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

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

:- dynamic activity_satisfies_definition/1.

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

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

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

```

```

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

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

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

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

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

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

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

temporal_mode(point_in_time).
temporal_mode(distribution_over_horizon).

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

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

```



```

%% --- Axis: skills / methodological toolkit (training, NOT intelligence) ---
scope_skill(mars_without_overfitting).
scope_skill(fit_vs_generalization_balance).    % balancing  $R^2$  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 Marktwert × Beleihungsauslauf als Näherungsgröße für den Zwangsversteigerungswert

Die vorangegangenen Abschnitte haben dargelegt, wozu die Wertermittlung dient, was sie hervorbringt und wo ihre doktrinen und regulatorischen Grenzen liegen; dieser Abschnitt erreicht die Spannung, auf die sie zugelaufen sind.

Eine der tieferen praktischen Realitäten hinter Hypothekenkreditvergabe und Sicherheitenanalyse besteht darin, dass ein in der Kreditprüfung verwendetes Marktwertgutachten in seiner tatsächlichen wirtschaftlichen Funktion oft als implizite Näherungsgröße für die erwartete Verwertbarkeit der Sicherheit über einen künftigen Haltezeitraum dient, obwohl das Gutachten formal als Einschätzung des gegenwärtigen Marktwerts zu einem bestimmten Bewertungsstichtag gerahmt ist. (Foster und Van Order 1984; Kau u. a. 1995; Deng, Quigley und Van Order 2000)

Ein Beleihungsauslauf (Loan-to-Value, LTV) von 80% ist nicht willkürlich. Historisch entwickelte er sich teilweise als Risikopuffer, der Kreditgeber schützen sollte vor:

- moderaten Marktrückgängen,
- Zwangsversteigerungskosten,
- Liquidationsabschlägen,
- Haltekosten,
- Rechtskosten,
- aufgelaufenen Zinsen
- sowie der Unsicherheit während der Zeit, die nach einem Ausfall für die Verwertung der Sicherheit benötigt wird. (Foote, Gerardi und Willen 2008; Qi und Yang 2009)

Der Puffer ist auch der Grund, warum der Darlehensbetrag selbst — das Produkt aus Marktwert und zulässigem Beleihungsauslauf — als Arbeitsschätzung des Kreditgebers für den erzielbaren Zwangsversteigerungswert gelesen werden kann: Die Deckelung des Darlehens auf einen Bruchteil des Marktwerts ist im Ergebnis eine Wette darauf, dass der Erlös aus einem Zwangsverkauf nicht unter den Darlehenssaldo fällt. Wie groß der Abschlag (Haircut) jedoch sein *sollte*, ist weit schwerer zu bestimmen, als die runde Zahl von 80% nahelegt. Schätzungen des Zwangsversteigerungsabschlags variieren stark über Märkte, Methoden und Zeiträume; eine

häufig zitierte Studie zu Zwangsverkäufen beziffert ihn auf etwa 27% unter vergleichbaren Preisen aus Transaktionen unter fremden Dritten (Campbell, Giglio und Pathak 2011), während andere Analysen — unter Korrektur für Objektzustand, Lage und Selektion — deutlich andere Größenordnungen berichten. (Pennington-Cross 2006; Clauretie und Daneshvary 2009)

Die konventionelle 80%-Schwelle ist demgegenüber im Wesentlichen ein einziger landesweiter Parameter, angewandt auf enorm heterogene lokale Märkte und aufeinanderfolgende Phasen des Kreditzyklus. Sie kann den Abschlag keines bestimmten Markts nachzeichnen und fungiert daher als sehr grobe Einheitsnäherung statt als präzise Abbildung vom Marktwert auf den Zwangsversteigerungserlös. Wenn überhaupt, ist sie eine knappe: Ein Abschlag von 20% liegt *unter* dem von jener Studie berichteten Abschlag von 27%, sodass das maximale konventionelle Darlehen den Bruttoerlös eines Zwangsverkaufs schon übersteigen kann, bevor die oben aufgezählten Verwertungskosten abgezogen werden.

Wo das Eigenkapital des Kreditnehmers zu gering ist, um den Puffer zu bilden, tritt eine Hypothekenversicherung an dessen Stelle. Dies ist weniger als politisches Stellrad zu verstehen denn als Abwägung in der Risikoallokation: Ein Segment von Kreditnehmern ist vermögensarm, aber einkommensstabil, und das zusätzliche Ausfall- und Verlustrisiko einer Kreditvergabe an sie oberhalb von 80% LTV wird von einem Versicherer prognostiziert und getragen — privat, über eine private Hypothekenversicherung (PMI) bei einem konventionellen Darlehen mit hohem Beleihungsauslauf, oder öffentlich, über die Federal Housing Administration (FHA, die US-Bundesbehörde für Wohnungsbaufinanzierung). Der Markt, in dem diese Abwägung Routine ist, ist allerdings eine Schöpfung der bundesstaatlichen Wohnungspolitik: Den staatsnahen Unternehmen (Government-Sponsored Enterprises), die die konventionelle Kreditvergabe dominieren, ist es per Charta untersagt, Darlehen über 80% LTV ohne eine solche Kreditverbesserung anzukaufen (United States Code 1970). Die Bedingungen, zu denen die Versicherung endet, sind ebenfalls institutionelle Entscheidungen und keine Risikoberechnungen, und sie unterscheiden sich zwischen den Programmen erheblich.¹

Obwohl der Gutachter also offiziell den „gegenwärtigen Marktwert“ schätzt, interpretiert der Kreditgeber diesen Wert wirtschaftlich durch eine zukunftsgerichtete Risikolinse. Im Ergebnis stellt der Kreditgeber eher eine Frage wie:

„Wenn dieser Kreditnehmer irgendwann in absehbarer Zukunft ausfällt, wird die Sicherheit dann voraussichtlich ausreichen, um die ausstehende Schuld und die mit der Zwangsversteigerung verbundenen Verluste zu decken?“

Das ist ihrem Wesen nach eine zukunftsgerichtete Frage.

Dies beruht auf einer bewussten begrifflichen Trennung zwischen:

- der Einschätzung des „gegenwärtigen Marktwerts“ durch den Gutachter und
- der praktischen Verwendung dieser Einschätzung durch den Kreditgeber als Prognosepuffer gegen künftige Wertminderung der Sicherheit.

Historisch und rechtlich versuchten die Aufsichtsbehörden zu verhindern, dass Gutachter zu spekulativen Prognostikern werden, weil:

- spekulative Prognosen zu früheren Kreditkrisen beigetragen hatten,
- Gerichte Gutachten verlangten, die an beobachtbare Evidenz gebunden sind,
- und Bankenaufseher standardisierte, verteidigungsfähige Bewertungspraktiken anstrebten.

¹Die vom Kreditnehmer bezahlte private Hypothekenversicherung konventioneller Darlehen endet nach dem Homeowners Protection Act von 1998 automatisch bei 78% LTV und kann bei 80% beantragt werden. (United States Congress 1998) Die Prämien der Federal Housing Administration sind strenger: Bei Darlehen, die am oder nach dem 3. Juni 2013 mit einem ursprünglichen Beleihungsauslauf über 90% (z. B. 3,5% Anzahlung) vergeben wurden, läuft die Jahresprämie über die gesamte Laufzeit des Darlehens, während sie bei Darlehen mit höchstens 90% nach elf Jahren entfällt. (U.S. Department of Housing and Urban Development 2013) Viele FHA-Kreditnehmer werden die Hypothekenversicherung daher nur los, indem sie in ein konventionelles Darlehen nahe 80% LTV umschulden.

So entwickelte sich die Bewertungsdoktrin hin zu:

- Gegenwartswert,
- aktueller Marktevidenz,
- definierten Bewertungsstichtagen
- und eingegrenzten Annahmen.

Unterdessen verwenden Kreditgeber, Investoren und Sicherheitenanalysten ebendiese Gutachten routinemäßig in probabilistischen Zukunftsrisiko-Rahmenwerken.

In Wirklichkeit ist die Hypothekenkreditprüfung ein integriertes System, das umfasst:

1. aktuelle Marktbewertung,
2. Schätzung künftiger Risiken,
3. Ausfallwahrscheinlichkeit,
4. zeitlichen Verlauf der Zwangsversteigerung,
5. Schwere der Liquidationsverluste,
6. makroökonomische Erwartungen
7. und Portfolio-Stressanalyse.

Der Gutachter trägt formal in erster Linie zu Punkt (1) bei, während der Kreditgeber dieses Ergebnis intern mit den übrigen kombiniert.

In der Praxis verschwimmt die Unterscheidung. Wenn ein Kreditgeber ein Darlehen nur vergeben würde, sofern er glaubt, dass die Sicherheit ihren Wert voraussichtlich hinreichend in die Zukunft bewahrt, dann fungiert das Gutachten mittelbar als Teil eines Modells zum Schutz künftigen Werts — auch wenn die Bewertungsstandards es formal vermeiden, es so zu beschreiben.

Diese Spannung ist noch ausgeprägter geworden mit:

- automatisierten Bewertungsmodellen (AVMs),
- institutioneller Portfolioanalytik,
- Verbriefungsmodellen,
- Bewertungssystemen mit maschinellem Lernen,
- der Basel-Kapitalmodellierung
- und groß angelegten Hypotheken-Stresstests.

Die moderne Sicherheitenanalytik behandelt Immobilienwerte zunehmend probabilistisch über die Zeit statt als statische Punktschätzungen. (Basel Committee on Banking Supervision 2017; Steurer, Hill und Pfeifer 2021; Board of Governors of the Federal Reserve System 2021)

Man könnte argumentieren, dass die traditionelle Regulierung der Wertermittlung einen historischen Kompromiss widerspiegelt:

- Gutachter liefern disziplinierte, evidenzbasierte Anker des Gegenwartswerts,
- während Finanzinstitute die Verantwortung für die Interpretation künftiger Risiken übernehmen.

Ob diese Trennung in modernen Finanzsystemen noch vollständig kohärent ist, ist eine berechnete und zunehmend diskutierte Frage.

Daraus folgt eine praktische Konsequenz für die Bewertungspraxis selbst. Die meisten Gutachter sind sich dieser Rolle in der Sicherheitenbewertung nicht bewusst — die Verwendung des Marktwerts durch den Kreditgeber als zukunftsgerichteter Puffer ist aus dem Bewertungsauftrag heraus selten sichtbar — und behandeln ihre Marktwertschlussfolgerungen deshalb nicht mit der Vorsicht gegenüber künftigen Werttrends, die diese Rolle verlangt. Zu erkennen, wie die Gegenwartswert-Einschätzung tatsächlich konsumiert wird, ist ein erster Schritt, um diese Kluft zu schließen, und ein Feld, das diese Darstellung für weitere Arbeiten öffnen soll.

```
%% 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 Schlussfolgerung

Die fünf in den vorstehenden Wissensbasen kodierte Passagen beschreiben einen einzigen Berufsstand aus fünf komplementären Blickwinkeln. Für sich genommen liest sich jede Passage als eigenständige Darstellung: die gesellschaftlichen und gemeinwohlbezogenen Belange, die sie tragen, die wirtschaftlichen Funktionen, die Gutachten unterstützen, ihre operative Rolle in der Sicherheitenrisikoanalyse, die regulatorische Grenze, die sie von benachbarter quantitativer Arbeit trennt, und die Spannung zwischen ihrer formalen Rahmung und ihrer praktischen Verwendung. Zusammen genommen, und mit der in Prolog explizit gemachten Struktur, beschreiben sie etwas Interessanteres — einen Berufsstand, dessen offizielle Doktrin und wirtschaftliche Funktion bewusst auseinandergehalten wurden und dessen innere Kohärenz durch die Analysewerkzeuge, die inzwischen auf seine Produkte angewandt werden, zunehmend auf die Probe gestellt wird.

Die ersten drei Wissensbasen teilen ein gemeinsames Schema: `effect(Quality, Horizon, Domain, Effect)`. Jede rahmt die Bewertungsqualität als kausalen Input, dessen Folgen sich über kurz-, mittel- und langfristige Horizonte fortpflanzen, und unterscheidet sich nur in der Domäne, durch die diese Folgen gelesen werden. Die ökonomische Linse liest sie durch Kreditmärkte, Investitionsallokation, Besteuerung und Kapitalbildung; die soziale Linse durch Gerechtigkeit, öffentliches Vertrauen und Wirkungen auf das Gemeinwesen; die Sicherheiten-Linse durch Kreditprüfung, Portfoliomanagement und Stabilität des Finanzsystems. Das gemeinsame Schema ist kein Zufall. Es spiegelt eine inhaltliche These wider, die sich durch alle drei Passagen zieht: Die Qualität der Bewertungsarbeit hat Folgen, diese Folgen skalieren mit der Zeit, und dieselben zugrunde liegenden Fehlermodi äußern sich unterschiedlich, je nachdem, wessen Interessen im Vordergrund stehen. Gemeinsame Atome auf der Argumentseite der Prädikate — insbesondere die Vokabulare für Qualität und Horizont — machen die drei Vokabulare verknüpfbar, sodass ein einzelnes Versagen wie systematische Überbewertung gleichzeitig durch ökonomische Fehlbepreisung, soziale Ungerechtigkeit und Fehlmessung von Sicherheiten verfolgt werden kann.

Die vierte Wissensbasis wechselt das Register. Sie beschreibt nicht die Folgen der Bewertungsqualität; sie beschreibt, was Wertermittlung methodisch und rechtlich *ist*. Ihre Prädikate sind taxonomisch statt kausal, und ihre Inferenzmaschinerie beantwortet Fragen der Zuständigkeit statt der Wirkung. Dieser Wechsel ist notwendig, weil die folgenorientierte Rahmung der ersten drei Passagen eine Definition der zu bewertenden Tätigkeit voraussetzt, und diese Definition ist selbst umstritten. Die Scope-Prädikate halten fest, was Gutachter tun — evidenzbasierte Gegenwartswert-Einschätzungen mit eng begrenzten zukunftsgerichteten Komponenten —, während die Regulierungs-Prädikate die Bedingungen festhalten, unter denen diese Tätigkeit in die lizenzierte Praxis fällt. Die Vorwärtsverkettungsregeln (`requires_license/1`, `exempt_from_appraisal_license/1`, `regulated_appraisal/1`) liefern einen Mechanismus, um zu fragen, ob eine gegebene Bewertungsarbeit innerhalb oder außerhalb des regulierten Perimeters liegt — eine Frage, von der die ersten drei Wissensbasen annehmen, dass sie bereits beantwortet ist.

Die fünfte Wissensbasis vertritt eine These. Wo die anderen beschreiben, klassifizieren oder Folgen nachzeichnen, argumentiert die Proxy-Wissensbasis, dass die sorgfältige doktrinaire Trennung zwischen Gegenwartswert-Bewertung und Zukunftswert-Prognose — die Trennung, die die Regulierungs-Wissensbasis dokumentiert — in Spannung zur tatsächlichen wirtschaftlichen Funktion von Gutachten in der Hypothekenkreditvergabe steht. Die 80%-Beleihungsauslauf-Konvention, die sieben spezifischen Risiken, gegen die sie abpuffert, die wörtlich festgehaltene zugrunde liegende Frage des Kreditgebers („Wenn dieser Kreditnehmer irgendwann in absehbarer Zukunft ausfällt, wird die Sicherheit dann voraussichtlich ausreichen...“) und das integrierte Kreditprüfungssystem aus sieben Komponenten stützen alle dieselbe Behauptung: Das Gutachten ist formal eine Gegenwartswert-Einschätzung und wirtschaftlich ein Zukunftswert-Puffer. Der historische Kompromiss, der diese Anordnung hervorbrachte — Gutachter als disziplinierte

Gegenwartswert-Anker, Kreditgeber als Ort der Interpretation künftiger Risiken —, war eine vertretbare Antwort auf die Bedingungen, unter denen die Gutachterlizenzierung entstand; doch die in der Passage aufgezählten modernen Verstärker (AVMs, Bewertungssysteme mit maschinellem Lernen, Basel-Kapitalmodellierung, Verbriefungsanalytik, groß angelegte Stresstests) behandeln Immobilienwerte probabilistisch über die Zeit. Diese Werkzeuge ziehen doktrinär getrennte Schichten wieder in eine einzige rechnerische Pipeline zusammen, und die begriffliche Trennung zwischen den Schichten wird dadurch schwerer aufrechtzuerhalten.

Ob diese Versöhnung erreichbar ist, und zu welchen Bedingungen, ist die Frage, mit der die Proxy-Passage schließt. Die Wissensbasen beantworten sie nicht. Sie liefern jedoch eine explizite, abfragbare Darstellung der für ihre Beantwortung relevanten Erwägungen: die Folgen der Bewertungsqualität über Horizonte und Domänen hinweg, die Zuständigkeitsgrenze, die bestimmt, welche Folgen in die regulierte Praxis fallen, und die strukturellen Gründe, aus denen diese Grenze nun unter Druck steht. Diese Darstellung ist das Fundament, auf dem spezifischere Argumente — über Methodik, über die Reform von Standards, über das Verhältnis zwischen Bewertungspraxis und dem breiteren System der quantitativen Bewertung — aufgebaut werden können.

Anmerkungen zur Wissensrepräsentation

Drei strukturelle Muster verdienen es, ausdrücklich festgehalten zu werden, da sie aus dem Kodierungsprozess hervorgingen und nicht aus einer einzelnen Passage.

Erstens ist die Familie der Wissensbasen nicht flach. Drei der fünf teilen ein Schema; eine liefert die Zuständigkeitsbedingungen, die bestimmen, wann dieses Schema anwendbar ist; eine vertritt eine These über die Kluft zwischen formaler Rahmung und wirtschaftlicher Funktion. Eine Brückenschicht, die alle fünf kombiniert, behandelt sie nicht als fünf Spalten derselben Matrix. Sie behandelt die ersten drei als Beschreibungen von Folgen, die vierte als Eingrenzung dessen, wo diese Beschreibungen wirksam sind, und die fünfte als Identifikation der Stelle, an der die Eingrenzung selbst unter Druck steht. Diese asymmetrische Struktur ist ein Merkmal, kein Mangel: Sie spiegelt die tatsächliche Organisation der Berufspraxis wider, in der deskriptive Aussagen, Zuständigkeitsaussagen und kritische Aussagen genuin unterschiedliche epistemische Rollen einnehmen.

*Zweitens leistet die Präfixdisziplin, die die Module unterscheidet (*economic_*, *social_*, *collateral_*, *scope_/regulation_*, *proxy_/underwriting_*), nur dort echte Arbeit, wo Kollisionsgefahr besteht.* Prädikate mit direkten Gegenständen in anderen Modulen verdienen ihr Präfix; Prädikate, die einer einzigen begrifflichen Domäne eigen sind, nicht. Das gemeinsame Wertevokabular auf der Argumentseite — die Atome für Qualität, Horizont und Domäne — ist genau das, was den Modulen die Kommunikation erlaubt, und eine Präfigierung dieser Werte würde das vereiteln. Das entstandene Muster ist allgemein: Präfix auf den Prädikaten, gemeinsame Atome auf den Argumenten, mit selektiver statt einheitlicher Präfigierung, wo das Prädikat nur einer Linse eigen ist.

Drittens ist die in der Proxy-Wissensbasis identifizierte zentrale Spannung eng verwandt mit der in der Scope-/Regulierungs-Wissensbasis identifizierten regulatorischen Asymmetrie, aber von ihr verschieden. Beide sind durch parallele zweistellige Prädikate festgehalten — *regulatory_asymmetry/2* und *proxy_doctrine_practice_gap/2* — mit bewusst unterschiedlichen Namen, weil sie unterschiedliche Achsen derselben zugrunde liegenden Anordnung beschreiben. Die regulatorische Asymmetrie betrifft, wer reguliert wird und wie stark; die Kluft zwischen Doktrin und Praxis betrifft, was über das Produkt des Gutachters formal gesagt wird gegenüber dem, was wirtschaftlich damit getan wird. Die beiden Spannungen verstärken einander. Die Regulierung drängt Gutachter zur Gegenwartswert-Doktrin; die wirtschaftliche Funktion zieht ihr Produkt zur Zukunftswert-Verwendung. Das Gesamtbild ist das eines Berufsstands, dessen offizielle Selbstbeschreibung immer schwerer mit den analytischen Pipelines zu vereinbaren ist, die seine Produkte speisen.

Literatur

- Agarwal, Sumit, Itzhak Ben-David und Vincent Yao (2015). „Collateral Valuation and Borrower Financial Constraints: Evidence from the Residential Real Estate Market“. In: *Management Science* 61.9, S. 2220–2240. DOI: [10.1287/mnsc.2014.2002](https://doi.org/10.1287/mnsc.2014.2002).
- Alexandrov, Alexei, Laurie Goodman und Michael Neal (Jan. 2023). *Reengineering the Appraisal Process: Better Leveraging Both Automated Valuation Models and Manual Appraisals*. Techn. Ber. 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 (März 2021). *Dodd-Frank Act Stress Test 2021: Supervisory Stress Test Methodology*. Techn. Ber. 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 und Parag Pathak (2011). „Forced Sales and House Prices“. In: *American Economic Review* 101.5, S. 2108–2131. DOI: [10.1257/aer.101.5.2108](https://doi.org/10.1257/aer.101.5.2108).
- Clauretie, Terrence M. und Nasser Daneshvary (2009). „Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time“. In: *Real Estate Economics* 37.1, S. 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 und Patrick M. McAllister (2010). „Means, Motive and Opportunity? Disentangling Client Influence on Performance Measurement Appraisals“. In: *Journal of Property Research* 27.2, S. 181–201. DOI: [10.1080/09599916.2010.499014](https://doi.org/10.1080/09599916.2010.499014).
- Deng, Yongheng, John M. Quigley und Robert Van Order (2000). „Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options“. In: *Econometrica* 68.2, S. 275–307. DOI: [10.1111/1468-0262.00110](https://doi.org/10.1111/1468-0262.00110).
- Diaz Julian, III und Marvin L. Wolverson (1998). „A Longitudinal Examination of the Appraisal Smoothing Hypothesis“. In: *Real Estate Economics* 26.2, S. 349–358. DOI: [10.1111/1540-6229.00749](https://doi.org/10.1111/1540-6229.00749).
- Ebrahim, M. Shahid und Sikandar Hussain (2010). „Financial Development and Asset Valuation: The Special Case of Real Estate“. In: *Journal of Banking & Finance* 34.1, S. 150–162. DOI: [10.1016/j.jbankfin.2009.07.011](https://doi.org/10.1016/j.jbankfin.2009.07.011).
- Eriksen, Michael D. u. a. (2019). „The Influence of Contract Prices and Relationships on Appraisal Bias“. In: *Journal of Urban Economics* 111, S. 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 und Paul S. Willen (2008). „Negative Equity and Foreclosure: Theory and Evidence“. In: *Journal of Urban Economics* 64.2, S. 234–245. DOI: [10.1016/j.jue.2008.07.006](https://doi.org/10.1016/j.jue.2008.07.006).
- Foster, Chester und Robert Van Order (1984). „An Option-Based Model of Mortgage Default“. In: *Housing Finance Review* 3.4, S. 351–372.
- Friedman, Jerome H. (1991). „Multivariate Adaptive Regression Splines“. In: *The Annals of Statistics* 19.1, S. 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, S. 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, S. 327–345. DOI: [10.1007/BF00161933](https://doi.org/10.1007/BF00161933).
- Ghent, Andra C., Walter N. Torous und Rossen I. Valkanov (2019). „Commercial Real Estate as an Asset Class“. In: *Annual Review of Financial Economics* 11, S. 153–171. DOI: [10.1146/annurev-financial-110118-123121](https://doi.org/10.1146/annurev-financial-110118-123121).

- Griffin, John M. und Gonzalo Maturana (2016). „Who Facilitated Misreporting in Securitized Loans?“ In: *The Review of Financial Studies* 29.2. Evidence of appraisal inflation and its consequences, S. 384–419. DOI: [10.1093/rfs/hhv130](https://doi.org/10.1093/rfs/hhv130).
- Hendershott, Patric H. und Edward J. Kane (Juli 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, S. 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. Aufl. New York: Pearson. ISBN: 978-0-13-447208-9.
- International Association of Assessing Officers (2017). *Standard on Mass Appraisal of Real Property*. International Association of Assessing Officers, Kansas City, MO. URL: https://www.iaao.org/wp-content/uploads/Standard_on_Mass_Appraisal.pdf.
- International Valuation Standards Council (2025). *International Valuation Standards (IVS)*. International Valuation Standards Council, London. Published 31 January 2024; effective 31 January 2025. URL: <https://www.ivsc.org/>.
- K’Akumu, Owiti A. und 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. u. a. (1995). „The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment“. In: *The Journal of Real Estate Finance and Economics* 11.1, S. 5–36. DOI: [10.1007/BF01097934](https://doi.org/10.1007/BF01097934).
- LaCour-Little, Michael und Stephen Malpezzi (2003). „Appraisal Quality and Residential Mortgage Default: Evidence from Alaska“. In: *The Journal of Real Estate Finance and Economics* 27.2, S. 211–233. DOI: [10.1023/A:1024728420837](https://doi.org/10.1023/A:1024728420837).
- Lee, Ming-Hsiang, Chien-Wen Peng und Hui-Fen Liao (2020). „An Analysis of Objectivity in the Real Estate Appraisal Process“. In: *International Real Estate Review* 23.4, S. 483–504.
- Mayer, Yanling G. und Frank E. Nothaft (2022). „Appraisal Overvaluation: Evidence of Price Adjustment Bias in Sales Comparisons“. In: *Real Estate Economics* 50.3, S. 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, S. 193–214. DOI: [10.1080/10835547.2006.12091177](https://doi.org/10.1080/10835547.2006.12091177).
- Qi, Min und Xiaolong Yang (2009). „Loss Given Default of High Loan-to-Value Residential Mortgages“. In: *Journal of Banking & Finance* 33.5, S. 788–799. DOI: [10.1016/j.jbankfin.2008.09.010](https://doi.org/10.1016/j.jbankfin.2008.09.010).
- Quan, Daniel C. und 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, S. 127–146. DOI: [10.1007/BF00173120](https://doi.org/10.1007/BF00173120).
- Steurer, Miriam, Robert J. Hill und Norbert Pfeifer (2021). „Metrics for Evaluating the Performance of Machine Learning Based Automated Valuation Models“. In: *Journal of Property Research* 38.2, S. 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): Rahmenwerk & Protokoll

Deutsche Übersetzung der aktualisierten Fassung des ursprünglich am 5. März 2025 auf Zenodo veröffentlichten Beitrags. Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

William Bert Craytor, SRA

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

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

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

Dieser Open-Access-Artikel wird unter den Bedingungen der [Creative-Commons-Namensnennung-4.0-Lizenz](#) verbreitet.

Zusammenfassung. Eine wesentliche Einschränkung konventioneller Wertermittlungsmethoden liegt in ihrer Abhängigkeit von veralteten manuellen Techniken wie der Matched-Pair-Analyse, die die Wertbeiträge einzelner Objektmerkmale übergehen, als existierten sie nicht: Das traditionelle Sales Grid (Vergleichsraster) enthält keine Spalte für Wertbeiträge, sondern nur für Anpassungen. Tatsächlich liefern Wertbeiträge die notwendige Grundlage für mathematische Restriktionen, die Über- und Unterbewertungen durch traditionelle Gutachter verhindern.

Der Residual Constraint Approach (RCA, dt. etwa: Residualrestriktionsansatz) erweitert den traditionellen Sales Comparison Approach (SCA, vergleichbar dem deutschen Vergleichswertverfahren) durch einen mehrphasigen Bewertungsprozess, der Multivariate Adaptive Regression Splines (MARS) mit strengen mathematischen Restriktionen für die Wertbeiträge von Objektmerkmalen verbindet — insbesondere für nicht gemessene Merkmale: latente Variablen wie Ästhetik, Design, Zustand und Qualität, deren Werte beim traditionellen Gutachter typischerweise subjektivem Urteil und Verzerrung unterliegen.

Die mathematischen Restriktionen des RCA beruhen auf der fachkundigen Anwendung der MARS-Regression zur Schätzung der Wertbeiträge messbarer Objektmerkmale zum Verkaufspreis. Diese Schätzung erklärt jedoch typischerweise nur etwa 80% des tatsächlichen Verkaufspreises; ausgenommen bleiben subjektiv beurteilte Merkmale wie Design, funktionale Nutzbarkeit, Zustand und Qualität. In der San Francisco Bay Area beträgt das Residuum — das diese subjektiven Komponenten repräsentiert — im Mittel rund 20% des gesamten Objektwerts. Der Wert von 80/20 spiegelt die Erfahrung des Autors in der San Francisco Bay Area unter typischen Datenbedingungen wider und ist nicht als marktinvariante Konstante zu verstehen. In Märkten mit niedrigerem erreichbarem R^2 besteht die Antwort des Rahmenwerks darin, nicht gemessene Werttreiber zu identifizieren und zu messen — die Entfernung zum Ozean in Küstenorten, die Höhenlage in Hanglagen und ähnliche, aus externen Quellen erhobene Variablen, sofern die MLS-Daten sie nicht liefern — statt die Einschränkung hinzunehmen und breitere Fehlerbänder auszuweisen. Während MARS-Residuen in anderen Kontexten häufig als Schätzfehler behandelt werden, dienen sie in der Immobilienbewertung als indirektes Maß für den Wert latenter Variablen. Auch wenn dies zunächst unpraktikabel erscheinen mag, zeigt ein Beweis, dass sich das Residuum sinnvoll in beschreibende Komponenten zerlegen lässt, sofern deren Beiträge in der Summe das Residuum ergeben — ohne das endgültige Bewertungsergebnis zu beeinflussen.

Diese Methodik ist das Ergebnis von zwei Jahrzehnten empirischer Anwendung durch den Autor und wurde durch iterative MARS-Implementierungen in R- und Python-Umgebungen entwickelt.

Schlüsselwörter: Immobilienbewertung, Wertermittlung von Wohnimmobilien, Residual Constraint Approach, MARS, Multivariate Adaptive Regression Splines, Sales Comparison Approach, Residualanalyse, latente Variablen, automatisierte Bewertung

Inhaltsverzeichnis

1	Einleitung	4
2	Hintergrund	6
3	Valuation Engineer	9
4	Annahmen	10
5	RCA-Workflow	13
5.1	Arten der MARS-Analyse	13
5.2	Verarbeitungs-Workflow	13
5.3	Daten-Workflow	15
5.3.1	Datenanforderungen für MARS	15
5.3.2	Protokolle: Listings abgeschlossener Verkäufe	16
5.4	Projektordnerstruktur	18
5.4.1	Protokolle: Mehrfache MARS-(earth-)Läufe	19
5.5	Die zentrale Excel-Arbeitsmappe	21
5.6	Datensicherungen	21
6	Notation der Arbeitsmappe	22
6.1	Deskriptive Dimensionen	22
6.1.1	Die Projektdimension („a“)	23
6.1.2	Dimension der Laufversionen („r“)	24
6.1.3	Workflow-Stufen („w“)	24
6.1.4	Blätter der Excel-Arbeitsmappe („s“)	25
6.1.5	Objekt-Listings („p“) - Zeilen der MLSData	25
6.1.6	MLS-Variablen („v“) Spalten der MLSData	25
6.1.7	Abschließendes Notationsbeispiel	26
6.1.8	Begründung	26
6.1.9	Matrizen?	27
6.1.10	Erste Vereinfachung	27
6.1.11	Zweite Vereinfachung	27
7	Das RCA Sales Grid	29

8 Variablen	33
8.1 Ontologie der Variablenherkunft	33
8.2 Ontologie der Variablenanwendung	34
8.3 Terme zweiten und dritten Grades	36
8.3.1 Anpassungsvariablen	36
8.3.2 Aggregationsvariablen	36
8.3.3 Spezielle Variablen & Notation	37
8.3.4 Aufschlüsselung der Residuen	37
9 Zentrale RCA-Beweise	39
10 MARS-Modell	41
10.1 „ g “-Funktionen für Modellvariablen und Interaktionen	41
10.2 Indizierung der g -Funktionen	41
10.3 Dimensionen der grafischen Darstellung	42
10.4 Die endgültige Modellgleichung	43
10.5 Anzahl der zur grafischen Darstellung eines Modells benötigten Diagramme	43
11 Datencharakteristika	46
11.1 Zweck des Residuums	46
11.2 Zustands- und Qualitätsbewertungen als approximative Regressoren	47
11.3 Datenfehler	47
11.4 Was ist Genauigkeit?	49
11.5 Das Bewertungsobjekt als Anomalie	51
11.6 Die Bedeutung des CVR2 im Umgang mit unerwarteten Daten	51
12 Residuen	52
12.1 Ranking & Scoring	52
12.2 CQA-Residuum-Kurve	55
12.3 Eigenschaften der CQA-Kurve	55
12.3.1 Analyse der Komponenten der CQA-Kurve	55
12.3.2 Marktdynamik und soziale Implikationen	56
12.4 Residuum des Bewertungsobjekts	56
12.5 RCA-Wertschlussfolgerung	58
12.5.1 Allgemeine Überlegungen	58
12.5.2 Überprüfung, Auditierung, Beschwerden	58
13 Schlussfolgerung	60

1 Einleitung

Diese Arbeit stellt die logische und mathematische Grundlage für einen modernen Ansatz zur genauen Schätzung des Marktwerts von Wohnimmobilien und anderen Immobilien dar. Die Methode, die im Laufe von 20 Jahren Erfahrung in der Anwendung von Multivariate Adaptive Regression Splines (MARS)-Software auf Wohnhäuser in Nordkalifornien — insbesondere in Gemeinden der San Francisco Bay Area — entwickelt wurde, bietet einen erheblichen Fortschritt in der Präzision der Bewertung. MARS verbessert die Qualität der Bewertung, indem es nichtlineare Beziehungen zwischen Objektmerkmalen und Wert geschickt erfasst und zugleich die einflussreichsten Variablen effizient identifiziert und priorisiert. Diese Fähigkeit erlaubt es Gutachtern, Wertdynamiken in neuen Marktgebieten weitaus schneller aufzudecken, als es traditionelle Methoden gestatten. MARS bildet das Rückgrat des vorgeschlagenen Residual Constraint Approach (RCA) (dt. etwa: Residualrestriktionsansatz). Die RCA-Methode verdankt ihren Namen ihrem rigorosen Rahmenwerk:

- Sie erlegt den Wertbeiträgen sowohl gemessener als auch nicht gemessener Objektmerkmale strikte Restriktionen auf:
 1. Alle aus MARS abgeleiteten Wertbeiträge, einschließlich des Residuums, müssen sich zum Nettoverkaufspreis jedes Vergleichsobjekts summieren.
 2. Das Residuum kann in Teilkomponenten mit zugewiesenen Werten zerlegt werden, sofern deren Summe dem Residuum des Vergleichsobjekts entspricht.
- RCA minimiert Verzerrungen bei der Schätzung subjektiver Merkmale — etwa ästhetischer Anmutung, Zustand und Qualität —, indem ihr Gesamtwert an das Residuum gebunden wird, das genau quantifiziert wird, wenn ein kompetenter Valuation Engineer das MARS-Modell konstruiert.

Während RCA in erster Linie für Wohnimmobilien konzipiert ist, bei denen subjektive Merkmale den Marktwert erheblich beeinflussen, birgt der Ansatz Potenzial für breitere Anwendungen in der Vermögensbewertung. Genaue Gutachten sind angesichts des auf \$45 Billionen geschätzten Werts des US-amerikanischen Wohnimmobilienmarkts von entscheidender Bedeutung. Fannie Mae stellt fest, dass typische Gutachten um etwa 5% vom tatsächlichen Verkaufspreis abweichen; diese Zahl verschleiert jedoch Verzerrungen, da Gutachter häufig unter Druck stehen, sich an den Verkaufspreisen auszurichten. Da zudem 60%–70% der Gutachten Refinanzierungen dienen — bei denen vergleichbare Verkaufspreise fehlen —, wird der Bedarf an Präzision, wie sie RCA bietet, offenkundig. Die Analyse der Federal Reserve aus dem Jahr 2012 schätzte, dass 15%–20% der vor der Krise von 2008 erstellten Gutachten um mehr als $\pm 10\%$ von den Verkaufspreisen abwichen — eine Spanne, die für Refinanzierungsgutachten heute je nach Objekttyp und Markt vermutlich weiterhin relevant ist. Warum ist Genauigkeit wichtig? Bei Transaktionen mit niedrigem Beleihungsauslauf (Loan-to-Value, LTV) — z.B. wenn ein Käufer mit starker Bonität \$500,000 auf ein Haus von \$1 Million anzahlt — verzichten Kreditgeber mitunter auf Gutachten, da sie ein minimales Zwangsvollstreckungsrisiko annehmen. Käufer, die auf ein Gutachten verzichten, riskieren jedoch, zu viel zu zahlen und Mängel erst nach dem Kauf zu entdecken. Dies unterstreicht den Wert robuster Bewertungsmethoden wie RCA. Die Quantifizierung der Genauigkeit von Gutachten bleibt eine Herausforderung. Forschern fehlen direkte Messgrößen für Abweichungen vom „Marktwert“; sie stützen sich stattdessen auf indirekte Näherungsgrößen. Traditionelle Wertermittlungsmethoden, die auf subjektivem Urteil und variabler Kompetenz der Gutachter beruhen, erschweren die Analyse. Zu den Strategien zählen:

1. Tzioumis (2016) untersuchte 4,782 Gutachter anhand von 259,436 Gutachten in 11 Bundesstaaten und stellte fest, dass ein Drittel innerhalb von 1% des Vertragspreises lag, die Hälfte zwischen 1%–3% darüber und eine Minderheit im Durchschnitt 5% darüber — was auf eine Aufwärtsverzerrung hindeutet. Die Studie identifizierte zwei Gutachterttypen: „zum Marktwert“ (nahe am Vertragspreis) und „über dem Marktwert“ (durchweg höher).
2. Krivorotov und LaCour-Little (2020) stellte fest, dass 90% der Kaufgutachten den Vertragspreis erreichten oder überschritten, und hob die „Ankerverzerrung“ hervor — die Tendenz, sich an einem gewünschten Wert auszurichten —, insbesondere im Refinanzierungskontext.

Diese Studien offenbaren Verzerrungen in der Gutachterpraxis, verstärkt durch die Grenzen traditioneller Werkzeuge. Wären Gutachter einheitlich kompetent und unverzerrt, würden ihre Bewertungen konvergieren; anhaltende Überbewertung durch bestimmte Gruppen deutet auf systemische Probleme hin.

Die oben zitierten Studien zur Verzerrung und das systemische Muster, auf das sie hinweisen, spiegeln ein tieferliegendes Merkmal der traditionellen Praxis des Verkaufspreisvergleichs wider, dem RCA begegnen soll. Die traditionelle Methodik operiert auf zwei Kanälen mit sehr unterschiedlichem epistemischem Status. Auf dem Kanal der messbaren Merkmale — GLA, Zimmeranzahl, Grundstücksgröße, Alter, Lage — wird vom Gutachter erwartet, dass er Anpassungswerte objektiv ableitet, durch Matched-Pair-Analyse oder lokale lineare Regression auf Vergleichsverkäufen. Kompetente Gutachter leisten diese Arbeit, und die resultierenden Anpassungen sind gegenüber den Daten zumindest annähernd vertretbar. Doch selbst hier verzeichnet das traditionelle Sales Grid (Vergleichsraster) nur Anpassungen; es besitzt nicht einmal eine Spalte für Wertbeiträge. Ebenso wenig werden traditionelle Anpassungen als Differenzen zwischen Wertbeiträgen berechnet — sie werden direkt geschätzt, aus Matched Pairs oder ähnlichen Verfahren, ohne dass die zugrunde liegenden Beiträge jemals berechnet würden. In RCA hingegen ist jede Anpassung die Differenz zwischen den Wertbeiträgen des Bewertungsobjekts und des Vergleichsobjekts für das jeweilige Merkmal — die Struktur, deren Konsequenzen Beweis II präzisiert. Auf dem Kanal der subjektiven Merkmale — Zustand, Bauqualität, Design, funktionale Nutzbarkeit, Ästhetik — reichen die verfügbaren Stützungsverfahren von dünn bis fadenscheinig. Anpassungen für subjektive Merkmale können durch Matched-Pair-Analysen von Objekten außerhalb des unmittelbaren Vergleichsobjektbestands, durch Berufung auf anerkannte Praxis der Berufskollegen oder durch Verweis auf die dokumentierte Markterfahrung des Gutachters gestützt werden. Alle drei sind formal mit den Berufsstandards vereinbar. Keines dieser Verfahren erreicht die statistische Strenge, die der messbare Kanal routinemäßig erzielt.

Das Ergebnis ist eine prozedurale Asymmetrie zwischen dem, was die Standards verlangen, und dem, was die Methodik leisten kann. Die Standards verlangen eine Stützung für jede Anpassung, ob messbar oder subjektiv. Die Methodik liefert robuste Stützung auf dem einen Kanal und ein Spektrum schwächerer Verfahren auf dem anderen, das von einem anekdotischen Matched Pair mit Stichprobenumfang eins bis zur Berufung darauf reicht, was andere Gutachter üblicherweise ansetzen. Der ehrliche, kompetente Gutachter, der sorgfältige Matched-Pair-Arbeit an einer Aussichtsanzpassung leistet, liefert eine bedeutend bessere Stützung als einer, der sich ohne Grundlage auf die Praxis der Kollegen beruft; ein Prüfer, der die beiden Gutachten auf dem Standardformular vergleicht, kann sie jedoch kaum unterscheiden, weil das Formular den Anpassungswert festhält, nicht die Strenge seiner Herleitung. Die Methodik toleriert die Asymmetrie, weil den praktizierenden Gutachtern kein stärkeres Verfahren für subjektive Merkmale allgemein zur Verfügung stand.

RCA verbessert beide Kanäle, jedoch auf kategorial unterschiedliche Weise. Auf dem messbaren Kanal erfasst die MARS-Regression nichtlineare Beziehungen, Variableninteraktionen und Basisfunktionsstrukturen, die Matched-Pair-Analyse und gewöhnliche lineare Regression nicht darstellen können. An die vollständige Menge der qualifizierten Verkäufe im Marktgebiet

angepasst — typischerweise mehrere Hundert — statt an die drei bis neun Vergleichsobjekte des traditionellen Sales Grid, sind die aus MARS abgeleiteten Wertbeiträge sowohl objektiv als auch wesentlich genauer als die Methodik, die sie ersetzen. Auf dem subjektiven Kanal leistet RCA etwas, das sich der Art nach von allen traditionellen Stützungsverfahren unterscheidet. Es wird gar nicht erst versucht, Anpassungswerte für einzelne subjektive Merkmale zu schätzen. Die Beiträge aller subjektiven Merkmale — latenter Variablen — erscheinen kollektiv als das Residuum: die Differenz zwischen dem tatsächlichen Nettoverkaufspreis jedes Vergleichsobjekts und seinem von MARS vorhergesagten Preis, berechenbar aus den Daten ohne Zutun des Gutachters. Die Zerlegung dieses Residuums in benannte Komponenten — Design, funktionale Nutzbarkeit, Aussicht, Zustand, Qualität und so weiter — ist der Restriktion unterworfen, sich zum Gesamtresiduum zu summieren, und nach Beweis II (Abschnitt 9) liefert jede Zerlegung, die diese Restriktion erfüllt, dieselbe Wertschlussfolgerung. Das Urteil, das im Arbeitsablauf verbleibt, ist die Platzierung des Bewertungsobjekts in der nach Rängen geordneten Residualverteilung der Vergleichsobjekte, vorgenommen anhand eines überprüfbaren Rankings und im Gutachten ausdrücklich dokumentiert. Ein schwaches RCA-Gutachten bleibt möglich — ein unerfahrener VE kann ein überangepasstes MARS-Modell erzeugen, das Bewertungsobjekt unplausibel platzieren oder Anomalien in den Daten übersehen —, doch es ist auf eine Weise schwach, die ein Prüfer durch Inspektion des Modells und des Rankings erkennen kann, und nicht auf eine Weise, die eine Prüfung anhand des Standardformulars nicht aufdecken könnte.

Über die Refinanzierung hinaus schulden Gutachter den Eigentümern genaue Bewertungen, insbesondere bei hochpreisigen Käufen, bei denen Fehler Zehn- oder Hunderttausende kosten können. Dennoch scheuen viele davor zurück, \$600–\$1,000 für Gutachten auszugeben, weil sie an deren Mehrwert zweifeln. Fortgeschrittene Methoden wie RCA sind zunächst ressourcenintensiv, versprechen jedoch größere Präzision. Mit Verfeinerung und wachsender Erfahrung dürften ihre Effizienz — und ihre Verbreitung — zunehmen.

2 Hintergrund

Der Bildungshintergrund des Autors liegt in den Bereichen Mathematik, Informatik und Statistik. Von Beruf ist er Softwareingenieur mit Nebenerfahrung in Wertermittlung und Rechnungswesen. Seit 2002 erstellt er mit Unterbrechungen Gutachten und wendet seit etwa 2004 MARS zur Bestimmung von Objektanpassungen und Preistrends im Wohnimmobilienbereich an.

Das Erstellen von Gutachten mit MARS bedeutet „tiefgreifende Erfahrung“ in der Wertermittlung. MARS erzeugt Modelle, die detaillierter als jede andere statistische Methode sowohl die Variablen, die zum Marktwert beitragen, als auch die Auswirkungen von Änderungen dieser Werte auf den Verkaufspreis abbilden. Somit untersucht der Gutachter bei der Wertermittlung mit MARS fortwährend kausale Zusammenhänge zwischen Variablenwerten und Verkaufspreisen, wie sie von MARS-Modellen erzeugt werden.

Im Interesse der Transparenz sei angemerkt, dass die RCA-Methodik (Residual Constraint Approach, dt. etwa: Residualrestriktionsansatz) am besten auf Wohngebiete mit folgenden Eigenschaften angewendet wird:

1. Komplexe Wohnviertel mit Objekten, die sich deutlich voneinander unterscheiden
 - Ältere Häuser
 - Unterschiedliche Qualität der Modernisierungen
 - Viele verschiedene Architekturen und Bauweisen
 - Variation von Objektattributen wie GLA, Grundstücksgröße, Höhenlage und Aussicht.

2. Komplexe Marktbedingungen

- Sich schnell ändernde Marktbedingungen
 - Viele verschiedene Käufertypen
3. Gute Datenquellen: Das RCA-Protokoll basiert auf Data Mining und benötigt reichlich qualitativ hochwertige Daten. Es kann zwar auch für Gebiete mit spärlichen Daten verwendet werden, jedoch lohnt sich dies angesichts des erforderlichen Zeit- und Kostenaufwands möglicherweise nicht.
 4. Genaue Schätzungen des Marktwerts sind erwünscht. RCA sollte in den meisten Fällen eine Genauigkeit von $\pm 1\text{-}2\%$ liefern können. Der Bericht des Valuation Engineer sollte die Bedingungen für die Genauigkeit der abschließenden Wertschlussfolgerung erläutern. Fehlen Daten oder Informationen oder wurden durch Data Mining keine konsistenten Muster im Markt aufgedeckt, so leidet die Genauigkeit darunter, und der Auftraggeber sollte über die Situation informiert werden. Konkret sind ein hohes R^2 (Bestimmtheitsmaß) und CVR2 (kreuzvalidiertes R^2 , berechnet durch wiederholte Aufteilung des MLS-Datensatzes in zufällige Trainings- und Testpartitionen), ein aussagekräftiges MARS-Modell sowie ein starkes Muster der CQA-Merkmale (condition-quality-appeal; Zustand, Qualität, Attraktivität) der Objekte in den nach Rang geordneten Residuen die wichtigsten Mittel zur Beurteilung von Genauigkeit und Zuverlässigkeit der Wertschlussfolgerung.

Die Entwicklung dieser RCA-Protokolle über die vergangenen mehr als 20 Jahre basiert auf der Verfügbarkeit von MLSListings Inc. (Sunnyvale, CA) und anderen MLS, die derzeit über MLSListings zugänglich sind, was über 95% der Datenanalyseerfahrung des Autors umfasst. Die Expertise des Autors beruht auf umfangreicher Gutachtertätigkeit in 15 primären Countys Nordkaliforniens sowie zusätzlicher Erfahrung in 10 sekundären Countys.

Die RCA-Methodik ist am wirksamsten in Märkten, die zwei zentrale Kriterien erfüllen:

1. Reichhaltige Verfügbarkeit hochwertiger Wertermittlungsdaten
2. Komplexe, heterogene Wohngebiete, gekennzeichnet durch:
 - Vielfältige Architekturstile
 - Unterschiedliche Grade der Objektmodernisierung
 - Über mehrere Jahrzehnte gewachsene Bebauungsmuster

Erfahrung in kalifornischen Countys

Die in diesem Beitrag getroffenen Aussagen beruhen auf Erfahrungen bei der Bewertung von Objekten in den folgenden nordkalifornischen Countys:

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

Die vorgenannten Gebiete umfassen im Wesentlichen das Silicon Valley und die umliegenden Countys und erstrecken sich nach außen in ländlichere Gebiete, einschließlich des Central Valley und der Ausläufer der Sierra. Es sei erwähnt, dass der Autor stets Zugang zu reichlichen und

relativ genauen Daten hatte, verglichen mit dem, was vielen Gutachtern in anderen, weniger entwickelten Gebieten der USA zur Verfügung steht.

Die verwendeten MLS-Daten sind RESO-zertifiziert und entsprechen dem RCF (RESO Common Format). Trotz RCF-Zertifizierung fehlen Daten von üblicherweise sekundärer Bedeutung bei einigen Objekten mitunter oder sind fehlerhaft, sei es durch ursprüngliche Eingabefehler des Steuergutachters (tax assessor) oder durch nachträgliche MLS-Eingabefehler der Verkaufsagenten. Häufig folgen diese Fehler einem Muster, etwa indem schlicht keine Eingabe erfolgt, wenn das Merkmal nicht vorhanden ist. Beispielsweise wird, statt 0 für „carport spaces“ einzugeben, wenn kein Carport vorhanden ist, nichts eingetragen. In Gebieten, in denen Carports selten sind, kann der Analyst 0 für fehlende Daten einsetzen, um eine brauchbare Anpassung für Carports zu erhalten. Zwar können R/earth und andere MARS-Implementierungen mit fehlenden Daten umgehen, doch obliegt es dem Valuation Engineer (VE) zu entscheiden, wie damit zu verfahren ist.

MLSListings-Daten sind auch aus den MLS einiger anderer Bundesstaaten verfügbar, von Florida bis Hawaii. Dies kann nützlich sein, um zu testen, wie gut Protokolle und Code in verschiedenen Bundesstaaten funktionieren.

3 Valuation Engineer

Definition 1: Valuation Engineer

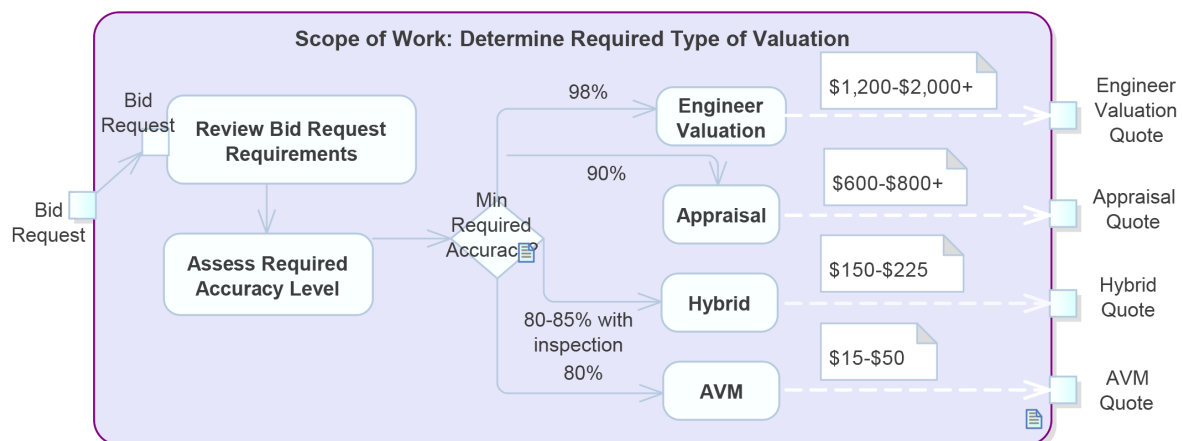
Zur Vermeidung von Mehrdeutigkeit wird der Begriff „Valuation Engineer“ (kurz „VE“) verwendet, um eine Person zu bezeichnen, die über die Fachkompetenz eines lizenzierten Gutachters verfügt, verbunden mit fortgeschrittenen Kenntnissen, Fähigkeiten und Erfahrungen im Einsatz von Data Mining, Statistik, Programmierung, detaillierten Protokollen und künstlicher Intelligenz zur Ermittlung des Marktwerts komplexer Objekte. Es ist ferner wichtig anzuerkennen, dass die Bewertung eine Aufgabe ist, die aufgrund des erwiesenen Betrugspotenzials eine Form der Lizenzierung erfordert.

Kreditgeber bieten üblicherweise mehrere Bewertungsoptionen an:

1. Automated Valuation Models (AVMs) - kostengünstigste Option
2. Hybride Gutachten - Einsatz nicht lizenzierter Besichtiger unter Aufsicht lizenzierter Gutachter
3. Standardgutachten - durchgeführt von lizenzierten Gutachtern

Bei der üblichen Vergabe von Wertermittlungsaufträgen bleiben Gutachten auf Ingenieurniveau (Engineer-level Appraisals) unberücksichtigt. Es gibt schlicht nicht viele Gutachter mit den erforderlichen Fähigkeiten, diese zu erstellen. Ein Bedarf dafür sollte jedoch bestehen. Sie können eine deutlich höhere Genauigkeit bieten und durch den Einsatz der RCA-Methodik (MARS und Restriktionen) eine recht gute Objektivität nahezu garantieren.

Für konforme Objekte in neueren Wohnsiedlungen genügen in der Regel Standardgutachten. Für ältere Häuser in Gebieten mit erheblicher, sich über mehr als 30 Jahre erstreckender Renovierungsgeschichte werden jedoch Ingenieurbewertungen empfohlen. Die Wahl zwischen diesen Methoden sollte einem strukturierten Entscheidungsprozess auf Grundlage der Objekteigenschaften und Marktbedingungen folgen.



4 Annahmen

Annahme 1: Geographisches Gebiet & Objekttyp

Der Gegenstand dieses Beitrags konzentriert sich in erster Linie auf Einfamilienhäuser und Eigentumswohnungen in den Metropol- und Stadtgebieten Kaliforniens. Zwar ist der Sales Comparison Approach (vergleichbar dem deutschen Vergleichswertverfahren) auch auf andere Objekttypen und geographische Regionen anwendbar, jedoch spielen Ertrags- und Sachwertverfahren typischerweise eine dominantere Rolle bei der Bewertung von landwirtschaftlichen Flächen, Mehrfamilienobjekten und Gewerbeimmobilien. In solchen Fällen kann der RCA-Ansatz einer Modifikation bedürfen oder nur begrenzt anwendbar sein.

Annahme 2: Datenspeicherung

CSV und Excel werden von MLS-Systemen intensiv genutzt und gelten als der hauptsächliche Weg, Daten in RCA-Programme einzuspeisen. Broker-Feeds können allerdings verwendet werden, um Daten kontinuierlich in SQL-Datenbanken einzuspeisen. Für Sicherungen und Zwischenspeicherung werden häufig SQL-Datenbanken verwendet; zwei, die eine große Rolle spielen, sind die freien und quelloffenen Systeme SQLite und PostgreSQL. Text- oder RTF-Dateien werden zudem für Berichtsfragmente verwendet, PNG-, JPG- und TIFF-Dateien für Grafiken. Es sei angemerkt, dass auch Microsofts SQL Server von vielen verwendet wird, jedoch ist die freie Edition „SQL Server Express“ stark eingeschränkt (z.B. kann nur 1 CPU-Kern genutzt werden), und andere Versionen sind mit erheblichen Lizenzgebühren für die kommerzielle Nutzung verbunden. Immobiliendatenbanken in Kalifornien können sehr groß sein, und Joins können sehr zeitaufwendig sein, sodass ein Rechner mit 8+ Kernen und 128G RAM für große Joins nützlich sein kann. SQL-Datenbanken sind ebenfalls wichtig, wenn alle Objekte in einem beliebig begrenzten Gebiet gefunden werden müssen. Wurden also beispielsweise Wohnviertel mittels Clusteranalyse definiert, so läge die Verwendung einer SQL-Datenbank nahe. Allerdings verfügen verschiedene Datenbanken über unterschiedliche eingebaute Funktionen zur Verarbeitung geographischer Informationen und sind daher untereinander nicht kompatibel. Für diese Artikelreihe über RCA wird PostgreSQL für komplexe SQL-Aufgaben und SQLite für einfache Aufgaben angenommen.

Annahme 3: „R/earth“ wird für die MARS-Verarbeitung verwendet.

Bis 2017 war die MARS-Software von Salford Systems mit \$270 pro Jahr eine erschwingliche Lösung. Nach der Übernahme von Salford Systems durch Minitab wurde MARS jedoch in deren SPM-Paket zu \$16,000 jährlich gebündelt — ein Preisniveau, das für die meisten kleinen Unternehmen prohibitiv ist. Glücklicherweise hat sich das „earth“-Paket (Milborrow 2024a) der Programmiersprache R, gepflegt von Stephen Milborrow, dahingehend entwickelt, dass es die Fähigkeiten von Minitabs SPM MARS erreicht oder übertrifft. Das earth-Paket bietet zusätzliche Vorteile: Es ist frei verfügbar, integriert sich nahtlos in das umfassende statistische Ökosystem von R und kann über Python mittels Pandas angesprochen werden, was ausgezeichnete Flexibilität für unterschiedliche Arbeitsabläufe bietet. Erörterungen im Zusammenhang mit MARS setzen die konkrete Verwendung von R/earth voraus.

Annahme 4: Notation

Rohdaten aus externen Quellen wie CSV-Dateien und Excel-Tabellen werden mit lateinischen Standardzeichen dargestellt. Dies umfasst Objektdaten des Multiple Listing Service (MLS), MARS-Konfigurationseinstellungen (Multivariate Adaptive Regression Splines) unter Verwendung des R/earth-Pakets sowie projektspezifische Daten. Diese Notation erstreckt sich auch auf die entsprechenden Data-Frame-Darstellungen in R oder Python. Bei der Erörterung von *Objektmerkmalen* und ihren verschiedenen Typen und Teilmengen wechselt die Notation zu griechischen Buchstaben, um Präzision und analytische Klarheit zu erhöhen. Auch wenn diese mathematische Notation strenger erscheinen mag, erfüllt sie einen entscheidenden Zweck: die Unterscheidung abstrakter Objekteigenschaften von ihren Rohdatendarstellungen. Als Konvention bezeichnen griechische Zeichen durchgängig Objektmerkmale, während lateinische Zeichen Tabellendatenstrukturen kennzeichnen. Beispielsweise stellt das Symbol M die MLS-Datenmatrix dar, organisiert mit Objekten als Zeilen und ihren entsprechenden Merkmalen als Spalten.

Annahme 5: Programmiersprachen

In diesem Bereich sind R und Python mit Pandas die dominierenden Sprachen, wobei C oder C++ verwendet wird, wenn Geschwindigkeit wichtig ist. R/earth selbst ist in R und C geschrieben.

Annahme 6: Indizierung

Die Indizierung kann zwischen Programmiersprachen und der Mathematik variieren, was verwirrend sein kann:

- Mathematik und R beginnen die Indizierung bei 1.
- C, C#, C++ und Python beginnen bei 0.

In diesem Beitrag beginnt zur Wahrung von Konsistenz und Einfachheit sämtliche Indizierung bei 0 (wie in Python oder C). Dies funktioniert wie folgt:

- Für Listen, Matrizen oder Data Frames:
 - Index 0 ist für das „Bewertungsobjekt“ (das untersuchte Hauptobjekt) reserviert.
 - Beispiel: In einer Tabelle, deren Zeilen Objekte darstellen, ist Zeile 0 das Bewertungsobjekt, und Spalte 0 enthält eindeutige IDs für alle Objekte.
 - Ein praktischer Hinweis: Die Objekte sollten so angeordnet werden, dass das Bewertungsobjekt zuerst steht (Zeile 0); die übrigen folgen vom neuesten zum ältesten Verkaufsdatum und werden fortlaufend nummeriert (0, 1, 2 usw.).
- Wenn kein Bewertungsobjekt vorhanden ist:
 - Enthält eine Datenstruktur keine Objekte (z.B. nur Variablen oder Funktionen), wird Index 0 übersprungen, und die Indizierung beginnt bei 1.
 - Besonderer Hinweis für R-Anwender: Da R eine 1-basierte Indizierung verwendet, ist beim Kodieren der Mathematik dieses Beitrags in R zu allen Indizes 1 zu addieren, um von der 0-basierten auf die 1-basierte Indizierung umzustellen.

5 RCA-Workflow

Das Verständnis des RCA-Workflows ist für eine effektive Analyse unerlässlich. Der Prozess folgt einer strukturierten Abfolge analytischer Aufgaben und Entscheidungspunkte

5.1 Arten der MARS-Analyse

Der Arbeitsablauf einer MARS-Analyse hängt davon ab, ob lediglich eine Marktanalyse benötigt wird oder ob die Bewertung bzw. das Gutachten eines bestimmten Objekts erforderlich ist. Zu beachten ist, dass ein Gutachten für ein bestimmtes Objekt in der Regel eine Analyse seines Marktgebiets voraussetzt.

- Marktanalyse: konzentriert sich auf großräumige Gebiete wie Countys, Städte oder definierte Regionen. Dieser Ansatz umfasst kein bestimmtes Bewertungsobjekt und erfordert daher keine Wertschlussfolgerung. Der Arbeitsablauf konzentriert sich auf die Modellentwicklung, die Validierung und die Erstellung unterstützender Analysen. Der Prozess endet, sobald das optimale Modell identifiziert und dokumentiert ist.
- Die Einzelobjektanalyse umfasst alle Elemente der Marktanalyse sowie zusätzlich die Ermittlung des Wertes eines bestimmten Objekts. Dies unterscheidet sich grundlegend von traditionellen Wertermittlungsmethoden. Bei traditionellen Ansätzen wird der Wert typischerweise abgeleitet, indem 3-9 Vergleichsverkäufe angepasst werden und ein gewichteter Durchschnitt berechnet wird. Im Gegensatz dazu ermittelt der RCA den Wert primär über zwei Komponenten:
 1. den vom MARS-Modell prognostizierten Wert
 2. das geschätzte objektspezifische Residuum (mit dem zugehörigen CQA-Score)

Das Sales Grid dient im RCA primär dazu, Wertunterschiede zwischen dem Bewertungsobjekt und den Vergleichsobjekten durch die Darstellung der Variablenbeiträge zu erklären. Ein wesentlicher Vorteil des RCA ist seine Stabilität: Während Auftraggeber über die Auswahl der Vergleichsobjekte diskutieren mögen, bleibt die Wertschlussfolgerung unabhängig davon konsistent, welche Vergleichsobjekte für die Darstellung ausgewählt werden, da alle Verkäufe (ob 15 oder 600) mit demselben Modell angepasst werden. Ältere Verkäufe werden zwar automatisch an die aktuellen Marktbedingungen angepasst, dennoch sind jüngere Verkäufe in der Regel vorzuziehen. Die Fähigkeit des Modells, Preisfaktoren zu erfassen, nimmt mit zeitlicher Distanz ab, da nicht gemessene Marktcharakteristika über längere Zeiträume an Bedeutung gewinnen.

5.2 Verarbeitungs-Workflow

Der gesamte Verarbeitungsablauf lässt sich durch folgende Schritte beschreiben:

1. Festlegung des Leistungsumfangs (Scope of Work).
2. Einrichtung der initialen Projektordner und -dateien.
3. Herunterladen der MLS-Daten. Zur erforderlichen Anzahl an Objektdatensätzen und zur Festlegung von Beschränkungen der Datumsbereiche sei auf die Protokolle 1.1–1.3 verwiesen.
4. Ausfüllen der Konfigurationsblätter: Regression Variables, Interactions, Aggregations, Calculations.

5. Durchführung der MARS-Regression (R/earth) zur Erstellung des Preismodells.
6. Erzeugung der Wertbeiträge & Residuen der Objekte (reines Residuum und Residuum/SF) aus dem MARS-Modell.
7. Ranking der Objekte nach Residuum/SF.
8. Erstellung der CQA-Scores für jeden Vergleichsverkauf.
9. Unter der Annahme, dass ein Qualitätsranking der Objekte erstellt wurde, das die Gesamtattraktivität der Objekte tatsächlich widerspiegelt, ist durch Untersuchung der übergreifenden Mustertrends im Ranking die beste Position für das Bewertungsobjekt im Ranking zu ermitteln und diesem ein CQA-Score zuzuweisen.
10. Zuweisung des entsprechenden Residuums/SF an das Bewertungsobjekt als Residuum/SF · GLA (des Bewertungsobjekts).
11. **Addition der MARS-Preisschätzung und des zugewiesenen Residuums zur Wertschlussfolgerung für das Bewertungsobjekt.**
12. Aufgliederung des Residuums des Bewertungsobjekts in deskriptive Komponenten und Wertbeiträge.
13. Berechnung der modellbasierten Wertbeiträge & Anpassungen der Vergleichsobjekte sowie aller Zwischensummen und angepassten Verkaufspreise.
14. Die angepassten Verkaufspreise entsprechen der endgültigen Wertschlussfolgerung für das Bewertungsobjekt.
15. Auswahl der 6-12 ähnlichsten Objekte für das Sales Grid.
16. Aufgliederung der gesamten Residualbeträge der Vergleichsobjekte in Wertbeiträge der Residualkomponenten, wobei der Residual-Gesamtwert nach Abschluss der Aufgliederung auf 0 gesetzt wird.
17. Neuberechnung aller Zwischensummen und der angepassten Verkaufspreise der Vergleichsobjekte im Sales Grid. Überprüfung der Werte.
18. Übernahme des Sales Grid in den Bericht.

Zu beachten ist, dass sich beim RCA die endgültige Wertschlussfolgerung NICHT direkt aus den Werten der Vergleichsobjekte im Sales Grid ableitet. Bei der RCA-Methode wird das Sales Grid ausschließlich zu einem Instrument, um zu erklären, warum Objekte, die dem Bewertungsobjekt ähnlich erscheinen, zu einem höheren oder niedrigeren Preis verkauft werden als die endgültige Wertschlussfolgerung bzw. der geschätzte Wert des Bewertungsobjekts. Dies muss aus einer anderen Perspektive betrachtet werden: Die Wertschlussfolgerung für das Bewertungsobjekt hängt von der Preisschätzung des MARS-Modells und dem vom Valuation Engineer geschätzten Residuum ab. Diese beiden Größen sind jedoch in hohem Maße von sämtlichen Objekten abhängig, die in die R/earth-Regression eingegangen sind, welche das Ranking der Objektresiduen erzeugt hat, das den CQA-Score des Bewertungsobjekts bestimmt und damit dessen Residuum durch die Platzierung im Ranking durch den Valuation Engineer festgelegt hat. Die Anpassungen im Sales Grid werden in Wirklichkeit durch die auferlegten mathematischen Restriktionen des RCA bestimmt; obwohl die Residuenzerlegung in gewissem Maße dem Ermessen des VE unterliegt, müssen sich die Summen der aufgegliederten Wertbeiträge stets zum zugehörigen Residuum addieren. Aus Beweis 2 ist bekannt, dass zwar die tatsächlichen Werte der Residualkomponenten je nach Urteil des VE variieren können, ihre Summen für ein gegebenes Objekt jedoch vollständig restringiert sind und daher den angepassten Verkaufspreis des zugehörigen Objekts nicht verändern, unabhängig davon, wie der VE die Residuenzerlegung abändert.

5.3 Daten-Workflow

Es sei darauf hingewiesen, dass in diesem Beitrag nur ein grober Arbeitsablauf für die RCA-Methode dargestellt wird. Weitere Artikel werden sich eingehend mit den verschiedenen Stufen und Teilstufen der RCA-Analyse befassen.

5.3.1 Datenanforderungen für MARS

Es müssen ein Anfangsdatum für die Verkaufstransaktionen sowie ein Enddatum auf Grundlage des Wertermittlungstichtags der Wertermittlung bzw. Bewertung festgelegt werden. Dabei werden üblicherweise die geschätzte Anzahl der Regressionsvariablen sowie die Interaktionen berücksichtigt, die für jeden Interaktionsgrad über 1 benötigt werden.

5.3.2 Protokolle: Listings abgeschlossener Verkäufe

Protokoll 1.1: Bestimmung der erforderlichen Anzahl an MLS-Listings abgeschlossener Verkäufe

Dieses Protokoll beschreibt, wie die empfohlene Mindestanzahl an Listings abgeschlossener Verkäufe berechnet wird, die in die MARS-Regression eingehen sollen. Zunächst ist eine grobe Schätzung vorzunehmen; anschließend wird MARS in Läufen ersten, zweiten und gegebenenfalls dritten Grades ausgeführt, um das initiale Modell zu erzeugen, und danach wird die Anzahl der Zeilen anhand der folgenden Regeln überprüft, um festzustellen, ob weitere Zeilen benötigt werden. Dieses Protokoll geht davon aus, dass die unabhängigen Variablen zu 100% unabhängig sind, was in der Immobilienwirtschaft selten der Fall ist; beispielsweise sind GLA und Zimmeranzahl in der Regel hoch korreliert, und GLA dürfte auch mit der Grundstücksgröße korrelieren. Dies führt zur Berechnung der „Degrees of Freedom (DOF)“ (Freiheitsgrade). Jede wirklich unabhängige Variable erhält einen Freiheitsgrad. Bestehen jedoch Abhängigkeiten zwischen den sogenannten unabhängigen Variablen, erfordert eine genaue Berechnung der DOF eine Analyse, die über den Rahmen dieses Beitrags hinausgeht. Es ist lediglich zu bedenken, dass die nachstehenden Schätzungen aufgrund wahrscheinlicher Kollinearität zwischen den Variablen vermutlich eher zu hoch angesetzt sind.

1. Für Terme erster Ordnung (ohne Interaktionen) bewährt sich die Regel von ~ 15 Zeilen pro Variable
2. Für jede zugelassene Zwei-Wege-Interaktion sollten etwa 15-30 Objekte hinzugefügt werden
3. Für Interaktionen höherer Ordnung (3+ Variablen) sollten jeweils 30+ Objekte vorhanden sein.

Dies ist jedoch keine feste Regel. Die tatsächlich benötigte Anzahl kann abhängen von:

- dem Signal-Rausch-Verhältnis der Daten
- der Komplexität der zu modellierenden Zusammenhänge
- der Variabilität der Zielvariablen
- der Qualität der Daten

Beispiel:

Liegen also vor:

- 10 Hauptvariablen
- 5 zugelassene Zwei-Wege-Interaktionen
- 2 zugelassene Drei-Wege-Interaktionen

so könnte die Mindestanzahl an Verkäufen wie folgt berechnet werden:

$$\begin{aligned}
 N &= (10 \text{ Variablen} * 15 \text{ Zeilen/Variable}) + \\
 &\quad (5 \text{ Zwei-Wege-Interaktionen} * 22 \text{ Zeilen/Zwei-Wege-Interaktion}) + \\
 &\quad (2 \text{ Drei-Wege-Interaktionen} * 30 \text{ Zeilen/Drei-Wege-Interaktion}) \\
 &= 320 \text{ Zeilen}
 \end{aligned}$$

Alternativ kann wie folgt vorgegangen werden: Nach einem Lauf von R/earth auf einem Datensatz kann die Anzahl der Basisfunktionen im Modell plus eins als Schätzung für die Freiheitsgrade verwendet werden; diese Zahl wird dann mit 20-30 multipliziert, um die Anzahl der benötigten Eingabedatensätze zu erhalten.

Protokoll 1.2: Verkaufszeiträume sollten möglichst geringe Veränderungen der Käuferpräferenzen umfassen

Bei der Erhebung historischer Objektdaten besteht keine strikte zeitliche Begrenzung, sofern Veränderungen der Marktpreferenzen durch geeignete Variablen erfasst werden können. Für den Umgang mit der architektonischen Entwicklung im jeweiligen Markt sind folgende Kernpunkte zu berücksichtigen: Veränderungen der vorherrschenden Architekturstile (etwa ein Wandel von California Ranch zu Mediterranean) können modelliert werden mittels:

- Subdivision-Kennungen
- MLS-Gebietscodes
- Baujahresbereichen

Können diese Stilveränderungen jedoch nicht angemessen durch spezifische Variablen erfasst werden, ergibt sich eine wichtige Einschränkung: Während das Verkaufsdatum typischerweise zur Anpassung an die Marktbedingungen verwendet wird, würde seine gleichzeitige Verwendung zur Kontrolle architektonischer Präferenzen diese beiden unterschiedlichen Effekte konfundieren. In solchen Fällen besteht der bevorzugte Ansatz darin, entweder:

1. die Datenerhebung auf den Zeitraum nach der Einführung des neueren Architekturstils zu begrenzen
2. die Auswahl der Vergleichsobjekte auf Subdivisions mit demselben Architekturstil zu beschränken

Dies stellt sicher, dass die Analyse die Konsistenz der Objektmerkmale wahrt und zugleich eine sachgerechte Anpassung an die Marktbedingungen im Zeitverlauf ermöglicht. Alternativ könnte, zumindest in diesem Fall und sofern sich der Architekturstil aus den Daten extrahieren lässt, die Interaktion zwischen diesen beiden Variablen berücksichtigt werden. Eine solche Extraktion ist jedoch nicht immer möglich.

Protokoll 1.3: In Gebieten mit wenigen Verkäufen das Bestmögliche erreichen

Bei der MARS-Modellierung von Verkaufsdaten werden für eine aussagekräftige Analyse etwa 15 Transaktionen benötigt. Liegen weniger Verkäufe vor (beispielsweise 5 Objekte), kann jeder Datensatz zweifach dupliziert werden, um den Schwellenwert von 15 Objekten zu erreichen. Wichtige Überlegungen zu diesem Vorgehen:

1. Es sollte ausschließlich eine Regression ersten Grades verwendet werden, um Überanpassung zu reduzieren
2. Die Kreuzvalidierung sollte ignoriert bzw. übersprungen werden, da der Datensatz zu begrenzt ist
3. Auch wenn MARS in diesem Szenario voraussichtlich Matched-Pairs-Verfahren oder der klassischen linearen Regression überlegen sein wird, ist zu beachten, dass die Duplizierung von Daten die Validität von Unsicherheitsmessungen und statistischer Signifikanz beeinträchtigt

Die beste langfristige Lösung besteht darin, den Datensatz um zusätzliche tatsächliche Verkäufe zu erweitern. Muss jedoch mit begrenzten Daten fortgefahren werden, bietet die Datenduplizierung eine praktikable vorübergehende Lösung, deren statistische Einschränkungen anzuerkennen sind.

5.4 Projektordnerstruktur

Es erscheint sinnvoll, die Projektordnerstruktur zu erörtern. Es empfiehlt sich, jeweils einen separaten Ordner für Marktanalysen und Einzelobjektgutachten anzulegen. Beide Ordner weisen eine ähnliche Unterordnerstruktur auf:

```

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
│   │       │           └── ...

```

Die Verarbeitung in Stufe 1 erstellt den Versionsordner mit seinen Unterordnern. Sie kopiert die Datei `MLS_Original.xlsx`, obwohl sich diese Tabelle *eigentlich* nie ändern sollte, sondern exakt das widerspiegeln soll, was zu einem bestimmten Datum und Zeitpunkt aus dem MLS heruntergeladen wurde; bei einem länger andauernden Projekt besteht jedoch die Möglichkeit, dass ein neuer MLS-Download erforderlich wird. Anschließend erstellt diese Stufe eine Kopie der erweiterten Arbeitsmappe `MLS.xlsx` mit ihren Konfigurationsparametern im Versionsordner. Es kann viele Versionsordner geben, die jeweils für einen vollständigen MARS-Lauf angelegt werden.

Die nachfolgenden Stufen erzeugen sodann Spaltenergänzungen, Sortierungen und weitere Änderungen an der Arbeitsmappe und schreiben eine aktualisierte Version als `MLS_Stage(N)` in den Datenordner des Projekts, wenn das Gutachten abgeschlossen ist. Nochmals: Die `MLS.xlsx` im Wurzelverzeichnis unterliegt ständiger Veränderung, sodass die Versionskopie eine fortlaufende Mutation der ursprünglichen `MLS`-Arbeitsmappe darstellt, mit Löschungen und Ergänzungen von Variablen sowie weiteren Änderungen. Der verarbeitende R- oder Python-Code wird höchstwahrscheinlich mit der Arbeitsmappe `MLS.xlsx` aus dem Data-Ordner als Eingabe ausgeführt. Somit liegt stets eine Kopie sämtlicher Daten vor, die jeder Version des RCA-Laufs zugrunde liegen.

Es werden so viele Versionen erstellt wie nötig, bis keine Verbesserungen mehr möglich sind; an diesem Punkt wird die beste Version als endgültige Version ausgewählt. Der Ordnername könnte durch Anhängen von „_Final“ geändert werden, sodass dies beim Durchsehen des Verzeichnisses ersichtlich ist. Einige ältere Versionen mit interessanten oder konkurrierenden Modellen sollten in der Regel aufbewahrt werden. Jeder Versionsordner enthält eine Reihe von Text-, Diagramm- und Grafikdateien, die in einen Bericht eingebettet werden können. Die endgültige Arbeitsmappe `MLS_Stage5.xlsx` enthält so viele Blätter, wie für den Bericht erforderlich sind. Insbesondere kann sie eine Tabelle zur Übertragung in ein Fannie-Mae-Sales-Grid enthalten. Sie wird zudem zwingend die Berechnungen bzw. Aufgliederungen hinter den Aggregationen enthalten. Dies ist potenziell der zeitaufwendigste Teil des RCA-Ansatzes — das Ringen um das perfekte Modell in einem schwierigen Marktgebiet.

Alles, was zur Rekonstruktion der Sales Grids benötigt wird, muss in den Versionsordner kopiert werden, einschließlich des verwendeten R- oder Python-Codes — sofern der Code häufigen Änderungen unterliegt.

5.4.1 Protokolle: Mehrfache MARS-(earth-)Läufe

Protokoll 2.1: Überanpassungsläufe zur gezielten Untersuchung der Preisfaktoren eines Marktgebiets

Bei initialen earth-Läufen sollte Überanpassung strategisch eingesetzt werden, um zugrunde liegende Preisfaktoren aufzudecken. Beispielsweise kann ein Marktgebiet, das durch 800 Verkäufe repräsentiert wird, nur 3 Objekte mit Werkstätten enthalten, und in diesen Einzelfällen kann die Werkstatt einen erheblichen Einfluss auf den endgültigen Verkaufspreis haben. Ein Lauf, der nur den Forward-Pass umfasst, erzeugt ein überangepasstes Modell, das den ungefähren Wertbeitrag dieser Werkstätten ausweist. Da es sich jedoch nur um 3 solcher Verkäufe von 800 handelt, werden die Werkstatt-Terme im Backward-Pass und erst recht unter Kreuzvalidierung entfernt. Der Wert, den der reine Forward-Pass den Werkstätten zuweist, kann notiert und in die Analyse der latenten Variablen übertragen werden, um für jedes Objekt mit Werkstatt verwendet zu werden, das im Sales Grid erscheint. Solche überangepassten Modelle liefern zudem detaillierte Schätzungen für viele weitere Merkmale mit dünner Datenlage, die durch das Pruning des Backward-Pass oder kreuzvalidierte Läufe eliminiert würden.

Jeder dieser explorativen Läufe sollte als eigener Versionsordner gespeichert werden: Die aufgezeichneten Wertbeiträge sind Belege, die der VE später bei der Aufgliederung der Residuen anführen möchte. Selbstverständlich muss sich der VE letztlich auf Pruning und Kreuzvalidierung stützen, um das endgültige Modell robust gegenüber Objekten zu machen, die der Regression nie vorlagen — etwa einem später hinzukommenden Bewertungsobjekt. Überangepasste Läufe sind Instrumente der Untersuchung; sie sind niemals das endgültige Modell.

Protokoll 2.2: Mehrere kreuzvalidierte Läufe unter verschiedenen earth()-Parametern

Marktgebiete unterscheiden sich in Struktur und Größe, und der VE muss in der Regel mit einer Reihe von Läufen experimentieren: um ein Verständnis dafür zu entwickeln, welche Variablen den Preis am stärksten beeinflussen; um zu beurteilen, ob der vorhandene Satz gemessener Variablen den Wert angemessen schätzt oder ob neue Daten hinzugefügt werden müssen (Entfernung zum Wasser, Höhenlage, Entfernung zu Einkaufszentren usw.); und um neue Arten von Interaktionen zu entdecken, während andere als ineffizient oder irreführend ausgeschlossen werden. Letztlich strebt der VE ein gutes R^2 auf den Trainingsdaten und das bestmögliche CVR2 auf den Testdaten an. Um dorthin zu gelangen, kann es erforderlich sein, neue Daten zu erheben, die Genauigkeit vorhandener Daten zu verbessern oder die earth()-Parameter auf Größe und Charakter des Marktgebiets abzustimmen (z. B. hinsichtlich der verfügbaren Freiheitsgrade — siehe Protokoll 1.1). Genau aus diesem Grund unterstützt die oben beschriebene Projektstruktur viele Lauf-Versionsordner: Jedes Kandidatenmodell wird mit seinen Parametern und seinem R^2 /CVR2 aufbewahrt, sodass der Verlauf über die Läufe hinweg verglichen werden kann. Eine Version, deren R^2 steigt, während ihr CVR2 stagniert oder fällt, passt sich an Rauschen statt an Marktsignal an, und erst die Versionshistorie macht dies sichtbar.

Protokoll 2.3: Prolog-Expansion von Bemerkungen und deskriptiven Merkmalsspalten

Obwohl sich diese Funktionalität noch in einer gewissermaßen experimentellen Phase befindet, wird earthUI inzwischen mit einer integrierten Prolog-Engine (vProlog) ausgeliefert und verwendet Definite Clause Grammars (DCGs), um die MLS-Freitextspalten zu parsen — allen voran Public Remarks, aber auch Maklerbemerkungen und kommagetrennte Merkmalslisten wie Raummerkmale oder vorhandene Baulichkeiten. Die Expansion überführt Formulierungen wie „owned solar“, „3 car garage“ oder „40x50 workshop“ in gewöhnliche Merkmalsspalten, die der Regression angeboten werden können, und auf diese Weise gewonnene Merkmale mit dünner Datenlage fließen direkt in die Analyse der latenten Variablen ein (siehe den späteren Unterabschnitt zu Zustands- und Qualitätsbewertungen als approximative Regressoren; die Funktionalität selbst wird im earthUI-Begleitartikel in dieser Ausgabe dokumentiert).

Die Textexpansion vervielfacht die Menge der Kandidatenprädiktoren, und dies erfordert mehrere earth-/RCA-Läufe zur Erprobung: Expandierte Spalten müssen sowohl im Modell als auch außerhalb des Modells getestet, bei Kollinearität zu abgeleiteten Zählwerten oder booleschen Größen konsolidiert und anhand des CVR2 daraufhin überprüft werden, ob sie Marktsignal und nicht Rauschen tragen. Jedes solche Experiment ist als eigener Lauf-Versionsordner zu behandeln, genau wie in den Protokollen 2.1 und 2.2. Und da sich die Parsing-Grammatik selbst noch in der Entwicklung befindet, sollte der VE die extrahierten Spalten anhand des Quelltextes überprüfen, bevor er sich in einem Modell auf sie verlässt.

5.5 Die zentrale Excel-Arbeitsmappe

Da die meisten MLS-Dienste Objektdaten in Excel-CVS-Arbeitsblätter herunterladen, ist Excel vermutlich die bequemste Möglichkeit, Daten und diverse Konfigurationsparameter zu speichern. Für jedes Projekt existiert eine zentrale Excel-Arbeitsmappe namens `MLS.xlsx`, die ein Blatt namens „`MLSData`“ zur Speicherung der aus dem MLS heruntergeladenen Listings abgeschlossener Objektverkäufe sowie mehrere Konfigurations- und Parameterblätter enthält:

```
Arbeitsmappe: MLS.xlsx/  
├─ Arbeitsblatt: Project Data  
├─ Arbeitsblatt: MLSData  
├─ Arbeitsblatt: Regression Variables  
├─ Arbeitsblatt: Calculations  
├─ Arbeitsblatt: Allowed Interactions  
├─ Arbeitsblatt: Aggregations  
└─ Arbeitsblatt: (MARS-Laufparameter)
```

Die Originaldaten werden in einer Arbeitsmappe mit einem Namen wie `MLS_Original.xlsx` gespeichert. Es sollte eine Kopie unter einem anderen Namen erstellt werden, etwa „`MLS.xlsx`“, die in das R- oder Python-RCA-Programm geladen wird. Beide Arbeitsmappen werden im Data-Ordner abgelegt. Die Arbeitsmappe `Data/MLS.xlsx` wird nun voraussichtlich viele Lauf-Iterationen durchlaufen, in denen verschiedene Variablenkombinationen, MARS-Parameter und Änderungen an weiteren Einstellungen erprobt werden. Jede Iteration erzeugt einen neuen Ein-/Ausgabeordner, der mit Datum und Uhrzeit im Format `VER_YYYYMMDDHHMMSS` benannt wird. Jeder Ordner enthält eine Kopie der Eingabe-Arbeitsmappe `MLS.xlsx` sowie die erzeugten Excel-Arbeitsmappen für jede der vier Verarbeitungsstufen (Stufen 2–5), außerdem die ausgegebenen Textdokumente, Diagramme und Grafiken. Zu beachten ist, dass Stufe 5, die letzte Stufe, je nach Art des Auftraggebers und des Leistungsumfangs (SOW) eine Reihe weiterer Blätter für den Bericht enthalten kann.

5.6 Datensicherungen

Beim Schreiben neuer oder geänderter Excel-Tabellen in die Versionsordner des Projekts können zusätzlich Datensicherungen in SQLite- oder PostgreSQL-Datenbanken im Projektordner angelegt werden. Bei den Datenbankdateien ist es weniger wahrscheinlich, dass sie in nachfolgenden Zeiträumen durch Prüfungs- oder Referenznutzung unbeabsichtigt beschädigt werden.

6 Notation der Arbeitsmappe

6.1 Deskriptive Dimensionen

Für eine Arbeitsmappe „M“ existieren im Wesentlichen sechs deskriptive Dimensionen:

1. Gutachten „a“
2. Laufversionen „r“
3. Workflow-Stufen „w“
4. Blätter „s“
5. Objekt-Listing „p“
6. Variablensatz „v“

Eine einzelne Zelle einer bestimmten Arbeitsmappe M lässt sich somit wie folgt beschreiben:

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

(Anmerkung: „*“ bezeichnet einen beliebigen gültigen Index)

oder:

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

wobei:

- n_p = # der Zeilen (abhängig vom Blatt)
- n_v = # der Spalten (abhängig vom Blatt)
- n_r = # der Laufversionen
- n_m = # der Workflow-Stufen
- n_g = # der Analyseprojekte
- n_h = # der Arbeitsmappenblätter

Es ist zu beachten, dass „W“ die gesamte Excel-Arbeitsmappe mit sechs zentralen Datenblättern bezeichnet, die jeweils eine eigene Tabelle enthalten und beim Ausführen des zugehörigen R- oder Python-Programms in separate interne Data Frames geladen werden. Sowohl R als auch Python verfügen über interne Funktionen, die zu einem gegebenen Namen ermitteln können, zu welcher Zeile oder Spalte des Data Frames dieser gehört. Gilt beispielsweise im Subskript v_j , dass $v = („GLA“, „LotSize“, „BathRms“, „BedRms“, \dots)$ und $j=0$, so ist $v_j = „GLA“$. R bzw. Python kann „GLA“ intern derjenigen Spalte zuordnen, in der der Wert gespeichert ist, was beispielsweise Spalte 15 des zugehörigen Data Frames sein könnte. In dieser Arbeit wird daher in Namens-Arrays indiziert: Ein Name wird an das laufende Programm übergeben, welches diesen Namen in den tatsächlichen Spalten- oder Zeilenindex des internen Data Frames umsetzt. Dies ist etwas kompliziert, im Sinne der Präzision jedoch notwendig und überaus zweckmäßig. Es handelt sich um eine „Trennung der Verantwortlichkeiten“ zwischen der Analystenrolle und der Entwicklerrolle, wenngleich in der Praxis der Valuation Engineer (VE) beide Rollen ausüben kann. Zu beachten ist, dass ein R-„data frame“ in Python als „dataframe“ bezeichnet wird.

Ferner ist zu beachten, dass das Blatt mit den MARS-Laufzeitparametern keine Dimension von W darstellt, da sich W auf die Version der in MARS eingegebenen Daten bezieht – getrennt von den Parametern, die festlegen, wie MARS/earth diese Daten verarbeitet.

Aus Gründen der Klarheit wird eine komplexe mehrdimensionale Notation für die Arbeitsmappe vorgestellt, die sich unmittelbar auf die internen R- bzw. Python-Data-Frames bezieht, in welche die Datenblätter der Arbeitsmappe geladen werden. Es handelt sich um eine „schwergewichtige“ Notation. Bei den Beweisen und dergleichen wird jedoch ein Großteil dieses Ballasts abgeworfen, indem Annahmen über den jeweiligen Untersuchungsgegenstand getroffen werden. Gleichwohl verhilft diese Notation von Beginn an zu einem besseren Überblick über den Arbeitsablauf und die Daten.

Zunächst wird die Notation für die Menge der zu verarbeitenden Arbeitsmappen eingeführt.

6.1.1 Die Projektdimension („a“)

Sobald ein Leistungsumfang (Scope of Work, SOW) genehmigt wurde, ist ein Projektordner einzurichten, in dem Eingabe- und Ausgabedaten abgelegt werden, einschließlich Konfigurationsparametern, Ausgabegrafiken, Diagrammen, Tabellenkalkulationen und Berichtsbausteinen. Der VE sollte alle Projekte mit einer einfachen fortlaufenden ID kennzeichnen, etwa einer ganzen Zahl, sodass das Fehlen eines Berichts sofort offensichtlich wäre. Wird ein Projekt storniert, sollte es eine einheitliche Möglichkeit geben, die Stornierung innerhalb des jeweiligen Projektordners zu kennzeichnen; der Ordner selbst sollte jedoch niemals gelöscht werden. Es wird angenommen, dass Projekt-IDs bei 0 beginnen und jeweils um 1 erhöht werden, wenngleich auch andere Varianten möglich sind. Ferner wird angenommen, dass alle gültigen Projekt-IDs in einer geordneten Menge „a“ gespeichert sind, die mit dem Verarbeitungsprogramm erzeugt wird.

Die Daten des Wertermittlungsprojekts befinden sich im Projekt-Arbeitsblatt der Arbeitsmappe, könnten jedoch auch in einer zentralen SQL-Datenbank gespeichert werden,

Nachfolgend die vorgeschlagenen Projektfelder:

1. Lokale ID: Eine kurze ganzzahlige ID, die für das Unternehmen fortlaufend erzeugt wird. Sie eignet sich hervorragend als Primärschlüssel für die Speicherung - und könnte von der Datenbank mit jedem weiteren Projekt automatisch generiert werden.
2. Globale GUID: Bei Bedarf kann eine GUID (Globally Universal ID) erstellt werden unter

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

3. Objektlage (Marktanalyse) oder Adresse
4. Wertermittlungstichtag
5. Auftragsdatum
6. Besichtigungsdatum
7. Datum des ursprünglichen Berichts
8. Datum des ersten aktualisierten Berichts
9. Datum des zweiten aktualisierten Berichts
10. Datum des dritten aktualisierten Berichts
11. Name des Auftraggebers
12. Adresse des Auftraggebers
13. Name des Kreditgebers
14. Adresse des Kreditgebers
15. Name des Eigentümers
16. Adresse des Eigentümers
17.

Beispiel einer Liste von Projekt-IDs:

```
a = („1“,
      „2“,
      „...“,
      „141“,
      „142“)
```

6.1.2 Dimension der Laufversionen („r“)

Als zweiter Schritt des Arbeitsablaufs wird die MARS-Regression auf die Eingabedaten angewendet. Dies umfasst in der Regel mehrere bis viele Ausführungen, in deren Verlauf verschiedene Parameter verändert werden, um das Modell zu verbessern. Jede Ausführung erzeugt einen neuen Versionsordner für die Ausgabe, dessen Name auf Datum und Uhrzeit der Ausführung basiert. Zum Beispiel:

```
r = („ver20241210_083124“,
     „ver20241209_151130“,
     „ver20241208_160944“)
```

```
latest_version : r1 = „ver20241210_083124“
```

Vielversprechende Versionen mit konkurrenzfähigen Modellen und Ausgaben sollten aufbewahrt werden, bis eine endgültige Version festgelegt ist, oder gegebenenfalls länger, sofern sie von Nutzen sind. Der Versionsordner kann zur Kennzeichnung der endgültigen Version angepasst werden, indem „Final“ an das Ende des Namens gesetzt wird.

6.1.3 Workflow-Stufen („w“)

Der RCA-Workflow gliedert sich in der Regel in fünf separate Stufen, wie nachfolgend aufgeführt. Jede Stufe muss erfolgreich abgeschlossen sein, bevor die nächste Stufe begonnen wird. Der VE sollte die Verarbeitung am Ende einer Stufe anhalten können, um die Ergebnisse vor dem Fortfahren zu überprüfen. Dies ist erforderlich, weil Marktgebiete recht komplex und häufig sehr individuell sind. Es können ungewöhnliche Probleme auftreten, die einen neuen Ansatz erfordern. Der VE muss in den R- oder Python-Code eingreifen können, um notwendige Änderungen vorzunehmen und auch neue Ansätze zu erproben. In solchen Fällen wird er unter Umständen experimentieren und verschiedene Ansätze ausprobieren wollen, was mehrere Läufe derjenigen Stufe erfordert, an der er gerade arbeitet. Gleichwohl sollte es dem VE ebenso freistehen, alle Stufen mit einem einzigen Befehl auszuführen, - mit einigen Einschränkungen (z. B. muss die Aggregation dritten Grades bereits durch einen früheren Lauf der Stufe 2 bestimmt worden sein). Dies bedeutet, dass jede Stufe nach ihrem Abschluss sämtliche internen Daten in Dateien oder Datenbanken ablegen sollte, sodass die nächste Stufe – durch erneutes Laden der Daten der vorherigen Stufe in den Speicher – so gestartet werden kann, als befände sich die vorherige Stufe noch im Speicher.

Die Stufen lauten:

```
w = („stage1“, -Einrichtung
     „stage2“, -MARS-Verarbeitung
     „stage3“, -Finales Sales Grid
     „stage4“, -Berichtsbausteine
     „stage5“) -Überprüfung/Bereinigung
```

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

6.1.4 Blätter der Excel-Arbeitsmappe („s“)

Die zentrale RCA-Excel-Arbeitsmappe enthält die folgenden Blätter, die jeweils in einen internen Data Frame abgebildet werden:

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

Aggregationen zweiten Grades werden deshalb auf einem separaten Blatt geführt, weil sie sich am einfachsten mit einer symmetrischen Tabelle definieren lassen, die alle möglichen Interaktionspaare zeigt und für jedes Paar die Angabe einer Aggregationsvariablen erlaubt, zu der die zugehörigen Wertbeiträge und Anpassungen aggregiert (addiert) werden. Bei Aggregationen dritten Grades ist dies nicht möglich, da hierfür eine dreidimensionale symmetrische Tabelle erforderlich wäre, was in Excel nicht realisierbar ist. Überdies umfasste eine 20x20x20-Tabelle 8,000 zu handhabende Zellen. Es ist einfacher, etwaige Termnamen dritten Grades in der Tabelle mit den Aggregationen ersten Grades aufzulisten, zum Beispiel:

Variable	Aggregationsziel
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 Objekt-Listings („p“) - Zeilen der MLSDData

Objekt-Listings bilden die Zeilen des Blatts MLSDData. In den übrigen Blättern treten sie erst in Stufe 4 auf, in der die endgültigen Berichtsbausteine mit der endgültigen Auswahl der Vergleichsobjekte erstellt werden. Nachfolgend eine beispielhafte Liste von Objekt-IDs:

```
p = („MLS823211“, „MLS823456“, „MLS823542“, „MLS823721“, „...“)
```

6.1.6 MLS-Variablen („v“) Spalten der MLSDData

Auf die Behandlung der Variablen wird später näher eingegangen. An dieser Stelle genügt der Hinweis, dass das Blatt MLSDData typischerweise 30-40 Variablen unterschiedlicher Typen

enthält, von denen in der Regel jedoch nur 10-20 als unabhängige Variablen für die MARS-Regression ausgewählt werden. Diese Variablen erscheinen auch in anderen Blättern, dort allerdings typischerweise als Liste aller möglichen Variablen, die für die Übergabe an MARS verwendet werden können. Ein Beispiel für eine Variablenliste ist:

v = („GLA“, „SalePrice“, „BedRms“, „BathRms“, „GarageSF“, „...“)

6.1.7 Abschließendes Notationsbeispiel

Die obige Notation wird nun in einem abschließenden Beispiel zusammengeführt. Unter Verwendung der obigen Listen sei

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

Es ist zu beachten, dass hier eine Nullindizierung verwendet wird, sodass für die ersten Elemente von Listen der Index 0 angenommen wird. Python verwendet ebenfalls eine Nullindizierung, R hingegen eine Einsindizierung.

Die obige Notation bezeichnet somit den Zellwert für:

- die Variable v_0 bzw. die Spalte „GLA“
- für das Objekt p_2 bzw. die ID „MLS823542“
- des Blatts s_0 bzw. „MLSDa“
- für die Workflow-Stufe w_1 bzw. „stage2“
- der Version r_1 bzw. „ver20241209_151130“
- des Projekts a_{141} bzw. „142“

Die obige Notation lässt sich dann auch wie folgt schreiben:

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

6.1.8 Begründung

„Warum werden in der Notation nicht einfach Indizes oder Zahlen verwendet?“ An unsere Notation werden zwei Anforderungen gestellt:

1. Teilmengen: Der Analyst wählt häufig Teilmengen einer Liste aus, etwa Teilmengen von Variablen. Beispielsweise kann das Arbeitsblatt MLSDData 35-40 Variablen enthalten, von denen für einen bestimmten Durchlauf nur 10-20 verwendet werden. Von einem Durchlauf zum nächsten ändert der VE häufig die an MARS/earth übergebenen Variablen.

2. Iteration, z. B. $i = 1, 2, \dots, n_1$

Unsere Algorithmen, Protokolle und Beweise müssen in der Regel alle Elemente einer geordneten Liste vom ersten bis zum letzten durchlaufen können - oder ein bestimmtes Element herausgreifen.

Es wird daher ein Buchstabe benötigt, der die Teilmenge bzw. die Liste von Elementen repräsentiert, und dieser muss über einen weiteren Subskript verfügen, mit dem über die Elemente der Liste iteriert werden kann, um die Zeichenketten zu erhalten, die zur Indizierung der Zeilen oder Spalten des Data Frames dienen.

6.1.9 Matrizen?

Und was ist mit den Matrizen der Linearen Algebra? Matrizen arbeiten ausschließlich mit Zahlen, seien es ganze oder reelle Zahlen. Matrizen als Datenstruktur werden sowohl in R als auch in Python unterstützt. Viele der Variablen, mit denen in MLS-Daten gearbeitet wird, sind jedoch sogenannte „Strings“ bzw. Zeichendaten. Selbst wenn manche Variablen nur aus Zahlen bestehen, handelt es sich häufig nicht um reelle Zahlen, sondern lediglich um Namen. Beispielsweise können die MLS-Gebietscodes einer Stadt („660“, „661“, „662“, „663“, ..., „672“) lauten. Die Earth-Regression erlaubt es, Variablen als „factor“-Variablen zu spezifizieren; in diesem Fall werden die Zahlen lediglich als Namen ohne jede Reihenfolge behandelt. In solchen Fällen liefert das erzeugte Modell lediglich einen Wertbeitrag für jede Zahl anstelle einer Funktion, die über die Zahlenfolge verläuft. Weitere Beispiele sind „C1“, ..., „C6“ für den Zustand und „Q1“, ..., „Q6“ für die Qualität.

6.1.10 Erste Vereinfachung

Die obige Notation für die zentrale Excel-Projektarbeitsmappe eignet sich gut, um die Zusammenhänge herzustellen, ist jedoch zu schwerfällig, um sie bei der Erläuterung von Algorithmen oder mathematischen Beweisen mitzuführen. Bei der eigentlichen Detailarbeit kann davon ausgegangen werden, dass sie innerhalb des gegebenen Projekts, der gegebenen Version, Stufe und des gegebenen Arbeitsblatts stattfindet. Es wird lediglich angegeben, womit gearbeitet wird, sofern dies nicht bereits implizit ist. Alles, was von der obigen Notation tatsächlich benötigt wird, sobald angegeben ist, womit gearbeitet wird, ist:

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

wobei: n_p = # Objekte
 n_v = # Eingabevariablen
 n_s = # Eingabeblätter

6.1.11 Zweite Vereinfachung

Es ist zweckmäßiger, die Arbeitsblätter nach den jeweiligen Arten von Data Frames umzubenennen, die sie unterstützen, statt Indizes zu verwenden. Für jedes Arbeitsblatt haben die Zeilen und Spalten unterschiedliche Funktionen. Die Menge W der Data Frames wird daher wie folgt in die Data Frames M, V, A, I und C aufgeteilt:

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

wobei: $n_p = \# \text{ Objekte}$
 $n_v = \# \text{ Eingabevariablen}$

Hinweis: Das Bewertungsobjekt steht in Objektlisten standardmäßig stets an erster Stelle

Im Excel-Arbeitsblatt „MLSDData“ und dem zugehörigen Data Frame ist die erste Zeile stets dem Bewertungsobjekt vorbehalten. Sie wird niemals mit den übrigen darunter aufgeführten MLS-Objekten sortiert. Sie geht niemals als unabhängige Daten in die MARS-Regression ein, sondern vielmehr als Zieldaten bzw. als abhängige Variable. Dasselbe gilt für Objektlisten. In allen Objektlisten wird stets angenommen, dass das erste Element mit dem Index 0 die ID des Bewertungsobjekts ist, sofern nichts anderes angegeben ist.

Selbst wenn die Aufgabe darin besteht, ein Modell für ein Marktgebiet ohne Bewertungsobjekt zu erstellen, wird die erste Zeile dennoch mit Platzhalter-Objektdaten befüllt. Dies erleichtert das Schreiben von Programmen, das Spezifizieren von Algorithmen und das Führen von Beweisen erheblich. Wird also c zur Darstellung der Objektliste verwendet, so bezeichnet c_0 das Bewertungsobjekt, und c_i , $i = 1, \dots, n_p$, sind die Vergleichs-Listings, die in die MARS/earth-Regression eingehen.

◇ **Regressionsvariablen-Spezifikationen**

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

wobei: $n_v = \# \text{ verfügbare Variablen}$
 $n_s = \# \text{ Spezifikationen}$

Man beachte, dass $n_s \leq n_v$ gilt, da in der Regel eine Teilmenge der verfügbaren Variablen für die Eingabe in die Regression ausgewählt wird.

◇ **Berechnungen**

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

wobei: $n_v = \# \text{ Variablen}$
 $n_c = \# \text{ Berechnungsspezifikationen}$

◇ **Interaktionen**

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

wobei: $n_f = \# \text{ „from“-Variablen}$
 $n_t = \# \text{ „to“-Variablen}$

Man beachte, dass der Data Frame I ein symmetrischer Data Frame ist, in dem die Variablen für die „from“-Zeilen und „to“-Spalten alphabetisch aufgeführt sind. Die Zellen werden mit einer „1“ markiert, um anzuzeigen, dass eine Interaktion zwischen dem zugehörigen Variablenpaar zulässig ist. Andernfalls zeigt eine „0“ an, dass keine Interaktion gestattet ist. Bei einer Interaktion zweiten Grades ist dies vergleichsweise einfach, da nur eine Interaktion möglich ist. Bei einer Regression dritten Grades bestehen drei Paarinteraktionen zwischen allen drei Variablen, wobei jede der Paarinteraktionen zulässig sein muss, damit der Interaktionsterm (bzw. die Interaktionsvariable) aus drei Variablen erzeugt werden kann.

◇ **Aggregationen**

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

wobei: $n_v = \# \text{ Variablen}$

7 Das RCA Sales Grid

Es ist zu beachten, dass das RCA-Protokoll ein R- oder Python-Programm erfordert, das auf R/earth zugreifen und sämtliche detaillierten Schritte ausführen kann. Das Verfahren ist datenintensiv, und die Arbeit könnte keinesfalls manuell durchgeführt werden. Zugleich ist, unter der Annahme, dass das Programm fehlerfrei ist, ein Softwareprogramm erforderlich, um sicherzustellen, dass die erforderlichen Arbeiten präzise und ohne menschliche Fehler ausgeführt werden.

Ebenso ist es für Valuation Engineers (VEs), die diese Arbeit ausführen, unerlässlich, derartige Programme bei Bedarf anpassen zu können, insbesondere dann, wenn sie mit neuen, unvorhergesehenen Marktgebieten, Nachbarschaften, Objekten oder sonstigen neuen Anforderungen oder Problemen konfrontiert sind.

Innerhalb der RCA-Verarbeitung existieren drei Arten von Sales Grids:

1. Das RCA Processing Sales Grid, in diesem Beitrag als „MLSDData“ bezeichnet, ein sehr großer Data Frame mit zugehöriger Kalkulationstabelle, der zahlreiche Zwischendetails enthalten kann, die schließlich einen Aggregationsprozess durchlaufen, um zu dem unten erwähnten RCA Sales Grid zu gelangen. Ein Beispiel wäre zu umfangreich, um es hier darzustellen. Höchstwahrscheinlich werden Valuation Engineers diese Kalkulationstabelle nach ihren eigenen Präferenzen gestalten. Der Verfasser beabsichtigt, zu einem späteren Zeitpunkt einen gesonderten Beitrag über die Entwicklung eines solchen Grids zu verfassen.
2. Das RCA Final Sales Grid, das in Nicht-GSE-Berichten verwendet wird und in etwa wie [Table 2](#) aussieht. Dieses ahmt beispielsweise das Formular von Fannie Mae hinreichend nach, um dem Leser eine gewisse Grundvertrautheit zu vermitteln. Es ist jedoch zu beachten, dass es je Vergleichsobjekt mehr Spalten enthält als das Fannie-Mae-Raster: Jedes Vergleichsobjekt verfügt neben seinen Spalten „Messwert“ und „Anpassung“ über eine Spalte „Wertbeitrag“, und auch das Bewertungsobjekt besitzt eine solche — Spalten, die dem Fannie-Mae-Formular fehlen. Eine einfachere Form zeigt zudem [Table 1](#).
3. Das GSE Upload Sales Grid, das dazu dient, das Sales Grid in Software wie Alamode hochzuladen. Es ist jedoch anzumerken, dass nicht alle Softwarepakete für Gutachtenformulare die Möglichkeit bieten, Sales-Grid-Daten aus einer Kalkulationstabelle hochzuladen. Zwangsläufig können solche Grids keine gesonderte Spalte für Wertbeiträge enthalten, für die GSEs wie Fannie Mae kein Verständnis besitzen. Die Strategie besteht darin, ihnen das von ihnen geforderte Formular zu liefern, dieses jedoch durch ein RCA Sales Grid sowie durch weitere unterstützende Dokumentation, etwa Details zu etwaigen Variablenaggregationen, zu untermauern.

Die RCA Processing Sales Grids ähneln nicht den Fannie-Mae-Sales-Grids, sondern stellen, grob gesagt, eine Transponierte des Fannie-Mae-Sales-Grid dar, bei der Zeilen und Spalten vertauscht sind: Jede Zeile enthält Daten zu einem bestimmten Objekt, und jede Spalte enthält Daten zu einer bestimmten Objektvariablen. Es können ohne Weiteres Hunderte oder Tausende von Objektzeilen und 200 oder mehr Spalten vorliegen. Die erste Zeile ist den Variablennamen vorbehalten und die zweite einem Bewertungsobjekt, selbst wenn lediglich eine Marktanalyse durchgeführt wird. Wird die Kalkulationstabelle in einen Data Frame eingelesen, so wird die erste Zeile des Blattes zum names-Attribut, und die zweite Zeile für das Bewertungsobjekt wird zur ersten Zeile des Data Frame.

Die Spalten des RCA Sales Grid umfassen Identifikationsspalten sowie primäre, externe und synthetische Variablen, die kombiniert und als unabhängige Variablen in MARS eingespeist werden. MARS gibt aus diesen Variablen diejenigen aus, denen es einen signifikanten Einfluss auf

den Verkaufspreis, auch als abhängige Variable oder Zielvariable bezeichnet, beimit. Bei einer Regression zweiten Grades können sich Paare dieser Variablen ergeben, die als neue Variablen wirken. Bei einer Regression dritten Grades können sich zudem Variablentripel ergeben, die als sogenannte unabhängige Variablen wirken, obwohl sie selbstverständlich mit den zugrunde liegenden Variablen hoch korreliert sind.

Tabelle 1: Beispielhafte RCA-Kalkulationstabelle: Vertikales Layout

Bewertungsobjekt			Vergleichsobjekt 1			Vergleichsobjekt 2		
Messwert	Wertbeitrag		Messwert	Wertbeitrag	Anpassung	Messwert	Wertbeitrag	Anpassung
Verkaufspreis			\$1,150,000			\$1,500,000		
Basiswert	\$500,000			\$500,000	\$0.00		\$500,000	\$0
Gemessene Merkmale			Gemessene Merkmale			Gemessene Merkmale		
Wohnfläche	1,530	\$463,500	1,350	\$382,500	\$81,000	1,700	\$540,000	-\$76,500
Grundstücksgröße	6,000	\$200,000	6,500	\$220,000	-\$20,000	7,000	\$240,000	-\$40,000
Alter	53	\$63,600	57	\$68,400	-\$4,800	51	\$61,200	\$2,400
Badezimmer	2	\$15,000	3	\$30,000	-\$15,000	3	\$30,000	-\$15,000
Schlafzimmer	3	\$6,000	4	\$9,000	-\$3,000	4	\$9,000	-\$3,000
Summe gemessener Merkmale		\$1,248,100.00		\$1,209,900.00	\$38,200		\$1,380,200.00	-\$132,100
Residualmerkmale			Residualmerkmale			Residualmerkmale		
Residuum	CQA 5.8	\$30,000	CQA 2.5	-\$59,900	\$89,900	CQA 8.2	\$119,800	-\$89,800
Aufschlüsselung des Residuums			Aufschlüsselung des Residuums			Aufschlüsselung des Residuums		
Design		7,000		-\$15,000	\$22,000		\$35,000	-\$28,000
Zustand		5,000		-\$8,000	\$13,000		\$20,000	-\$15,000
Qualität		8,000		-\$10,000	\$18,000		\$30,000	-\$22,000
Funkt. Nutzbarkeit		5,000		-\$12,000	\$17,000		\$15,000	-\$10,000
Ausstattung		5,000		-\$14,900	\$19,900		\$19,800	-\$14,800
Zwischensumme		30,000		-59,900	\$89,900		119,800	-\$89,800
Gesamt			Gesamt			Gesamt		
Wertbeitrag gesamt		\$1,278,100.00		\$1,150,000.00			\$1,500,000.00	
Angepasster Verkaufspreis					\$1,278,100.00			\$1,278,100.00

8 Variablen

Auf der Eingabeseite der MARS-Ausführung liegen ausschließlich einfache Variablen vor, nämlich:

1. Objektmerkmale, die aus der MLS-Tabelle geladen werden.
2. Abgeleitete Variablen, die vom VE erstellt werden, um die MLS-Variablen zu verfeinern oder Transformationen bestehender Variablen hinzuzufügen. Beispielsweise kann der VE es vorziehen, das Verkaufsdatum in „DaysOffMarket“ zu transformieren, das dem Wertermittlungstichtag des Gutachtens abzüglich des Verkaufsdatums entspricht.
3. Der VE kann Daten aus Drittanbieterquellen importieren, etwa Hypothekenzinssätze. Fehlen Breitengrad-, Längengrad- oder Höhendaten in den MLS-Daten, kann er diese aus anderen Quellen importieren. Diese Variablen können als „externe Variablen“ bezeichnet werden.

Auf der Ausgabeseite von MARS wird mit einem Preismodell gearbeitet, das eine Kombination von Termen ist, die entweder:

1. nur eine der Eingabevariablen enthalten und die Gleichung für den Wertbeitrag dieser Variablen liefern. Diese werden Terme ersten Grades oder (Ausgabe-)Variablen genannt. Der Term selbst wird als Wertbeitragswert behandelt und erhält einen Namen.
2. zwei der Eingabevariablen enthalten und die Gleichung für den Wertbeitrag der Interaktion der beiden gegebenen Variablen darstellen. Diese Terme können Terme zweiten Grades oder (Ausgabe-)Variablen genannt werden.
3. drei der Eingabevariablen enthalten und die Gleichung für den Wertbeitrag der Interaktion zwischen den drei Variablen darstellen. Diese Terme können in dieser Arbeit Terme dritten Grades oder (Ausgabe-)Variablen genannt werden.

Die Ausgabevariablen von MARS sind somit Wertbeiträge, einschließlich des Residual-Wertbeitrags. Sie werden ihrerseits als Variablen in den Berechnungen des Sales Grid behandelt, etwa bei der Berechnung von Anpassungen oder der Erstellung von Grafiken.

Es lassen sich verschiedene Typen von Variablen unterscheiden, die in zwei grundlegende Ontologien fallen: (1) Variablenherkunft und (2) Variablenanwendung, d. h. nicht-residual vs. residual

8.1 Ontologie der Variablenherkunft

1. Variablen ersten Grades
 - 1.1 MLS-Regressionsvariablen
 - 1.2 Externe Regressionsvariablen
 - 1.3 Abgeleitete Regressionsvariablen

2. Faktorvariablen:

Zur Behandlung von Faktorvariablen wird eine Funktion benötigt, die diese identifizieren kann. Da sie in der Excel-Projektkonfigurationsarbeitsmappe spezifiziert werden, wird eine Funktion erstellt, die true zurückgibt, wenn eine ihr übergebene Variable als Faktortyp klassifiziert wurde.

3. Interaktionsvariablen: Diese werden von MARS bei der Modellerzeugung erstellt, wenn der Grad auf 2 oder höher gesetzt ist. Mehr als 3 interagierende Variablen können jedoch nicht empfohlen werden, und zwar aufgrund der Schwierigkeit der Interpretation, der Wahrscheinlichkeit einer Überanpassung sowie der Auswirkungen auf die Freiheitsgrade und der daraus resultierenden erhöhten Anzahl an Objektdatensätzen, die zur Sicherstellung eines robusten Modells erforderlich ist.
 - 3.1 Variablen zweiten Grades: Diese sind häufig; gute Beispiele wären GLA:BathRms, GLA:LotSize, DateOfSale:BathRms, Latitude _Longitude.
 - 3.2 Variablen dritten Grades: Seltenere Interaktionen, die zur Vorsicht mahnen und auf mögliche Überanpassung hindeuten.
4. Residualaufschlüsselungsvariablen

8.2 Ontologie der Variablenanwendung

1. Von MARS erzeugte Wertbeiträge
 - 1.1 Variablen ersten Grades
 - 1.1.1. MLS-Regressionsvariablen
 - 1.1.2. Externe Regressionsvariablen
 - 1.1.3. Abgeleitete Regressionsvariablen
 - 1.2 Interaktionsvariablen
 - 1.2.1. Variablen zweiten Grades
 - 1.2.2. Variablen dritten Grades
 - 1.3 Residualaufschlüsselungsvariablen
 - 1.4 Aggregationsvariablen
2. Vom Ingenieur erzeugte Variablen & Werte
 - 2.1 Residualaufschlüsselungsvariablen
 - 2.1.1. Design/Ästhetik
 - 2.1.2. Landschaftsgestaltung
 - 2.1.3. Funktionale Nutzbarkeit
 - 2.1.4. Aussicht
 - 2.1.5. Zustand
 - 2.1.6. Qualität
 - 2.1.7. ...

Diese Variablengruppen erhalten im nächsten Abschnitt ihre Notation und werden dort ausführlicher beschrieben

Protokoll 3.1: Faktorvariablen

Faktorvariablen, oder kurz Faktoren, können in allen Termen vorkommen, einschließlich der Interaktionen. Sie sind wichtig, weil sie für die MARS-Regression und die grafische Darstellung gesondert behandelt werden müssen.

Während Variablen mit Zeichenwerten notwendigerweise Faktoren sind, müssen numerische Variablen als Faktorvariablen gekennzeichnet werden, damit MARS sie als solche erkennt. Beispielsweise werden MLS-Gebiete häufig durch Nummern identifiziert. Insbesondere werden die Nachbarschaftsgebiete in der San Francisco Bay Area oft mit eindeutigen ganzen Zahlen identifiziert. Werden diese Werte als Faktoren an MARS übergeben, so versucht MARS zu ermitteln, ob bestimmte Gebiete einen durchschnittlichen Wertbeitrag zu den Objektverkäufen innerhalb ihrer Grenzen leisten, ohne Rücksicht auf irgendeine Beziehung zu anderen Gebieten. Bei normalen Variablen wird dagegen unterstellt, dass eine gewisse Kontinuität der Wertbeiträge besteht, wenn sich der Variablenwert ändert. Dies ist eine wichtige Unterscheidung. Darüber hinaus können diese Faktorvariablen zusammen mit Nicht-Faktorvariablen in Modelltermen auftreten, was die grafische Darstellung erschwert. Konkret werden bei der grafischen Darstellung von Termen mit einer einzigen Variablen, die Faktoren enthalten, alle diese Faktoren zusammengefasst und ihre — positiven oder negativen — Beiträge einfach als Balkendiagramm dargestellt.

Treten die Faktoren zusammen mit anderen Variablen in Termen auf, so erscheinen sie stets mit genau einem möglichen Wert, denn Faktorvariablen können für ein gegebenes Objekt nur einen Wert annehmen: Hat ein Gebiet den Wert 662, kann es definitionsgemäß keinen anderen Wert für dieses Gebiet geben. Sollte dies verwirrend erscheinen, kommt erschwerend hinzu, dass es im von MARS erzeugten Modell keinen einzelnen Variablennamen für den Faktor gibt; MARS übernimmt den Stammmamen der als Eingabe übergebenen Faktorvariablen und erzeugt dann neue Spalten mit Variablennamen, die durch Anhängen der möglichen Werte gebildet werden. So werden für eine einzelne Spalte in der Excel-Tabelle, etwa `MlsArea`, mehrere Data-Frame-Spalten für jeden möglichen Wert erzeugt, etwa `MlsArea660`, `MlsArea661` usw. Eine Variable wie `MlsArea661` hat für Objekte im Gebiet 661 den Wert 1, andernfalls 0. MARS regressiert dann auf diese Spalten, als repräsentierten sie boolesche Variablen. Diese modifizierten Namen werden in dem von MARS gelieferten Modell verwendet.

In der Regel empfiehlt es sich, diese Modellnamen wieder in ihre ursprünglichen Namen zurückzuführen, um die Grafiken für Kunden verständlicher zu machen, die es gern schlicht und einfach haben (zumindest so einfach wie möglich). Statt `MlsArea661=1` erscheint dann `MlsArea=661`. Terme lassen sich anschließend grafisch darstellen, indem in der Grafiküberschrift beispielsweise „Für MLS Area = 662“ angegeben wird. Somit werden Terme N-ten Grades mit M Faktorvariablen als Terme (N-M)-ten Grades dargestellt, und die Faktordimension verschwindet in den Titel.

Es ist zu beachten, dass zwei oder mehr Faktorvariablen in einem einzigen Term vorkommen können, sodass vermutlich für jedes mögliche Wertepaar eine eigene Grafik benötigt wird. Treten die Variable x mit drei verschiedenen Werten und die Variable y mit zwei verschiedenen Werten zusammen mit der Nicht-Faktorvariablen z in demselben Term auf, so werden $2 \cdot 3 = 6$ Grafiken benötigt. Ein Beispiel wäre „`MlsArea`“ + „`Architectural Style`“, d. h. viktorianische Häuser in San Francisco können je nach Stadtgebiet von San Francisco mehr oder weniger wert sein. Glücklicherweise gibt es in den meisten Gebieten üblicherweise nur 1-3 Architekturstile, die als Wertbeitragsträger bedeutsam genug sind, um im Modell verwendet zu werden (auch wenn es 6+ verschiedene Architekturstile geben mag).

8.3 Terme zweiten und dritten Grades

Bei einer Regression zweiten oder dritten Grades können einige der Terme des Ausgabemodells zwei bzw. drei Variablen enthalten. Zum Beispiel:

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

Es ist zu beachten, dass diese Terme benannt werden, indem die Variablennamen alphabetisch sortiert und, etwa mit einem Unterstrich, verkettet werden. Ebenso wird bei Termen dritten Grades mit den im Modell erzeugten Variablennamen-Tripeln verfahren. Zum Beispiel: „BathRms_GLA_LotSz.“

Diese Variablen werden in Stufe 3 zu den primären Eingabevariablen hinzugefügt, wenn das RCA-Programm die Beiträge aller Variablen für jedes Objekt berechnet, einschließlich der Interaktionsvariablen.

8.3.1 Anpassungsvariablen

Das RCA-Programm berechnet in Stufe 3, nachdem die Wertbeiträge berechnet wurden, außerdem die Anpassungen für alle Variablen und Objekte, indem der entsprechende Wertbeitrag des Vergleichsobjekts für ein Merkmal vom Wertbeitrag des Bewertungsobjekts für dieses Merkmal subtrahiert wird.

8.3.2 Aggregationsvariablen

Aggregationsvariablen werden definiert, um Teilmengen der Variablen zu ersetzen, indem alle ihre Wertbeiträge und Anpassungen in einer gemeinsamen Variablen zusammengefasst werden, die im Bericht verwendet wird. Beispielsweise können sowohl die ausgebaute als auch die nicht ausgebaute Wohnfläche in feinere Detailvariablen aufgeteilt worden sein:

1. Oberirdische genehmigte ausgebaute Fläche
2. Unterirdische genehmigte ausgebaute Fläche
3. Oberirdische nicht genehmigte ausgebaute Fläche
4. Unterirdische nicht genehmigte ausgebaute Fläche
5. Oberirdische nicht ausgebaute Fläche
6. Unterirdische nicht ausgebaute Fläche

Ja, es gibt Gegenden, in denen dies durchaus sinnvoll ist.

Die MARS-Regression regressiert auf alle diese Variablen und erzeugt ein Modell, das Beiträge und Anpassungen für sie liefert. Für einen Formularbericht müssen diese verschiedenen Beiträge jedoch vermutlich zu weniger Variablen kombiniert bzw. aggregiert werden. Die Einzelheiten der Aggregation sollten in einem Addendum angegeben werden. Ein häufigeres Beispiel wäre die Aggregation von Interaktionsanpassungen in eine der Anpassungen der primären Variablen. Beispielsweise könnte die Anpassung für GLA_Lotsize zu GLA oder zu LotSize aggregiert oder sogar im Verhältnis 50:50 auf beide aufgeteilt werden, wie durch das Aggregation-Arbeitsblatt festgelegt.

8.3.3 Spezielle Variablen & Notation

π_i	Tatsächlicher Nettoverkaufspreis für Objekt $i = 1, 2, \dots, n_p$; für das Bewertungsobjekt bezeichnet π_0 die Wertschlussfolgerung (siehe Beweis I)
α_i	Angepasster Nettoverkaufspreis für Objekt $i = 1, 2, \dots, n_p$
ρ_i	Durch MARS geschätzter Nettoverkaufspreis für Objekt $i = 0, 1, \dots, n_p$
β	Basiswert des Modells
$\varphi_{i,j}$	Wertbeitrag der Regressionsvariablen $j = 1, 2, \dots, n_r$ für Objekt $i = 0, 1, \dots, n_p$
$\delta_{i,j}$	Anpassung für Regressionsvariable $j = 1, 2, \dots, n_r$ für Objekt $i = 1, 2, \dots, n_p$
ϵ_i	Gesamtresiduum für Objekt $i = 0, 1, \dots, n_p$
$\epsilon_{i,c}$	Wertbeitrag der Residualkomponente c, $c = 1, 2, \dots, n_c$ für Objekt $i = 0, 1, \dots, n_p$
$\xi_{i,c}$	Anpassung für Residualkomponente c, $c = 1, 2, \dots, n_c$ für Objekt $i = 0, 1, \dots, n_p$

wobei

n_p = Anzahl der Vergleichsobjekte

n_r = Anzahl der Regressionsvariablen

n_c = Anzahl der Residualaufschlüsselungsvariablen

8.3.4 Aufschlüsselung der Residuen

Sobald die MARS-Regression ein Modell erstellt hat, erzeugt sie aus den von MARS ausgewählten Eingabevariablen die geschätzten Verkaufspreise. Dieser Wert wird vom entsprechenden Nettoverkaufspreis subtrahiert, um für jedes Objekt einen Residualwert zu erhalten. Das RCA-Protokoll schreibt nun vor, dass der VE, sobald er die endgültigen Vergleichsobjekte festgelegt hat —

üblicherweise 6-12 Objekt-Listings —, jedes Listing durchgehen und das berechnete Residuum in diejenigen Merkmale (latente Variablen) aufschlüsseln muss, die seiner Einschätzung nach den Wert am wahrscheinlichsten beeinflussen, wobei die Möglichkeit zufälliger Datenfehler außer Acht gelassen wird. Für ein gegebenes Objekt könnten beispielsweise folgende Aufschlüsselungsvariablen vorliegen:

1. Design
2. Funktionale Nutzbarkeit
3. Verkauf über/unter Marktniveau
4. Zustand
5. Qualität

Verschiedene Objekte weisen vermutlich teilweise unterschiedliche Residualaufschlüsselungsvariablen auf, die in einer geordneten Menge zusammengeführt werden müssen.

Schätzung der Wertbeiträge seltener Merkmale. Einige wertrelevante Merkmale sind zu selten für den Regressionskanal: Eine Werkstatt, die in 3 von 780 Verkäufen vorhanden ist, kann keinen belastbaren Koeffizienten erlangen, und der Vorwärtsdurchlauf wird eine solche Variable in der Regel vollständig übergehen. Dies sind Residualmerkmale, und ihre Beiträge können *im Residualraum* geschätzt werden: Man passt das Modell ohne das Merkmal an und regressiert anschließend die Modellresiduen mittels gewöhnlicher Kleinst-Quadrate-Schätzung auf die ausgeschlossene(n) Spalte(n). Für ein binäres Flag ist der Koeffizient ein Pauschalbeitrag (äquivalent: der mittlere Residual-Offset der gekennzeichneten Verkäufe); für eine numerische Spalte ist er ein Beitrag je Einheit. Werden mehrere korrelierte Ausstattungsmerkmale *gemeinsam* geschätzt (z. B. die Fläche von Nebengebäuden zusammen mit einem Werkstatt-Flag), teilt die Regression den Wert zwischen ihnen auf, statt dieselben Dollarbeträge in zwei Aufschlüsselungszeilen doppelt zu zählen. Schätzungen, die auf weniger als etwa zehn Beobachtungen beruhen, sind lediglich indikativ und sollten durch abgeschriebene Herstellungskosten oder eine erweiterte Stichprobe (z. B. aus öffentlichen Anmerkungen gewonnene Flags) untermauert werden; die Residualrestriktion selbst begrenzt den letztlich zugewiesenen Wert. earthUI (v0.11+) implementiert diese Methode direkt als Reiter *Residual Contribution*, einschließlich der gemeinsamen Schätzung, der Residualevidenz je Verkauf und eines Dokumentationsvermerks für die Arbeitsakte.

Notation: Unter Verwendung der Notation aus obiger Tabelle ergibt sich:

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

wobei n_c die Anzahl der Residualaufschlüsselungskomponenten ist.

Der VE muss das Residuum nach eigenem Urteil in die gegebenen Komponenten aufschlüsseln. Man könnte meinen, dies erlaube das Eindringen von Verzerrungen in die endgültige Wertschlussfolgerung. Dies ist jedoch nicht der Fall, denn es liegt der mathematische Beweis vor, dass die Aufschlüsselungsanpassungen, sofern sie sich zum Residuum summieren, die Wertschlussfolgerung nicht beeinflussen. Zweck der Aufschlüsselung ist nicht die Wertschlussfolgerung, sondern die Erklärung, warum das Residuum im Vergleich zu den Residuen anderer Objektverkäufe so ausfällt, wie es ausfällt. Die einzige Stelle, an der die Verzerrung des Ingenieurs die endgültige Wertschlussfolgerung beeinflussen kann, ist die Schätzung des Residuums für das Bewertungsobjekt, da es für das Bewertungsobjekt naturgemäß keinen tatsächlichen Verkaufspreis gibt, der zur Berechnung eines Residuums erforderlich wäre.

Jedes Objekt besitzt seine eigene Menge von Residual-Wertbeiträgen; diese müssen jedoch über alle Objekte hinweg zu einer Menge C zusammengeführt werden:

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

wobei $[\epsilon_{i,c}]$ der Variablenname für den Wert $\epsilon_{i,c}$ ist, A_i die Menge solcher Variablen bezeichnet und n_c die Anzahl der Werte in jedem ϵ_i ist.

Stehen also die Daten des Bewertungsobjekts und der Vergleichsobjekte in getrennten Spalten, so sollte unterhalb aller Wertbeiträge der Regressionsmodellvariablen für jedes Objekt die Menge C der nc entsprechenden Residual-Wertbeiträge zu sehen sein.

9 Zentrale RCA-Beweise

Es seien:

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

Beweis I: Jeder einzelne angepasste Verkaufspreis eines Vergleichsobjekts ist gleich dem geschätzten Verkaufspreis des Bewertungsobjekts.

Für ein beliebiges Vergleichsobjekt i ist der angepasste Verkaufspreis α_i die Summe aus dem Netto-Verkaufspreis π_i und den Anpassungen: den Differenzen zwischen den Wertbeiträgen des Bewertungsobjekts und des Vergleichsobjekts für den Basiswert, die Regressionsvariablen $j = 1, 2, \dots, n_r$ und das Residuum. Für das Bewertungsobjekt, für das kein tatsächlicher Verkauf vorliegt, wird der Netto-Verkaufspreis π_0 als das Wertschlussfolgerung $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$ definiert. Dann gilt:

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

Beweis II: Eine Veränderung der Residuenzerlegung unter der Restriktion, dass sämtliche Wertbeiträge der Residualkomponenten in ihrer Summe das Residuum des Vergleichsobjekts ergeben, hat keinen Einfluss auf das endgültige Wertschlussfolgerung des RCA-Protokolls.

Dies ist nahezu trivial. Ohne Betrachtung der Mathematik wird es jedoch leicht übersehen, da traditionelle Gutachter ihr Augenmerk auf die Anpassungen statt auf die Wertbeiträge richten. Es

sind die Wertbeiträge, die sich hier gegenseitig aufheben — ein jahrzehntelanges Versäumnis. Für ein beliebiges Vergleichsobjekt k , $k > 0$, ist die gesamte Residuumsanpassung gleich der Summe der n_c Residualkomponenten der Zerlegung, die ihrerseits gleich der Summe der Differenzen der Wertbeiträge ist, welche wiederum gleich dem Gesamtresiduum des Bewertungsobjekts abzüglich des Gesamtresiduums des Vergleichsobjekts ist. Und dies ist unabhängig davon, wie viele Residualkomponenten gebildet werden oder was sie repräsentieren. Solange der Wert der Residualkomponenten sowohl für das Bewertungsobjekt als auch für das gewählte Vergleichsobjekt dem Gesamtresiduum entspricht, bleibt ξ_k unverändert.

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

10 MARS-Modell

10.1 „g“-Funktionen für Modellvariablen und Interaktionen

Ein Grundprinzip der RCA-Methode besteht darin, dass der „erklärenden Argumentation“ höchste Bedeutung zukommt. Leider kann die Bereitstellung aussagekräftiger Grafiken für die verschiedenen Wertbeitragsgleichungen, insbesondere wenn Interaktionen und Faktorvariablen verwendet werden, eine Herausforderung darstellen. Es wird daher als unerlässlich angesehen, ein klares und striktes Protokoll für die Erstellung dieser Grafiken bereitzustellen. Dies geschieht hiermit; es sei jedoch angemerkt, dass sich das Protokoll an den Entwickler richtet, der den R- oder Python-Code zur Erzeugung der Grafiken schreibt, und für Personen, die nicht mit der Programmierung befasst sind, möglicherweise nur von oberflächlichem Interesse ist.

Das von MARS erzeugte Preismodell ist eine Summe aus einer Basiskonstanten und zusätzlichen Termen, die jeweils 1 bis 3 Variablen umfassen. Zur Analyse wird das Modell zunächst in einzelne Terme zerlegt. Anschließend werden mehrere Terme mit derselben Variablenmenge, die eine einzelne Interaktion repräsentieren, zu einzelnen Interaktionsausdrücken zusammengefasst. Es sei angenommen, dass der maximale Regressionsgrad 3 beträgt. Betrachtet wird ein Regressionsmodell dritten Grades, bestehend aus einer Basiskonstanten, n_1 Ausdrücken ersten Grades, n_2 Interaktionsausdrücken zweiten Grades und n_3 Interaktionsausdrücken dritten Grades. Daraus ergeben sich vier verschiedene Gruppen von Ausdrücken, bei denen es sich entweder um Einzelvariablen oder um Interaktionen mehrerer Variablen handelt und die jeweils in Abhängigkeit von zwei Faktoren unterschiedlich grafisch dargestellt werden: (1) ob sie eine Faktorvariable (eine kategoriale Variable) enthalten, und (2) wie viele eindeutige Variablen sie enthalten. Im Einzelnen:

- Ausdrücke ersten Grades (1 Variable) werden in 2D dargestellt.
- Ausdrücke zweiten Grades (2 Variablen) werden in 3D dargestellt.
- Ausdrücke dritten Grades (3 Variablen) werden als *Menge* von 3D-Grafiken dargestellt.
- Enthält ein Ausdruck eine Faktorvariable, so verringert jeder eindeutige Faktor die Dimensionalität der Grafik um eins (z. B. wird aus einer 3D-Grafik eine 2D-Grafik, wenn eine der Variablen ein Faktor ist).

Die Indizierung jeder Funktion g liefert die Informationen, die erforderlich sind, um zu bestimmen, wie sie grafisch darzustellen ist, indem sie die Anzahl der eindeutigen Variablen und das Vorhandensein von Faktorvariablen widerspiegelt.

10.2 Indizierung der g -Funktionen

Jede Funktion g wird als $f_{j,k} g_k^j$ indiziert, wobei gilt:

- $f_{j,k}$ ist der erste Index (ein hochgestellter Index vor g):
 - j läuft von 0 bis 3 und bezeichnet die Gruppe:
 - * $j = 0$: die Basiskonstante (eine Ausnahme von der üblichen Indizierung).
 - * $j = 1$: Ausdrücke ersten Grades.
 - * $j = 2$: Ausdrücke zweiten Grades.
 - * $j = 3$: Ausdrücke dritten Grades.

- k ist der Index einer bestimmten Funktion innerhalb ihrer Gruppe (z. B. $k = 1$ für die erste Funktion in Gruppe j).
- Der Wert von $f_{j,k}$ gibt die Anzahl der eindeutigen Faktorvariablen in der Funktion an:
 - * $f_{j,k} = 0$: keine Faktorvariablen.
 - * $f_{j,k} > 0$: die Anzahl der Faktorvariablen (z. B. 1, 2 oder 3).
- Nachgestellte Indizes an g :
 - Der hochgestellte Index j nach g (z. B. g^j) wiederholt zur Verdeutlichung die Gruppennummer.
 - Der tiefgestellte Index k nach g (z. B. g_k) entspricht dem k in $f_{j,k}$ und identifiziert die Funktion innerhalb ihrer Gruppe.

Obwohl die Indizes von f und g numerisch übereinstimmen, unterscheiden sich ihre Rollen: $f_{j,k}$ zählt die Faktorvariablen, was für die Entscheidungen zur grafischen Darstellung entscheidend ist, während j und k an g dessen Gruppe und Position festlegen.

10.3 Dimensionen der grafischen Darstellung

Die Dimension d der Grafik für eine Funktion $f_{j,k} g_k^j$ berechnet sich als:

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

wobei gilt:

- j ist die Anzahl der eindeutigen Variablen (bzw. die Gruppennummer, wenn man den Fall der Konstanten außer Acht lässt).
- $f_{j,k}$ ist die Anzahl der Faktorvariablen.
- d ist die Anzahl der stetigen (Nicht-Faktor-)Variablen, die grafisch darzustellen sind, mit einem Maximum von 3 (da der maximale Grad 3 beträgt). Eine Funktion mit d stetigen Variablen wird in $d + 1$ Dimensionen gezeichnet: $d = 1$ ergibt eine 2D-Grafik, $d = 2$ eine 3D-Grafik (oder eine 2D-Heatmap) und $d = 3$ eine *Menge* von 3D-Grafiken.

Beispiele:

- Ist $j = 3$ (3 Variablen) und $f_{j,k} = 0$ (keine Faktorvariablen), so ist $d = 3 - 0 = 3$, was eine Menge von 3D-Grafiken erfordert, bei denen jeweils eine der Variablen festgehalten wird.
- Ist $j = 3$ und $f_{j,k} = 1$ (eine Faktorvariable), so ist $d = 3 - 1 = 2$, was für jeden im Modell auftretenden Wert der Faktorvariablen eine 3D-Grafik (oder eine 2D-Heatmap) erfordert.
- Ist $j = 0$ (die Konstante), so ist in der Regel keine grafische Darstellung erforderlich, da es sich um einen einzelnen Wert handelt.

Bei $d = 3$ (z. B. ein Ausdruck dritten Grades ohne Faktorvariablen) kann die Visualisierung der vollständigen Funktion mehrere 3D-Grafiken erfordern, bei denen jeweils eine Variable auf verschiedene Werte festgesetzt wird, um eine Teilmenge des Funktionsverhaltens darzustellen.

10.4 Die endgültige Modellgleichung

Das MARS-Modell wird in vier Gruppen von Ausdrücken zerlegt (Konstante sowie Ausdrücke ersten, zweiten und dritten Grades). Die Indizierung $f_{j,k} g_k^j$ gibt an:

- j : die Gruppe und die Anzahl der Variablen.
- $f_{j,k}$: die Anzahl der Faktorvariablen.
- $d = j - f_{j,k}$: die Dimension der Grafik.

Diese Struktur bestimmt, wie jede g -Funktion visualisiert wird, und stellt sicher, dass die Darstellungsmethode der Komplexität des Ausdrucks und den Variablentypen entspricht.

Die vom MARS-Modell gelieferte Verkaufspreisschätzung $\hat{P}$ lässt sich ausdrücken als:

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

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

10.5 Anzahl der zur grafischen Darstellung eines Modells benötigten Diagramme

Wie viele Grafiken sind erforderlich, um alle Funktionen eines Modells zu visualisieren? Die Antwort hängt von den Variablen, ihren Interaktionen und davon ab, ob es sich um Faktor- oder Nicht-Faktor-Variablen handelt.

Der VE kann in der Excel-Konfigurationsdatei für R/earth die Variablen und Interaktionen einschränken und damit beeinflussen, welche Interaktionen zulässig sind. Ein praktikabler Ansatz besteht darin, R/earth alle möglichen Interaktionen bewerten zu lassen und jene mit signifikanten Beiträgen zur Verkaufspreisschätzung auszuwählen, wobei auf Probleme wie Überanpassung (Over-Fitting) oder Kollinearität zu prüfen ist. Einige Interaktionen können ausgeschlossen werden, was sich auf Interaktionen höheren Grades auswirkt (wird z.B. ein Paar in einer Interaktion dritten Grades ausgeschlossen, so ist die gesamte Kombination ausgeschlossen).

Bei Faktorvariablen (z. B. MLS Area mit 12 Werten in Interaktion mit GLA) wird die grafische Darstellung komplexer. Dies könnte 12 Grafiken erfordern—eine pro Gebiet. Interagiert ein weiterer Faktor wie Architectural Style (6 Werte) mit MLS Area und GLA, könnten bis zu $12 \cdot 6 = 60$ Grafiken erforderlich sein. R/earth stuft jedoch selten alle Kombinationen als signifikant ein; typischerweise rechtfertigen nur wenige zentrale Interaktionen (z. B. bestimmte Gebiete und Stile) eine grafische Darstellung, wie sie das Modell identifiziert.

Es folgt eine vereinfachte Übersicht über die Anzahl der benötigten Grafiken:

1. **Funktionen einer Variablen (Nicht-Faktor):** 1 Grafik.
2. **Funktionen einer Variablen (Faktor):** 1 Balkendiagramm, das alle Faktorwerte zusammenfasst.
3. **Basisfunktion:** Keine Grafik erforderlich.

4. **Interaktionen zweiten Grades (keine Faktoren):** 1 Grafik (z. B. 2D-Heatmap mit Farben für den Verkaufspreis).
5. **Interaktionen zweiten Grades (1 Faktor):** 1 Grafik pro im Modell auftretendem Faktorwert (z. B. 2 Werte = 2 Grafiken).
6. **Interaktionen zweiten Grades (2 Faktoren):** 1 Grafik mit einer Matrix der Faktorkombinationen (z. B. 3 Gebiete $\times$ 2 Stile = 6 Zellen).
7. **Interaktionen dritten Grades (keine Faktoren):** Mehrere Grafiken (z. B. 4 Grafiken für Bath count = 1, 2, 3, 4).
8. **Interaktionen dritten Grades (1 Faktor):** 1 Grafik pro im Modell auftretendem Faktorwert.
9. **Interaktionen dritten Grades (2 Faktoren):** 1 Grafik pro im Modell auftretender Kombination von Faktorwerten.
10. **Interaktionen dritten Grades (3 Faktoren):** 1 Grafik pro Wert der Faktorvariablen mit den wenigsten unterschiedlichen Werten, unter Verwendung einer Farbmatrix für die übrigen beiden.

Komplexere Szenarien können fortgeschrittenere Darstellungstechniken erfordern.

Es bestehen weitere Möglichkeiten mit voraussichtlich anspruchsvolleren Grafiken.

Ein Beispiel für ein vollständiges, von MARS erzeugtes Modell findet sich in [Figure 1](#).

Abbildung 1: Komponenten des MARS-Modells

Basis	$= {}^0g_1^0 = 4957150$
Age	$= {}^0g_1^1 = -3163.1809255 \cdot \text{Age}$
AreaNbr	$= {}^1g_2^1 = -623197.75104 \cdot I\{\text{AreaNbr} = 471\}$
Baths	$= {}^0g_3^1 = -111023.51496 \cdot \max(0, 2 - \text{Baths}) + 96735.52418 \cdot \max(0, \text{Baths} - 2)$
Beds	$= {}^0g_4^1 = -233489.66415 \cdot \max(0, \text{Beds} - 5)$
DaysOffMkt	$= {}^0g_5^1 = -2577.2188622 \cdot \max(0, 189 - \text{DaysOffMkt}) - 1282.8789534 \cdot \max(0, \text{DaysOffMkt} - 189)$
GLA	$= {}^0g_6^1 = -666.73587914 \cdot \max(0, 4027 - \text{GLA}) + 281.36312913 \cdot \max(0, \text{GLA} - 4027)$
Latitude	$= {}^0g_7^1 = -18141680.417 \cdot \max(0, \text{Latitude} - 37.573)$
LotSize	$= {}^0g_8^1 = -85.005683404 \cdot \max(0, 8000 - \text{LotSize}) - 23.562970183 \cdot \max(0, \text{LotSize} - 8000)$
Age_Baths	$= {}^0g_1^2 = -3916.3594872 \cdot \text{Age} \cdot \max(0, \text{Baths} - 3.5)$
Age_Garage	$= {}^0g_2^2 = -6503.1897339 \cdot \text{Age} \cdot \max(0, \text{Garage} - 2)$
AreaNbr_LotSize	$= {}^1g_3^2 = 78.54207825 \cdot I\{\text{AreaNbr} = 463\} \cdot \max(0, 8000 - \text{LotSize})$ $+ 87.263298012 \cdot I\{\text{AreaNbr} = 466\} \cdot \max(0, \text{LotSize} - 8000)$
AreaNbr_Baths	$= {}^1g_4^2 = -1070222.5995 \cdot I\{\text{AreaNbr} = 464\} \cdot \max(0, \text{Baths} - 2)$
AreaNbr_Latitude	$= {}^1g_5^2 = 17571072.548 \cdot I\{\text{AreaNbr} = 466\} \cdot \max(0, \text{Latitude} - 37.573)$
AreaNbr_GLA	$= {}^1g_6^2 = -328.79173933 \cdot I\{\text{AreaNbr} = 471\} \cdot \max(0, \text{GLA} - 2550)$ $+ 525.69503229 \cdot I\{\text{AreaNbr} = 471\} \cdot \max(0, 2550 - \text{GLA})$
DaysOffMkt_GLA	$= {}^0g_7^2 = 0.43839582286 \cdot \max(0, \text{DaysOffMkt} - 661) \cdot \max(0, 4027 - \text{GLA})$
FrPlcNbr_LotSize	$= {}^0g_8^2 = 22.059059591 \cdot \max(0, 2 - \text{FrPlcNbr}) \cdot \max(0, \text{LotSize} - 8000)$ $+ 39.084508159 \cdot \max(0, \text{FrPlcNbr} - 2) \cdot \max(0, \text{LotSize} - 8000)$
GLA_LotSize	$= {}^0g_9^2 = -0.45274160024 \cdot \max(0, \text{GLA} - 2994) \cdot \max(0, 8000 - \text{LotSize})$
ADU_Age_Garage	$= {}^0g_1^3 = -5.4578633919 \cdot \max(0, \text{ADU} - 0) \cdot \text{Age} \cdot \max(0, 2 - \text{Garage})$
AreaNbr_Baths_LotSize	$= {}^1g_2^3 = 60.607126605 \cdot I\{\text{AreaNbr} = 460\} \cdot \max(0, \text{Baths} - 2) \cdot \max(0, \text{LotSize} - 2480)$
AreaNbr_Baths_FrPlcNbr	$= {}^1g_3^3 = 593779.41722 \cdot I\{\text{AreaNbr} = 464\} \cdot \max(0, \text{Baths} - 2) \cdot \max(0, \text{FrPlcNbr} - 0)$
AreaNbr_GLA_Latitude	$= {}^1g_4^3 = 54199.925713 \cdot I\{\text{AreaNbr} = 464\} \cdot \max(0, \text{GLA} - 2670) \cdot \max(0, \text{Latitude} - 37.573)$ $+ 58934.565747 \cdot I\{\text{AreaNbr} = 471\} \cdot \max(0, \text{GLA} - 2550) \cdot \max(0, \text{Latitude} - 37.568)$
AreaNbr_DaysOffMkt_GLA	$= {}^1g_5^3 = 0.37173306643 \cdot I\{\text{AreaNbr} = 470\} \cdot \max(0, 661 - \text{DaysOffMkt}) \cdot \max(0, 4027 - \text{GLA})$
Beds_GLA_LotSize	$= {}^0g_6^3 = -5.1607264026 \cdot \max(0, 4 - \text{Beds}) \cdot \max(0, \text{GLA} - 2994) \cdot \max(0, 8000 - \text{LotSize})$
Age_AreaNbr_Baths	$= {}^1g_7^3 = -10759.444688 \cdot \text{Age} \cdot I\{\text{AreaNbr} = 460\} \cdot \max(0, \text{Baths} - 3.5)$

11 Datencharakteristika

Der folgende Abschnitt behandelt MARS. MARS ist vollständig von den Daten abhängig, die es erhält, und der Großteil dieser Daten stammt aus einem MLS (Multiple Listing Service) oder von Steuerveranlagungsbehörden (Tax Assessors). Die Daten können jedoch über andere Datenanbieter geleitet werden, die sie pflegen und verbessern und anschließend mit Gewinn verkaufen. In einigen Gebieten sind sie recht zuverlässig, in anderen weniger. In kritischen Fällen muss der Objektbesichtiger eigene Beobachtungen und Messungen vornehmen, deren Genauigkeit infrage gestellt werden kann.

Der Valuation Engineer muss verstehen, was die Daten repräsentieren und wie gut sie das repräsentieren, was sie repräsentieren sollen. Daraus ergibt sich eine ganze Reihe von Fragen:

- Wie lässt sich eine Variable als Messung klassifizieren?
- Wenn eine Variable keine Messung ist, was kann sie sonst sein?
- Wie kann Regression nützlich sein, wenn die erhaltenen Daten nicht genau sind?

Bei den meisten Daten, die für die Regression aus MLS-Diensten und von Steuerveranlagungsbehörden bezogen werden, handelt es sich um numerische Daten, die Messungen oder Zählungen implizieren. Ein Teil davon ist logischer Natur, wahr oder falsch (True/False).

11.1 Zweck des Residuums

Das Residuum dient als indirekter Maßstab für den Wertbeitrag derjenigen Variablen, die nicht im Regressionsmodell enthalten sind. In der Sprache der hedonischen Literatur sind diese nicht gemessenen Merkmale *latente Variablen*: Variablen, die den Verkaufspreis beeinflussen, aber nicht direkt gemessen werden, in den meisten Fällen, weil es keine praktikable Möglichkeit gibt, sie objektiv zu messen. Ihre Werte müssen stattdessen erschlossen werden, und in der Praxis wird der Wert einer latenten Variablen häufig durch subjektives Urteil geschätzt — im RCA durch ein Urteil, das im Residuen-Ranking verankert ist. Durch die Ordnung der Objekte nach ihrem Residuum bzw. ihrem Residuum pro Quadratfuß wird ein objektives Ranking von den Objekten mit der geringsten bis zu jenen mit der größten Attraktivität erreicht. Diese Ranking-Methode ist wirksam, sofern ein erfahrener Analyst mit tiefem Verständnis des lokalen Marktes das MARS-Modell (Multivariate Adaptive Regression Splines) erstellt. Darüber hinaus gelten die folgenden zusätzlichen Anforderungen:

- Qualität der Daten und des Modells: Ein gut konstruiertes Regressionsmodell und qualitativ hochwertige Daten sind Voraussetzungen für verlässliche Residuen. In der San Francisco Bay Area besteht das Ziel des Verfassers darin, ein Modell mit einem R^2 von etwa 75-80% und einem Kreuzvalidierungs- R^2 (CV R^2) von 55-70% zu entwickeln.
- Überanpassung vermeiden, das R^2 begrenzen: Der Valuation Engineer sollte in den meisten reifen Märkten in hochpreisigen Gebieten wie der SF Bay Area ein maximales R^2 von etwa 80% anstreben, da ein Großteil des Wettbewerbswerts einer Wohnimmobilie in ihrer subjektiven Attraktivität liegt, die nicht quantifizierbar ist und daher nicht direkt in Regressionsmodelle einbezogen werden kann. Ein R^2 deutlich über 80% kann auf ein überangepasstes Modell hindeuten und sollte Anlass zu weiterer Prüfung geben. Ein warnender Hinweis jedoch: Es gibt Marktgebiete mit R^2 -Werten über 90% ohne Überanpassung, und es gibt schwer zu bewertende Gebiete, in denen bestenfalls 50-60% oder weniger erreichbar sind.

- Hohes R^2 in bestimmten Gebieten: Es gibt allerdings Regionen, etwa Neubausiedlungen (Subdivisions) oder andere einheitliche Wohngebiete, in denen höhere R^2 -Werte aufgrund der Homogenität der Objekte überraschend häufig auftreten.

11.2 Zustands- und Qualitätsbewertungen als approximative Regressoren

Die Bewertungsskalen von Fannie Mae für Zustand und Qualität (C1–C6 und Q1–Q6) verdienen eine besondere Anmerkung. Diese Skalen sind zu grob, um als echte gemessene Merkmale behandelt zu werden, und da sie vom Gutachter vergeben werden, handelt es sich in der Praxis um subjektive Bewertungen — ihrerseits latente Variablen. Es ist nützlich, zwei Ebenen ausdrücklich zu unterscheiden: die grobe Bewertung des Gutachters, die als Regressor verwendet wird, *condition-coarse-grain* (condition-1) und *quality-coarse-grain* (quality-1), gegenüber den feinkörnigen latenten Variablen, die aus dem Residuum gewonnen werden, *condition-fine-grain* (condition-2) und *quality-fine-grain* (quality-2). Der VE kann die grobkörnigen Bewertungen des Gutachters als Näherungen akzeptieren und auf sie regressieren, wodurch approximative Wertbeiträge gewonnen werden, die als anfängliche Wertzuweisungen für die feinkörnigen latenten Variablen nützlich sind. Wo dies geschieht, sieht das Protokoll vor, die regressionsgeschätzten Wertbeiträge dieser beiden grobkörnigen Variablen von der Modellschätzung abzuziehen — sie also aus den gemessenen Attributen im Regressionsabschnitt des Sales Grid herauszunehmen — und sie dem Residuum wieder hinzuzufügen. Von dort werden sie anschließend aus dem Residuum entfernt und den entsprechenden feinkörnigen latenten Variablen `res_condition1` und `res_quality1` zugewiesen. Zusätzliche Informationen, die aus den öffentlichen MLS-Anmerkungen (`public_remarks`) und zugehörigen Zeichenkettenlisten extrahiert werden, dienen anschließend Mikroverfeinerungen dieser latenten Variablenwerte. Das Endergebnis ist, dass die feinkörnigen Wertbeiträge die grobkörnigen Wertbeiträge ersetzen: Die groben Bewertungen dienen lediglich als Gerüst für das approximative Modell, und die letztlich für Zustand und Qualität ausgewiesenen Werte sind die feinkörnigen, aus dem Residuum abgeleiteten.

11.3 Datenfehler

Ausweislich von Kommentaren in sozialen Medien sind die Daten des kalifornischen Multiple Listing Service (MLS) im Allgemeinen von höherer Qualität als in vielen anderen Regionen der Vereinigten Staaten. Typischerweise ist in den meisten reifen Metropolregionen mit MLS-Daten guter Qualität zu rechnen, in ländlichen Gebieten hingegen mit schlechterer.

Zwar treten aus verschiedenen Gründen gelegentlich Fehler in den Daten auf, doch sollten sich diese Fehler im Allgemeinen gegenseitig aufheben und bei ausreichender Datenmenge nur minimale Auswirkungen auf die in Preismodellen verwendeten Mittelwerte haben.

In bestimmten Gebieten können jedoch systematische Verzerrungen auftreten. Man betrachte beispielsweise eine Nachbarschaft, in der viele Häuser ursprünglich mit vollständigen zweiten Geschossen errichtet wurden, einigen Käufern jedoch die Option eingeräumt wurde, Räume im Obergeschoss zu entfernen, um in einem oder mehreren Räumen des Erdgeschosses gewölbte Decken zu schaffen, wodurch sich die Bruttowohnfläche (gross living area, GLA) erheblich verringerte. Diese Änderungen wurden aus dem einen oder anderen Grund nicht in den Unterlagen der Steuerveranlagungsbehörde nachgeführt. Folglich weist das MLS möglicherweise 2,400 Quadratfuß aus, während der gemessene Wert nur 2,000 Quadratfuß beträgt. In solchen Szenarien sollten die MLS-Daten vor der Durchführung der MARS-Regression (Multivariate Adaptive Regression Splines) korrigiert werden. Zu beachten ist, dass die Beschaffung dieser Informationen für Vergleichsobjekte einen gewissen Ermittlungsaufwand erfordern kann. Abweichungen sollten für spätere Zwecke dokumentiert werden.

Angesichts des Vorstehenden ist die Annahme berechtigt, dass in Gebieten mit hochwertigen MLS-Daten die Fehler, die in das Residuum durchschlagen, im Vergleich zum Einfluss subjektiver und nicht gemessener Merkmale wie Design, Zustand und Qualität minimal sind. Dies ist eine wichtige Überlegung bei der Schätzung der Genauigkeit der RCA-Methode. Beträgt beispielsweise der R^2 -Wert für ein gegebenes Marktgebiet 80%, so entfallen etwa 20% des Preises auf das CQA-Residuum. Ein Fehler von $\pm 10\%$ bei der Einordnung des Bewertungsobjekts im Residuen-Ranking pflanzt sich daher als Fehler von $\pm 10\% \cdot 20\% = \pm 2\%$ in der endgültigen Wertschlussfolgerung fort. Auf diese Weise lässt sich die Genauigkeit steigern! In vielen Fällen sind genaue Wertschätzungen deutlich unter $\pm 1\%$ möglich.

Es ist daher vertretbar, Residuen als indirekte Indikatoren für den kollektiven Wert nicht modellierter Variablen in der MARS-Regression zu behandeln, sofern der Valuation Engineer über ausreichende lokale Markterfahrung verfügt, um zu erkennen, wo und warum systemische Messverzerrungen bestehen. Dieses Vorgehen stellt sicher, dass die Analyse die tatsächlichen Objektwerte zutreffend widerspiegelt und dabei die Grenzen und Feinheiten von Immobiliendaten berücksichtigt.

Ausgearbeitetes Beispiel: Messung nicht gemessener Werttreiber in Sea Ranch

Wenn das in einem Marktgebiet erreichbare R^2 niedriger ausfällt, als es die scheinbare Homogenität des Marktes erwarten ließe, besteht die Antwort des Rahmenwerks darin, die fehlende Variable zu identifizieren und zu messen, statt die Einschränkung hinzunehmen. Die Gemeinde Sea Ranch im Sonoma County veranschaulicht dies unmittelbar.

Ein konkretes Beispiel für die Maxime, zu messen statt hinzunehmen. Die Gemeinde Sea Ranch im Sonoma County ist eine Küstensiedlung, in der die Entfernung zum Ozean ein primärer Werttreiber ist: Ein Haus, das mehrere hundert Fuß von der Steilküste entfernt liegt, erzielt einen erheblich anderen Preis als eines eine Viertelmeile landeinwärts, bei ansonsten gleichen Merkmalen. Die MLS-Daten der Gemeinde erfassen die Ozeanentfernung nicht als Variable. Ein erstes MARS-Modell, das auf den verfügbaren MLS-Variablen angepasst wurde, lieferte eine geringere Erklärungskraft, als die scheinbare Homogenität des Marktes hätte stützen sollen, mit einer Residualstruktur, die die Objekte nach keinem in den Daten sichtbaren Merkmal hinreichend ordnete — ein Signal dafür, dass ein wichtiger Treiber im Modell fehlte, und nicht dafür, dass der Markt tatsächlich verrauscht war.

Die Abhilfe bestand in direkter Messung. Der Verfasser beschaffte eine Gemeindegkarte, die die Lage der Grundstücke relativ zur Küstenlinie zeigt, und maß die Ozeanentfernung für mehrere hundert Häuser. Die gemessenen Entfernungen wurden dem MLS-Datensatz als Variable hinzugefügt, und die MARS-Regression wurde erneut durchgeführt. Die neue Variable wurde vom Modell als signifikant ausgewählt, die Erklärungskraft stieg erheblich, und das Residuen-Ranking ordnete die Objekte fortan nach Mustern, die mit den verbleibenden subjektiven Merkmalen — Aussicht, Zustand, Qualität — konsistent waren, statt nach dem latenten Ozeanentfernungssignal, das es zuvor kontaminiert hatte. Die Höhenlage wurde in Hanglagen-Nachbarschaften, in denen die MLS-Variable fehlt oder unzuverlässig ist, in ähnlicher Weise behandelt.

Das methodische Prinzip ist allgemein. Fällt das erreichbare R^2 unter das, was die scheinbare Homogenität des Marktes stützen sollte, lautet die erste Hypothese, dass eine wichtige Variable im Modell fehlt, und die angemessene Reaktion besteht darin, zu identifizieren, um welche es sich handelt, und die Messungen direkt zu erheben. Zu den Quellen zählen Gemeindegkarten, GIS-Daten, Unterlagen der Steuerveranlagungsbehörden, Unterlagen der Bauämter, Satellitenbilder und Ortsbesichtigungen. Die Kosten solcher Messarbeiten sind real, aber begrenzt; ist die Variable für ein bestimmtes Marktgebiet einmal erhoben, kann sie in künftige Aufträge in diesem Gebiet übernommen werden. Der RCA nimmt neue Variablen problemlos auf, da das Rahmenwerk

indifferent gegenüber der Frage ist, welche Variablen in die MARS-Regression eingehen — die Residualstruktur zeigt dem VE, ob die neue Variable echtes Signal erfasst, und die Wertbeiträge passen sich entsprechend an.

Die alternative Reaktion — ein niedriges R^2 zu akzeptieren und breitere Fehlerbänder auszuweisen — ist vertretbar, wenn die fehlenden Treiber tatsächlich nicht messbar sind (käufer-spezifische Präferenzen, vorübergehende Marktstimmung, idiosynkratische Transaktionsumstände), nicht jedoch, wenn sie lediglich nicht erhoben wurden. Der erste Schritt des VE bei einem schwachen Modell sollte die Frage sein, ob das fehlende Signal Letzteres ist.

11.4 Was ist Genauigkeit?

Vollkommene Genauigkeit einer Bewertung erfordert:

1. Übereinstimmung mit dem Marktwert: Die hypothetische Verkaufstransaktion des Bewertungsobjekts entspricht zu 100% der gegebenen Definition des Marktwerts
2. Genaue und robuste Modellentwicklung: Es wird ein hochwertiges MARS-Modell mit einem R^2 von etwa 75-80% und einem CVR2 von rund 55-70%, je nach Lage, entwickelt.
3. Gute Datenfehleranalyse: Alle wahrscheinlichen Datenfehler der gemessenen Variablen sind verstanden und eingegrenzt, mit geringem Übergreifen auf die Residuen.
4. Berücksichtigung systemischer Verzerrungen: Jede systemische Verzerrung in den Daten sollte berücksichtigt und bereinigt werden.
5. Residuen-Rang-Analyse: Es ist sicherzustellen, dass die Objekte im Residuen-Ranking vom niedrigsten zum höchsten Rang ein plausibles Merkmalsmuster aufweisen. Anomalien wie Nachlassverkäufe, Short Sales und Auktionsverkäufe sollten untersucht, erklärt und gegebenenfalls entfernt werden.
6. Platzierung des Bewertungsobjekts: Auf der Grundlage eines gründlichen Verständnisses des Residuen-Rankings sollte das Bewertungsobjekt genau positioniert werden. Es sollten ein Objekt mit deutlich höherer und ein Objekt mit deutlich geringerer Attraktivität identifiziert, das Bewertungsobjekt zwischen ihnen platziert und mit einem Score versehen werden. Anschließend ist zu versuchen, den Rangabstand zwischen dem höheren und dem niedrigeren Objekt so weit wie möglich einzuengen.
7. Erklärende Begründung: Die erklärende Begründung darf niemals der Genauigkeit geopfert werden. Der Gutachter muss in der Lage sein, die Wertunterschiede zwischen dem Bewertungsobjekt und den Vergleichsobjekten zu erklären, zumindest gegenüber einer Person mit gewöhnlicher Verständigkeit. Beispielsweise kann ein gutes MARS-Modell mit optimalem R^2 und CVR2 verwendet werden, um die Wertbeiträge gemessener Merkmale zu erklären, während die Aufschlüsselung des Residuums durch den Valuation Engineer zur Erklärung der Wertbeiträge der übrigen Merkmale dienen kann.
8. Überanpassung: Eine Überanpassung der Modelle ist zu vermeiden, indem keine Objektmerkmale verwendet werden, die einzeln oder in Kombination ein oder mehrere Objekte identifizieren können. Insbesondere sind Faktorvariablen zu vermeiden, deren Werte eine kleine Zahl von Objekten spezifizieren können. Eine Faktorvariable, die in umfangreichen Datensätzen Nachbarschaften identifiziert, ist angemessen; eine Variable, die Objekte mit Schwimmbecken in einem Gebiet identifiziert, in dem nur sehr wenige Schwimmbecken existieren, muss hingegen mit Vorsicht verwendet werden. Ein Merkmal, das einzigartige Objekte identifiziert, wie exakte GIS-Koordinaten, Flurstücksnummern oder Adressen, ist absolut unzulässig.

Beträgt das R^2 der Regression 80%, ist zu erwarten, dass etwa 20% des Werts des Bewertungsobjekts auf das Residuum entfallen. Dies gilt allerdings nur näherungsweise. Aus der CQA-Residualkurve ist bekannt, dass der Großteil der Verteilung der Residuen an den unteren und oberen Extremen der CQA liegt. Auf der Grundlage der CQA-Residualkurve und der extremen Steigung bei niedrigen und hohen CQA-Werten wäre in diesen Bereichen eine weniger genaue Schätzung des Residuums zu erwarten. Dem stehen jedoch die relative Seltenheit von Objekten in diesen Bereichen sowie der größere Attraktivitätsunterschied zwischen benachbarten Objekten entgegen. Während also im mittleren Bereich von der Auswirkung einer Spanne von $\pm 5\%$ gesprochen werden kann, wäre an den Extremen eher eine etwas engere Spanne von $\pm 2 - \pm 3\%$ maßgeblich.

Die Berechnung der Genauigkeit des RCA ist im Allgemeinen wegen der Nichtlinearität der CQA-Residualkurve etwas schwierig. Sie ist jedoch im Einzelfall möglich, sofern ein plausibles Residuen-Ranking vorliegt. Der Wertebereich der CQA-Scores reicht von 0.0 bis 10.0. Mit einem guten MARS-Modell sollte das Residuen-Ranking ein klares Muster von der geringsten bis zur höchsten Attraktivität zeigen, und es sollte nicht schwerfallen, eine Einstufung innerhalb von $\pm 5\%$ der Gesamtspanne bzw. ± 0.5 des CQA-Scores vorzunehmen. Bei 100 Objekten in der Regression entspräche dies einer genauen Einstufung des Bewertungsobjekts innerhalb einer Gruppe von 10 Objekten. Unter Bezugnahme auf die Burlingame-Verkaufsdaten in Tabelle 3 sei daher die approximative Genauigkeit (bzw. der Fehler) für einen vom VE geschätzten CQA-Wert von 7.5, 5.1 und 1.9 berechnet, wobei dessen Wertschlussfolgerung für das Bewertungsobjekt \$3,000,000 beträgt. Da Tabelle 3 nur eine Stichprobe der mehreren hundert verwendeten Objekte enthält, muss das Residuum, wo erforderlich, interpoliert werden (unter Annahme von Linearität). Zur Klarstellung: Es wird angenommen, dass der VE in jedem Fall sehr sicher ist, dass seine Einstufung des Bewertungsobjekts innerhalb einer Spanne von ± 5 Objekten im Residuen-Ranking von nur 100 Objekten liegt.

CQA= 7.5 ± 0.5					
CQA	Residuum	Abs. Diff.	Erw. Fehler	Hyp. Verkaufspreis	Proz. Fehler
8.00	\$257,112				
7.50	\$197,113	\$59,999			
7.00	\$154,615	\$42,498	$\pm \$51,249$	\$3,000,000	$\pm 1.71\%$

CQA=5.1 ± 0.5					
CQA	Residuum	Abs. Diff.	Erw. Fehler	Hyp. Verkaufspreis	Proz. Fehler
5.60	\$50,911				
5.10	\$19,559	\$31,352			
4.60	-\$28,207	\$47,766	$\pm \$39,559$	\$3,000,000	$\pm 1.32\%$

CQA= 1.9 ± 0.5					
CQA	Residuum	Abs. Diff.	Erw. Fehler	Hyp. Verkaufspreis	Proz. Fehler
2.40	-\$228,411				
1.90	-\$276,381	\$47,970			
1.40	-\$326,938	\$50,557	$\pm \$49,264$	\$3,000,000	$\pm 1.64\%$

Letztlich besteht der einzige objektive Weg zur Verifikation der Genauigkeit darin, ein striktes Protokoll festzulegen (das dieser Beitrag nur teilweise in groben Zügen andeutet) und dessen Einhaltung zu überprüfen. Dies ist erforderlich, weil tatsächliche Verkaufspreise aufgrund von Unvollkommenheiten des Marktes vom Marktwert abweichen können. Der Vergleich von Gutachten mit nachfolgenden Verkäufen kann gewisse Hinweise auf die Genauigkeit liefern. Die Komplexität

und Instabilität des Immobilienmarktes erfordern jedoch die Expertise von Gutachtern oder hochqualifizierten Valuation Engineer, um Wertschätzungen auf der Grundlage etablierter Standards und Protokolle abzugeben. Diese sollten streng genug sein, um sicherzustellen, dass kompetente Valuation Engineer unabhängig voneinander zu nahezu derselben Wertschlussfolgerung gelangen. Dies sollte in den meisten Fällen, in denen gute Daten verfügbar sind, ein erreichbares Ziel sein.

11.5 Das Bewertungsobjekt als Anomalie

Da die Schätzung des Marktwerts eines Objekts die Analyse aktueller Verkaufstransaktionen erfordert, ist die Frage berechtigt, ob aktuelle Verkaufstransaktionen die verschiedenen Merkmale eines älteren Bewertungsobjekts, dem Modernisierungen und Reparaturen fehlen, gut repräsentieren. Dies gilt insbesondere für die San Francisco Bay Area, wo es sich bei aktuellen Verkaufstransaktionen typischerweise um Häuser handelt, die zur Steigerung ihres Verkaufspreises modernisiert wurden. Die durchschnittliche Qualität und der durchschnittliche Zustand der Vergleichsobjekte sind daher im Allgemeinen besser als bei älteren Häusern, die seit geraumer Zeit nicht mehr auf dem Markt waren. Dieser Wertunterschied sollte durch die Schätzung des Residuums für das Bewertungsobjekt — mit anderen Worten: durch dessen CQA-Score — erfasst werden. Dies ist eine wichtige Überlegung im Hinblick auf Datenqualität und Genauigkeit bei der Bewertung des Bewertungsobjekts.

11.6 Die Bedeutung des CVR2 im Umgang mit unerwarteten Daten

Die MARS-Regression beruht auf mehreren Iterationen von R/earth, angewandt auf eine zufällig ausgewählte Teilmenge von Objekten, die typischerweise 80% bis 90% des MLS-Datensatzes umfasst und als Trainingsdatensatz bezeichnet wird. Die verbleibenden 10% - 20% der Objekte dienen als Testdatensatz. In dieser Phase wird das MARS-Modell aus den Trainingsdaten erstellt und anschließend verwendet, um die Verkaufspreise der Objekte im Testdatensatz zu prognostizieren, die dann mit den tatsächlichen Verkaufspreisen verglichen werden. Diese Methodik gilt unter Data-Mining-Fachleuten als die wirksamste Strategie zur Entwicklung eines robusten Modells, das die Verkaufspreise noch nicht beobachteter Objekte prognostizieren kann.

Der primäre Nutzen der CVR2-Kennzahl liegt jedoch nicht darin, die faktische Genauigkeit des Regressionsmodells hinsichtlich tatsächlicher Verkäufe im Marktgebiet abzubilden, sondern vielmehr in der Optimierung der Prognosefähigkeit des Modells für hypothetische Objektverkäufe zu einem festgelegten Bewertungsstichtag. Diese Objekte sind nicht im Trainingsdatensatz enthalten und können neue oder unerwartete Merkmalskombinationen aufweisen, die sowohl in den Objektdaten des Trainings- als auch des Testdatensatzes fehlen.

Wie lässt sich vor diesem Hintergrund der Verkaufspreis eines Objekts schätzen, das in den vergangenen 15 Jahren nicht modernisiert wurde, unter der Annahme, dass es beispielsweise einen Monat vor dem Bewertungsstichtag gelistet wurde? Ein Modell mit einem starken CVR2 liefert mit größerer Wahrscheinlichkeit die genaueste Schätzung als ein Modell, das diese wiederholte Kreuzvalidierung nicht durchlaufen hat. Ebenso wesentlich ist, dass der Datensatz eine ausreichende Zahl von Vergleichsobjekten mit unterschiedlichem Grad an Aktualisierungen, Reparaturen und Modernisierungen enthält.

Verfügt man letztlich über ein robustes Modell, das die Wertbeiträge der verschiedenen Objektmerkmale im Datensatz der Vergleichsobjekte genau prognostiziert, so lässt sich das neue Bewertungsobjekt wirksam in das nach Residuen geordnete Ranking dieser Objekte durch das Modell einordnen. Diese Platzierung erlaubt es, dem Objekt einen angemessenen CQA-Score und ein angemessenes Residuum zuzuweisen.

12 Residuen

Wie bereits der Name „Residual Constraint Approach“ erkennen lässt, kommt den Residuen in dieser Methodik eine zentrale Rolle zu. Als Residuum wird hier die Differenz zwischen dem Nettoverkaufspreis einer Immobilie und dem durch die Regressionsanalyse prognostizierten Verkaufspreis bezeichnet.

Ein positives Residuum deutet darauf hin, dass das Objekt zu einem höheren Preis verkauft wurde, als auf Grundlage typischer messbarer Merkmale wie der Bruttowohnfläche (GLA) und der Grundstücksgröße zu erwarten war, was auf eine überdurchschnittliche Attraktivität hindeuten kann. Ein negatives Residuum zeigt an, dass das Objekt zu einem geringeren Preis verkauft wurde, was auf eine geringere Attraktivität oder andere nicht erfasste negative Merkmale hindeuten kann.

Da größere Häuser vermutlich größere Residuen aufweisen, kann für einen stärker normalisierten Vergleich das Residuum pro Quadratfuß auf Basis der GLA herangezogen werden. Eine vertiefte Erörterung dieser Frage bleibt einer weiteren Arbeit vorbehalten.

12.1 Ranking & Scoring

Nach der Erstellung des MARS-Modells (Multivariate Adaptive Regression Splines) für die gemessenen Variablen werden die Residuen für jedes Objekt berechnet. Diese Residuen stellen die Differenz zwischen den beobachteten Nettoverkaufspreisen und den vom Modell prognostizierten Verkaufspreisen dar. Zusätzlich sollte eine Spalte für das Residual/SF angelegt werden, das heißt das Residuum dividiert durch die Wohnfläche des Wohngebäudes, wobei zu berücksichtigen ist, dass das Residuum einzelner Objekte in der Regel mit der Größe der Wohnfläche zunimmt. Andere mögliche Kennzahlen können in manchen Wohngebieten besser geeignet sein, sind jedoch ein Thema für eine spätere Gelegenheit.

Annahme 7: „Rangbildung nach Residuen“

In dieser Arbeit ist unter „Rangbildung nach Residuen“ eine Rangbildung nach dem Residuum, nach dem Residual/SF oder nach einer anderen Funktion der Residuen zu verstehen, je nachdem, was der Valuation Engineer für das jeweilige Marktgebiet als am geeignetsten erachtet. Die Erfahrung des Autors in der San Francisco Bay Area zeigt, dass das Residual/SF in der Regel am zuverlässigsten ist, insbesondere bei einem Bewertungsobjekt mit ungewöhnlich kleiner oder großer GLA (Gross Living Area).

Anschließend werden die Objekte nach ihren Residuen vom negativsten bis zum positivsten Wert sortiert. Sodann wird für jedes Objekt ein Condition-Quality-Appeal-Score (CQA-Score; Zustand, Qualität und Attraktivität) im Bereich von 0.0 bis 10.0 erzeugt. Dieser Score wird wie folgt berechnet:

Der CQA-Score entspricht dem Perzentilrang des Residuums eines Objekts unter allen Objekten. Konkret wird der Score ermittelt, indem der prozentuale Anteil der Objekte bestimmt wird, deren Residuen niedriger sind als das des betreffenden Objekts, und dieser Prozentsatz durch 10 dividiert wird. Ist beispielsweise das Residuum eines Objekts höher als die Residuen von 50% der Objekte, so beträgt sein Score 5.0.

Aufgrund von Rundungen kann ein Objekt einen Score von 10.0 erhalten, nämlich das Objekt mit dem höchsten Residuum. Ein perfekter Score von 10.0 ist jedoch theoretisch unmöglich, da logischerweise kein Objekt ein Residuum aufweisen kann, das höher ist als das von 100% der Objekte einschließlich seiner selbst. Der niedrigstmögliche Score von 0.0 ist hingegen möglich, da für mindestens ein Objekt kein anderes Objekt mit einem niedrigeren Residuum existiert. Zu

beachten ist, dass aufgrund von Rundungen mehr als ein Objekt einen Score von 0.0 erhalten kann, wenn viele Objekte verglichen werden.

Der CQA-Score liefert somit ein intuitives Maß dafür, wie gut die Merkmale eines Objekts mit den Erwartungen des Modells übereinstimmen, wobei höhere Scores eine bessere Übereinstimmung mit den Modellprognosen anzeigen.

Es ist unbedingt zu berücksichtigen, dass der CQA-Score nur dann als indirekter Indikator für die Attraktivität eines Vergleichsobjekts gelten kann, wenn die Verkaufsbedingungen seinem Marktwert entsprechen. Der Valuation Engineer ist dafür verantwortlich, durch eine umfassende Erläuterung und gründliche Untersuchung der Verkaufstransaktion festzustellen, ob der Residualwert über- oder untermarktliche Elemente enthält. Deutlich über oder unter dem Markt liegende Verkäufe sind zwar nicht häufig, können jedoch unter Umständen wie Zwangsverkäufen, Nachlassabwicklungen, Geschäftsunfähigkeit, Kollusion usw. auftreten.

Tabelle 3 zeigt einen halbrealistischen Auszug aus MLS-Daten zusammen mit Residuen und CQA-Scores, jeweils insgesamt und pro Quadratfuß. Die Gesamtzahl der in dieser Studie analysierten Objekte betrug etwa 600. Bemerkenswert ist, dass mehrere Variablen, die in die Regression eingingen, nicht dargestellt sind. Mit anderen Worten: Die dargestellten Residuen ergeben sich, nachdem der Einfluss anderer Variablen, wie etwa des Verkaufsdatums, durch die MARS-Regression berücksichtigt wurde.

Tabelle 3: Nach Residuum sortierte Vergleichsverkäufe

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

12.2 CQA-Residuum-Kurve

Eine Auftragung des CQA gegen das Residual/SF für ein Wohngebiet führt ausnahmslos zu der in [Figure 2](#) dargestellten charakteristischen CQA-Residuum-Kurve. Diese konkrete Darstellung beruht auf Daten aus Burlingame, CA, aus der Zeit um 2022. Ähnliche Kurven ergeben sich in der Regel auch, wenn CQA-Scores gegen Residualwerte aufgetragen werden; der Autor bevorzugt jedoch das Residual/SF, da es den CQA in Bezug auf die Bruttowohnfläche (GLA) des Bewertungsobjekts kontextualisiert. Durch Multiplikation des Residual/SF mit der GLA lässt sich der CQA eines Objekts in einen entsprechenden Residualwert umrechnen.

Der Valuation Engineer muss sorgfältig prüfen, ob die Größe des Objekts, insbesondere die Bruttowohnfläche (GLA), die Höhe der Residuen wesentlich beeinflusst. Diese Beurteilung ist von entscheidender Bedeutung, da sie die genaueste Methode zur Berechnung der Objektwerte bestimmen kann. So können beispielsweise hochwertige Küchengeräte und luxuriöse Badezimmersausstattungen den Residualwert eines Hauses erheblich steigern. In Wohngebieten, in denen Größe und Grundriss von Küchen und Bädern weitgehend einheitlich sind, korrelieren solche Aufwertungen jedoch möglicherweise nicht stark mit der GLA. Ihr Einfluss auf den Residualwert hängt dann weniger mit der beanspruchten Fläche als vielmehr mit der Qualität zusammen, die sie repräsentieren.

Umgekehrt können teure Außenmerkmale und -materialien, die die äußere Anmutung eines Objekts deutlich verbessern, in einem direkteren Zusammenhang mit der Objektgröße stehen. Größere Häuser bieten Raum für aufwendigere Außenanlagen, etwa umfangreiche Gartengestaltung, hochwertige Dachmaterialien oder großzügige Terrassen, die die Gesamtattraktivität des Objekts und damit seinen Residualwert auf natürliche Weise steigern.

Daher ist es für Valuation Engineers unerlässlich, diese Dynamiken bei der Analyse der Residuen zu berücksichtigen. Sie sollten nicht nur den durch hochwertige Ausstattung und Merkmale geschaffenen intrinsischen Wert beurteilen, sondern auch verstehen, wie diese Aufwertungen mit den physischen Abmessungen des Objekts zusammenwirken. Dieser umfassende Ansatz stellt sicher, dass die Bewertungen sowohl die qualitativen als auch die quantitativen Merkmale eines Objekts zutreffend widerspiegeln.

Folglich empfiehlt es sich, sowohl die CQA-Residuum-Kurve als auch die CQA-Residual/SF-Kurve darzustellen und die Unterschiede zu untersuchen, die auf Muster in einem Marktgebiet hinweisen können. Der Valuation Engineer kann sein Programm beide Residuen berechnen lassen und einen gewichteten Mittelwert bilden, um das dem Bewertungsobjekt zuzuweisende Residuum zu ermitteln. Ob dies einen nennenswerten Unterschied ausmacht, kann im Einzelfall variieren.

12.3 Eigenschaften der CQA-Kurve

Die Condition-Quality-Appeal-Kurve (CQA-Kurve) bietet ein differenziertes Modell zum Verständnis dafür, wie Immobilienwerte in Abhängigkeit von Attraktivität, Zustand und Marktdynamik innerhalb eines bestimmten Gebiets schwanken. Dieses Modell erfasst das Zusammenspiel wirtschaftlicher und physischer Faktoren, die den Immobilienmarkt beeinflussen, und spiegelt die nuancierten Realitäten der Immobilienbewertung über verschiedene Marktsegmente hinweg wider.

12.3.1 Analyse der Komponenten der CQA-Kurve

1. Exponentieller Abfall auf der linken Seite der Kurve:

Sinkt der CQA-Score von etwa 2 auf 0, weisen die Objekte Merkmale rascher Verschlechterung auf. Dieser Teil der Kurve wird gut durch eine exponentielle Abklingfunktion abgebildet, was

verdeutlicht, wie schnell ein Objekt ohne angemessene Instandhaltung und Modernisierung an Attraktivität und Wert verlieren kann. Physischer Verfall wie morschtes Holz, verblässende Farben und rostendes Metall beschleunigt diesen Niedergang, insbesondere bei Häusern, die nicht regelmäßig instand gehalten werden.

2. Linearer Anstieg im mittleren Bereich:

Der Großteil des Marktes fällt typischerweise in diese Kategorie, in der die Attraktivität von Scores von etwa 2 bis 8 nahezu linear zunimmt. Dieser Abschnitt umfasst Objekte, die im Allgemeinen standardmäßig instand gehalten werden, jedoch keine außergewöhnlichen Qualitäten oder Lagen aufweisen. Er repräsentiert die breite Mittelschicht des Wohnungsbestands und damit den überwiegenden Teil des Wohnimmobilienmarktes, in dem Häuser auf Grundlage ihrer Standardmerkmale und ihres Gesamtzustands wettbewerbsfähig bepreist werden.

3. Exponentielles Wachstum auf der rechten Seite der Kurve:

Hier steigt die Kurve von Scores von 8 bis 10 exponentiell an und spiegelt das attraktivere Ende des Marktes wider. Objekte in diesem Segment gehören häufig wohlhabenderen Personen, die nicht nur in überdurchschnittliche Instandhaltung, sondern auch in Aufwertungen investieren, welche die Attraktivität und den Wert des Objekts erheblich steigern. Das exponentielle Wachstum spiegelt die Knappheit solcher hochwertigen Objekte und die hohe Nachfrage vermögender Käufer wider. Dabei ist zu beachten, dass ein preisgünstigeres Haus hinsichtlich seiner Attraktivität sehr hoch eingestuft werden kann, während ein großes und teures Haus niedrig eingestuft werden kann. Diese Dynamiken werden durch den fachkundigen Einsatz von MARS aufgedeckt und bleiben den meisten Gutachtern in der Regel verborgen.

12.3.2 Marktdynamik und soziale Implikationen

Die CQA-Kurve erzählt implizit die Geschichte relativen Wohlstands und wirtschaftlicher Ungleichheit innerhalb verschiedener Regionen. In Gegenden wie San Mateo County veranschaulichen die Wertunterschiede zwischen Gebieten wie Burlingame und wohlhabenderen Wohnlagen wie Hillsborough, Woodside oder Atherton erhebliche sozioökonomische Gefälle. Diese Disparität wirkt sich auf alles aus, von der Instandhaltung der Objekte bis hin zu den Marktbewertungen, und beeinflusst letztlich die Käuferdemografie und die Trends am Immobilienmarkt.

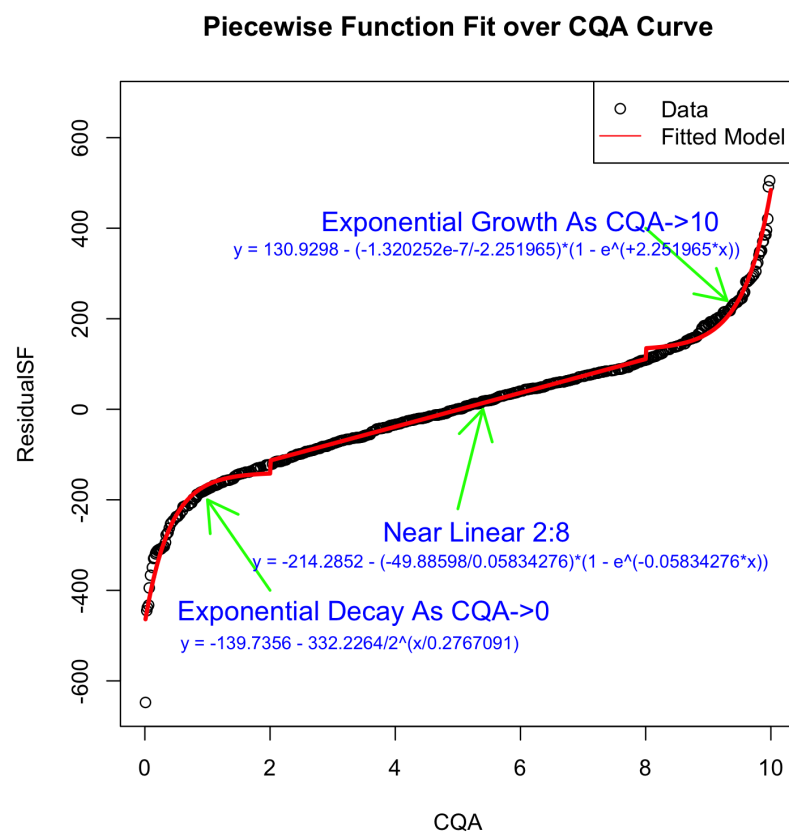
Darüber hinaus verdeutlicht die Anekdote über eine relativ wohlhabende Person, die in ein hochwertiges Haus in einem Wohngebiet mittlerer Preisklasse investiert, in dem sie sich wohlfühlt und aus verschiedenen Gründen bevorzugt leben möchte, einen weiteren zentralen Aspekt der Immobilienökonomie — das Konzept „überbauter“ Objekte („overbuilt“). Solche Objekte finden trotz hoher Attraktivität möglicherweise keine lokalen Käufer, die den Preis zu zahlen bereit sind, den sie in einem höherpreisigen Marktgebiet hätten erzielen können.

Die CQA-Residuum-Kurve ist für jedes Marktgebiet charakteristisch und sollte bei der Analyse nicht außer Acht gelassen werden.

12.4 Residuum des Bewertungsobjekts

Im Gegensatz zu den Vergleichsobjekten besitzt das Bewertungsobjekt kein objektiv bestimmtes Residuum; dieses muss vom Valuation Engineer (VE) geschätzt werden. Diese Schätzung führt ein gewisses Maß an Subjektivität in das Protokoll des Residual Constraint Approach (RCA) ein und markiert einen kritischen Punkt, an dem das persönliche Urteil eine entscheidende Rolle spielt.

Abbildung 2: Charakteristische CQA-Residuum-Funktion



Zur Schätzung des Condition-Quality-Appeal-Scores (CQA-Scores) analysiert der VE sorgfältig die in eine Rangfolge gebrachten Objekte und sucht nach erkennbaren Mustern innerhalb der Rangfolge. Diese Analyse umfasst häufig eine detaillierte Durchsicht der MLS-Fotos zu jedem Objekt. Eine erschöpfende Durchsicht ist bei einem großen Datensatz, etwa 600 Objekten, nicht praktikabel. Stattdessen wendet der VE praktisches Schlussfolgern an, um die wahrscheinliche Position des Bewertungsobjekts innerhalb der Gesamtrangfolge näherungsweise zu bestimmen. Objekte mit umfangreichem Reparaturbedarf, sogenannte „fixers“, sind in der Regel am unteren Ende positioniert, während Objekte mit überdurchschnittlichen Modernisierungen und hoher Attraktivität tendenziell am oberen Ende rangieren.

Anschließend geht der VE gezielter vor und ‚springt‘ innerhalb der Rangfolge, um Vergleichsobjekte zu identifizieren, die der Attraktivität des Bewertungsobjekts weitgehend entsprechen. Mit zunehmender Eingrenzung des Bereichs achtet der VE genauer auf Ähnlichkeiten zwischen den Objekten, insbesondere in Aspekten wie Qualität, Zustand, gestalterischer Ästhetik (einschließlich Elementen wie Anstrich, Holzarbeiten, Steinarbeiten und Fenstern), funktionaler Nutzbarkeit und anderen relevanten Merkmalen.

Es ist von entscheidender Bedeutung zu verstehen, dass die Urteile des VE grundlegend durch die vom Markt etablierten Rangfolgen geleitet werden. Dadurch wird sichergestellt, dass die subjektiven Elemente des Verfahrens in den Marktrealitäten verankert sind, mit dem Ziel einer fairen und genauen Bewertung des Bewertungsobjekts.

12.5 RCA-Wertschlussfolgerung

12.5.1 Allgemeine Überlegungen

Die Bestimmung eines CQA-Scores ist für die Festlegung des geschätzten Residualwerts eines Objekts von zentraler Bedeutung, da diese Elemente untrennbar miteinander verbunden sind. Zwar ist eine Beurteilung eines Objekts auch ohne CQA möglich, doch liefert diese Kennzahl wertvollen Kontext, indem sie verdeutlicht, wo das Objekt im Vergleich zu anderen im Marktgebiet steht. Ein CQA-Score von 3.2 zeigt beispielsweise an, dass das Objekt eine höhere Attraktivität aufweist als 32% der Vergleichsobjekte in der Umgebung. Bestehen Einwände gegen den zugewiesenen Score, müssen die Einwendenden die Rangfolgen konsultieren und vorgeschlagene Anpassungen anhand von Fotobelegen oder anderen substanziellen Daten begründen. Ist das Rangsystem robust und hat der Valuation Engineer (VE) das Objekt innerhalb dieses Rahmens zutreffend positioniert, ist es unwahrscheinlich, dass wesentliche Abweichungen auftreten.

In der Berichtsphase werden typischerweise 6 bis 12 aktuelle und relevante Vergleichsverkäufe ausgewählt, um die Wertschlussfolgerung zu stützen. Die Unterschiede zwischen diesen Vergleichsobjekten und dem Bewertungsobjekt werden analysiert, wobei jede Variable entsprechend angepasst wird. Diese Anpassungen werden anschließend auf die Nettoverkaufspreise der Vergleichsobjekte angewendet, um angepasste Verkaufspreise abzuleiten. Wie mathematisch gezeigt wurde, sollten diese angepassten Preise mit dem geschätzten Verkaufspreis des Bewertungsobjekts konvergieren.

12.5.2 Überprüfung, Auditierung, Beschwerden

Sobald das geschätzte Residuum für das Bewertungsobjekt bestimmt ist, können Residualanpassungen für alle Vergleichsobjekte berechnet werden. Diese Anpassungen dienen als Restriktionen für weitere manuelle Aufschlüsselungen durch den VE, zugeschnitten auf die jeweils als relevant erachteten CQA-Variablen. Es ist wichtig zu betonen, dass diese detaillierten Aufschlüsselungen erläuternden Charakter haben und die endgültige Bewertung nicht beeinflussen, wie in Beweis II

bestätigt.

Was Beschwerden von Nutzern betrifft, die mit dem Wertschlussfolgerung des Gutachtens nicht einverstanden sind, so muss ein traditioneller Gutachter typischerweise Zeit darauf verwenden zu erläutern, warum dieses oder jenes Objekt mit einem höheren oder niedrigeren Preis nicht die vom Nutzer erwartete Auswirkung auf den Wert hat. Beim RCA werden alle in Frage kommenden Vergleichsverkäufe (in der Regel 100–600) typischerweise auf exakt denselben Wert wie das Preisergebnis hin ausgewertet, sodass es wenig Aufwand erfordert, die Anpassungen für ein beliebiges Objekt aufzuzeigen. Der gesamte RCA-Prozess ist automatisiert und objektiv, mit Ausnahme der Bewertung des Residuums des Bewertungsobjekts. Ein weiterer Vorteil des RCA besteht somit darin, dass Kritik an Unvollkommenheiten des subjektiven Urteils des VE im Wesentlichen auf diesen einen Bereich beschränkt ist — und dort wirksam und effizient gehandhabt werden kann. Zudem führt, ausgehend von einem R^2 von etwa 80%, ein Fehler von $\pm 10\%$ bei der Schätzung des Residuums des Bewertungsobjekts lediglich zu einem Fehler von $\pm 2\%$ in der endgültigen Wertschätzung.

In vielen Fällen fügt sich das Bewertungsobjekt so passgenau in die Residuen-Rangfolge der Objekte ein, dass eine Genauigkeit innerhalb von $\pm 1\%$ erwartet werden kann.

Die Summe aus dem geschätzten Residuum und dem durch MARS geschätzten Verkaufspreis des Bewertungsobjekts ergibt das endgültige Wertschlussfolgerung, das als Residual Constraint Approach (RCA) Value Conclusion (RCA-Wertschlussfolgerung) bezeichnet wird. Dieses Verfahren gewährleistet einen umfassenden und transparenten Ansatz der Immobilienbewertung und stärkt das Vertrauen in die Genauigkeit und Fairness der Beurteilung.

13 Schlussfolgerung

Der Residual Constraint Approach ist ein Rahmenwerk mit drei substanziellen Beiträgen: einer strukturellen Anti-Ankereffekt-Eigenschaft, die sich daraus ergibt, dass das Modell auf dem gesamten Markt und nicht auf einer kleinen Menge von Vergleichsobjekten angepasst wird (Beweis I); einem Invarianzresultat, das es erlaubt, die Residuenzerlegung als erklärende Ebene zu nutzen, ohne das Wertschlussfolgerung zu beeinträchtigen (Beweis II und die darauf folgende Bemerkung); sowie einer empirischen Regelmäßigkeit — der charakteristischen CQA-Residuen-Kurve —, die in charakteristischer exponentiell-linear-exponentieller Form über Marktgebiete hinweg wiederzukehren scheint. Die Grenzen des Rahmenwerks sind gleichermaßen substanziell: Die Residuenzerlegung ist invariant, aber nicht identifizierend, der Schritt der CQA-Einordnung beruht auf einer VE-Einschätzung, die noch nicht durch eine formale Studie zur Interrater-Reliabilität validiert wurde, und die Aussagen zur Genauigkeit stehen noch vor einem Benchmarking gegenüber Alternativen.

Was diese Arbeit vorlegt, sind ein Rahmenwerk und ein Protokoll. Was sie noch nicht vorlegt, das größere, derzeit laufende Forschungsprogramm jedoch schon, ist eine ausführbare, prüfpfadfähige Pipeline, die das Protokoll in Software mit vertretbaren Standardeinstellungen operationalisiert. Diese Pipeline — das Paket earthUI auf CRAN für die Stufe der MARS-Basisfunktionen, das Paket glmnetUI für das regularisierte Strukturmodell auf diesen Basen und das Paket mgcvUI für die räumliche Glättung — ist Gegenstand von Begleitartikeln in dieser Ausgabe und bildet die Brücke von RCA als Methode zu RCA als lieferbarem Ergebnis. Die Lücke zwischen Doktrin und Praxis in der Regulierung der Wertermittlung (Craytor, 2026, in dieser Ausgabe) liefert den Grund, weshalb diese Brücke von Bedeutung ist: Die Ergebnisse der Wertermittlung werden ökonomisch als zukunftsgerichtete Risikoeingaben konsumiert, und ein Rahmenwerk, das die Gegenwartswert-Einschätzung prüfbar und reproduzierbar macht, ist eine Voraussetzung für jede redliche Auseinandersetzung mit dieser zukunftsgerichteten Verwendung. RCA ist ein Schritt in diesem größeren Vorhaben. Es ist nicht das gesamte Vorhaben, und diese Arbeit erhebt auch nicht diesen Anspruch.

Literatur

- Ambrose, Brent W. u. a. (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, S. 1–141.
- Hastie, Trevor, Robert Tibshirani und 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 u. a. (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 und 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 und 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*. Techn. Ber. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (2024c). *Variance Models in Earth*. Techn. Ber. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.
- Perry, Andre M., Jonathan Rothwell und David Harshbarger (2018). *The Devaluation of Assets in Black Neighborhoods*. Techn. Ber. Brookings Institution.
- Perry, Andre M., Jonathan Rothwell, J. Williamson u. a. (2024). *How Racial Bias in Appraisals Affects the Devaluation of Homes in Majority-Black Neighborhoods*. Techn. Ber. Brookings Institution.
- Property Appraisal and Valuation Equity Task Force (2022). *Action Plan to Advance Property Appraisal and Valuation Equity*. Techn. Ber. U.S. Department of Housing und Urban Development.
- Tzioumis, Konstantinos (2016). „Appraisers and Valuation Bias: An Empirical Analysis“. In: *Real Estate Economics* 45.3, S. 679–712. DOI: [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- Williamson, Jake und Mark Palim (2022). *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Techn. Ber. 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: Eine interaktive Benutzeroberfläche für das earth-(MARS-)Paket

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische

Fassung ist maßgeblich.

William Bert Craytor

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

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

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

Dieser Open-Access-Artikel wird unter den Bedingungen der [Creative-Commons-Namensnennung-4.0-Lizenz](#) verbreitet.

Zusammenfassung. earthUI ist eine grafische Benutzeroberfläche für das R-Paket **earth**, das Multivariate Adaptive Regression Splines (MARS) implementiert. Es bietet drei Verwendungsmodi — allgemeine prädiktive Modellierung, Immobilien-Wertermittlung und Marktgebietsanalyse — und führt den Anwender durch Datenimport, Modellkonfiguration, Modellanpassung sowie die Interpretation der Diagnostik und grafischen Ausgaben von earth. Dieser Artikel dokumentiert die Datenformat-Anforderungen, den Modellierungs-Workflow, die Ausgabedarstellungen und die vollständige Funktionsreferenz von earthUI.

Schlüsselwörter: MARS, Multivariate Adaptive Regression Splines, earth, R, grafische Benutzeroberfläche, Immobilienbewertung

Anmerkung des Verlags (Juli 2026): Dieser Artikel beschreibt earthUI 0.10.0. Kurz vor Drucklegung hat earthUI 0.11.0 die Begleitanwendungen glmnetUI und mgcvUI (in den beiden folgenden Artikeln beschrieben) übernommen und damit die MARS-, Elastic-Net- und GAM-Workflows in einer einzigen Anwendung zusammengeführt. Die hier beschriebene Benutzeroberfläche und der Workflow bleiben von der Zusammenführung unberührt.

Inhaltsverzeichnis

1	Einleitung	4
1.1	Was ist earthUI?	4
1.2	Ein Paket, drei Routinen	4
1.3	Drei Purpose-Modi	4
1.4	Immobilienpezifische Funktionen	5
1.5	Installation von earthUI und vProlog	5
1.6	Erste Schritte	6
2	Anforderungen an MLS-Eingabedaten	7
2.1	Dateiformat & Struktur	7
2.2	Erforderliche Spalten für den Appraisal-Modus	7
2.3	Besondere Konventionen für Spaltennamen	9
2.4	Datenqualität & Vollständigkeit	9
2.5	Platzierung der Zeile des Bewertungsobjekts	9

3	General Purpose Mode	11
3.1	Überblick	11
3.2	Der Arbeitsablauf in der Sidebar	11
3.3	Tabs des Hauptpanels	12
3.4	Persistenz der Einstellungen	12
3.5	Dark Mode	12
3.6	Locale & regionale Einstellungen	13
3.6.1	Unterstützte Länder	13
3.6.2	Übersteuerung des Papierformats	13
3.6.3	Speichern der Standardeinstellungen	14
4	Appraisal Mode (Gutachtenmodus)	15
4.1	Behandlung der Subject-Zeile	15
4.2	Effective Date & Sale Age	15
4.3	Special-Spalten-Ausweisungen	15
4.4	Residualer Wertbeitrag (Schätzung seltener Merkmale)	16
4.5	Überblick über die RCA-Anpassungen	17
5	Market Area Analysis Mode (Marktgebietsanalysemodus)	18
5.1	Unterschiede zum Appraisal Mode	18
5.2	Wann der Market Mode zu verwenden ist	18
6	Variablenauswahl	19
6.1	Zielvariable(n)	19
6.2	Die Prädiktortabelle	19
6.3	Datentyperkennung und Überschreibungen	20
6.4	Die Markierungen Factor und Linear	20
6.5	Referenz der speziellen Spaltentypen	20
6.6	Mehrere Zielvariablen	21
7	Parameterauswahl	22
7.1	Grundparameter	22
7.2	Vorwärtsdurchlauf (Forward Pass)	22
7.3	Allowed Interactions (zulässige Interaktionen)	23
7.4	Pruning	23
7.5	Kreuzvalidierung	23
7.6	Subset-Filterung	24
7.7	Empfohlene Parameterwerte	24
7.7.1	Parameter des Vorwärtsdurchlaufs	24
7.7.2	Pruning-Parameter	26
7.7.3	Kreuzvalidierungs-Parameter	26
7.7.4	Varianzmodell	27
7.7.5	Earth-Ausgabedatei	27
7.8	Voreinstellungen (Settings Defaults)	27
8	Anpassung des Earth-Modells	28
8.1	Die Schaltfläche „Fit“	28
8.2	Anpassungs-Modal & Trace-Ausgabe	28
8.3	Anpassungsprotokoll	28

9	Herunterladen der Daten	29
9.1	Ausgabespalten	29
9.2	Spaltenanordnung & Excel-Formatierung	29
9.3	CQA-Scores	30
9.4	Sortierung & Zeile des Bewertungsobjekts	30
9.5	Automatischer mgcvUI-Export	30
9.6	glmnetUI-Import	31
9.6.1	Wesentliche Vorteile	31
9.6.2	Wichtig: Die Gewichtungsspalte muss übereinstimmen	31
9.6.3	Arbeitsablauf	31
10	RCA-Berechnungen & Download	33
10.1	Öffnen des RCA-Dialogs	33
10.2	Interpolation des CQA-Scores	33
10.3	Ausgabespalten	33
10.4	Die Registerkarte RCA Adjustments	34
11	Sales Comparison Grid	35
11.1	Überblick & Zweck	35
11.2	Modaldialog zur Auswahl der Vergleichsobjekte	35
11.3	Rasterlayout	35
11.4	Sale Price, Concessions & Net SP	36
11.5	Gruppierte Zeilen (Location, Site, Age)	36
11.6	Modellvariablenzeilen	36
11.7	CQA Residual & Zeilen für Residualmerkmale	37
11.8	Formel für den Adjusted Sale Price	37
11.9	Blattschutz & editierbare Zellen	37
11.10	Arbeiten mit dem Raster in Excel	37
12	Herunterladen von Berichten	39
12.1	Generate Quarto Report	39
12.2	Convert Quarto Report	39
12.3	Berichtsinhalte	39
13	Die Post-Processing-Module: Elastic Net und GAM	41
13.1	Wozu unterstützende Anpassungen?	41
13.2	Start und gemeinsamer Zustand	41
13.3	Die Elastic-Net-Routine	41
13.4	Die GAM-Routine	41
13.5	Ein praktischer Arbeitsablauf	42
14	Demo-Datensatz: Appraisal_1.csv	43
14.1	Beschreibung	43
14.2	Spalten	44
14.3	Empfohlener Schnelleinstieg	44
15	Experimentell: Prolog-Textextraktion & Ableitungsregeln	46
15.1	Funktionsweise	46
15.2	Voraussetzungen: vProlog (Trealla), nicht SWI-Prolog	47
15.3	Aktivierung und Ausführung	47
15.4	Ableitungsregeln	47
15.5	Versionierte aktive Datendatei	48

15.6	Programmatische Schnittstelle & mitgelieferte Dateien	48
15.7	Verwandtes Experiment: Klimaregionen-Priors	48
15.7.1	Über Kalifornien hinaus: Jurisdiction Packs und Market-Area-Module . .	48

1 Einleitung

1.1 Was ist earthUI?

earthUI ist eine grafische Benutzeroberfläche für das R-Paket **earth**, das Multivariate Adaptive Regression Splines (MARS) implementiert (Friedman 1991). Die Modellierungsmaschine ist Stephen Milborrows Paket **earth** (Milborrow 2024), das von Milborrow (o.D.[a]) und Milborrow (o.D.[b]) ausführlich dokumentiert wird. Es läuft als lokale Shiny-Anwendung — es gibt keine Anmeldung, keinen Server und keine Benutzerkonten. Die Anwendung wird aus R heraus gestartet; anschließend wird ein Datensatz (CSV oder Excel) importiert, das Modell konfiguriert und interaktiv angepasst.

Die Anwendung stellt einen vollständigen Arbeitsablauf bereit: Datenimport, Variablenkonfiguration, Modellanpassung, Diagnosegrafiken, Variablenwichtigkeit, Modellgleichungen sowie herunterladbare Berichte im HTML-, Word- oder PDF-Format.

1.2 Ein Paket, drei Routinen

Seit der aktuellen Version *enthält* das Paket earthUI zudem seine beiden Post Processing Modules (Nachverarbeitungsmodule), die in früheren Versionen die separaten Pakete glmnetUI und mgcvUI waren. Mit der Installation von earthUI werden alle drei Routinen installiert; jede läuft als eigene lokale Shiny-Anwendung:

- **earthUI::launch()** — die in diesem Artikel beschriebene earth-(MARS-)Anwendung auf Port 7878. **earth()** ist die zentrale Bewertungsmaschine.
- **earthUI::launch_glmnet()** — das Elastic-Net Post Processing Module (Port 7879), das die Basisfunktionen des earth-Modells importiert und unter Lasso-/Elastic-Net-Regularisierung neu schätzt (siehe „glmnetUI Import“ weiter unten).
- **earthUI::launch_mgcv()** — das GAM Post Processing Module (Port 7880), das die Knotenpositionen des earth-Modells als Startpunkte für mgcv-Glättungsterme verwendet (siehe „mgcvUI Auto-Export“ weiter unten).

Die drei Routinen teilen sich einen Projektbaum, eine Einstellungsdatenbank und eine Kopie der unterstützenden Infrastruktur, sodass ein in einer Anwendung geöffnetes Projekt den anderen sofort zur Verfügung steht. Die glmnet- und mgcv-Routinen dienen als stützende Evidenz für die Wertschlussfolgerung des earth-Modells und stellen keine konkurrierenden Arbeitsabläufe dar: Das vorgesehene Vorgehen besteht darin, zunächst das earth-Modell anzupassen und zu verfeinern und anschließend — wenn eine Bestätigung gewünscht wird — die Module auf dasselbe Projekt anzuwenden.

1.3 Drei Purpose-Modi

earthUI organisiert die Arbeit in **Projekten** — benannten Einheiten, die Eingabedateien, Ausgaben und Einstellungen an einem festen Ort zusammenhalten. Jedes Projekt trägt einen **Purpose** (Verwendungszweck), der im New Project Wizard (Assistent zur Projektanlage) bei der Erstellung des Projekts gewählt wird. Es gibt drei Modi:

- **General** (Allgemein) — Earth-Regression für beliebige Populationen oder Datensätze. Dies ist der Standardmodus. Er stellt den vollständigen MARS-Modellierungsablauf ohne domänenspezifische Ergänzungen bereit.

- **For Appraisal** (für die Wertermittlung) — Earth-Regression, zugeschnitten auf die Wertermittlung von Immobilien. Ergänzt Funktionen speziell für die Bewertung einzelner Objekte, darunter die Behandlung des Bewertungsobjekts, spezielle Spaltenkennzeichnungen, den Residual Constraint Approach (RCA) (dt. etwa: Residualrestriktionsansatz) sowie die Erzeugung des Sales Comparison Grid (Vergleichsraster).
- **Market Area Analysis** (Marktgebietsanalyse) — Earth-Regression, zugeschnitten auf Marktgebietsstudien. Ergänzt Funktionen zur Analyse von Objektgruppen in einem definierten Markt anstelle der Bewertung eines einzelnen Bewertungsobjekts.

In allen drei Modi ist die zentrale Modellierungsmaschine identisch — es wird stets ein Earth-(MARS-)Modell angepasst. Die Purpose-Einstellung steuert, welche zusätzlichen Werkzeuge und Oberflächenelemente verfügbar sind.

Der Purpose ist je Projekt fest. Der Verwendungszweck eines Projekts — ebenso wie sein Land, Bundesstaat, County und seine Stadt — wird bei der Erstellung des Projekts festgeschrieben. Beim Wechsel zu einem anderen Projekt wird earthUI auf einen sauberen Ausgangszustand zurückgesetzt (importierte Daten, Modellergebnisse, Tabs, Variablenkonfiguration und earth-Parameter werden sämtlich gelöscht); anschließend wird die für den Purpose dieses Projekts gespeicherte Konfiguration wiederhergestellt.

1.4 Immobilienspezifische Funktionen

Wenn entweder **For Appraisal** oder **Market Area Analysis** gewählt ist, aktiviert earthUI mehrere Funktionen, die für die Immobilienanalyse konzipiert sind:

- **Spezielle Spaltenkennzeichnungen** — Jeder Prädiktor kann mit einer speziellen Rolle versehen werden, etwa `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` oder `display_only`. Diese Kennzeichnungen steuern, wie die Spalte bei der Modellanpassung, der Ausgabe und der Erzeugung des Sales Grid behandelt wird. Die vollständige Referenz findet sich in Kapitel 6.
- **Rundung von geografischer Breite und Länge** — Spalten, die als `latitude` oder `longitude` gekennzeichnet sind, werden automatisch auf 3 Dezimalstellen gerundet, um Überanpassung zu vermeiden.
- **Sale-Age-Spalte** — Wenn eine Spalte als `contract_date` gekennzeichnet ist und ein Effective Date (Wertermittlungstichtag) angegeben wird, berechnet earthUI eine Spalte `sale_age` (Tage zwischen Verkaufsdatum und Stichtag) und setzt sie als Prädiktor ein.
- **RCA-Berechnungen (nur Wertermittlung)** — Im Appraisal Mode kann earthUI nach der Modellanpassung Ausgaben des Residual Constraint Approach (RCA) berechnen; diese umfassen Anpassungen je Vergleichsobjekt, Zusammenfassungen der Netto-/Bruttoanpassungen sowie einen angepassten Verkaufspreis für das Bewertungsobjekt.
- **Sales Comparison Grid (nur Wertermittlung)** — Erzeugt eine Excel-Arbeitsmappe mit mehreren Blättern und Formeln für die Anpassungen der Vergleichsobjekte, Aufschlüsselungen der Residualmerkmale und angepasste Verkaufspreise. Siehe Kapitel 11.

***Hinweis:** Der Purpose-Modus General eignet sich für beliebige Datensätze — aus Finanzwesen, Wissenschaft, Ingenieurwesen oder Immobilienwirtschaft. Die Modi für Wertermittlung und Marktanalyse fügen lediglich Komfortfunktionen für Immobilienfachleute hinzu.*

1.5 Installation von earthUI und vProlog

earthUI ist auf CRAN veröffentlicht; die Standardinstallation besteht daher aus einer Zeile in R:

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

Dieser Artikel beschreibt die aktuelle Version. Sollte CRAN zum Zeitpunkt der Lektüre noch eine ältere Version führen, lässt sich die Entwicklungsversion von GitHub installieren (man beachte das `subdir` — das Paket liegt im Ordner `pkg/` des Repositorys):

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

vProlog (optional). Die experimentelle Funktion zur Verarbeitung von Anmerkungen (siehe „Prolog Text Extraction“ gegen Ende dieses Artikels) verwendet das Paket `vProlog` — eine eigenständige, eingebettete Prolog-Maschine. Es ist bewusst *keine* zwingende Abhängigkeit: `earthUI` lässt sich ohne `vProlog` installieren und ausführen, und die Prolog-Funktionen bleiben schlicht deaktiviert, bis das Paket vorhanden ist. `vProlog` wird über das `r-universe`-Repository des Autors verteilt, das vorkompilierte Binärdateien für Windows und macOS bereitstellt (keine Compiler-Werkzeugkette erforderlich):

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

Eine systemweite Prolog-Installation (SWI-Prolog oder anderes) ist nicht erforderlich — die Trealla-Maschine ist in das Paket selbst einkompiliert. Alles Übrige in diesem Artikel funktioniert ohne `vProlog`.

1.6 Erste Schritte

Zur Verwendung von `earthUI`:

1. **Installation des Pakets** von CRAN, wie oben beschrieben.
2. **Start der Anwendung:** In R wird `earthUI::launch()` ausgeführt, oder von der Kommandozeile aus `Rscript -e 'earthUI::launch()'`. Die Anwendung öffnet sich im Webbrowser auf Port 7878. Alternativ lässt sich die Anwendung direkt über `http://localhost:7878` im Browser aufrufen. (Die Post Processing Modules starten auf dieselbe Weise: `launch_glmnet()` auf Port 7879, `launch_mgcv()` auf Port 7880.)
3. **Erstellen oder Öffnen eines Projekts** in Section 1 der Seitenleiste. Der New Project Wizard erfasst den Projektnamen, den Purpose (General, For Appraisal oder Market Area Analysis), den Standort (Land, Bundesstaat, County, Stadt) und optional eine anfängliche Datendatei, die in den Ordner `in/` des Projekts kopiert wird.
4. **Import der Daten** in Section 2, wo die CSV- und Excel-Dateien im Ordner `in/` des aktiven Projekts aufgelistet werden. Dateien, die manuell über Finder oder Explorer hinzugefügt werden, erscheinen nach einem Klick auf Refresh (Aktualisieren).
5. **Konfiguration der Variablen** — Zielvariable(n) und Prädiktoren wählen, Datentypen festlegen und gegebenenfalls spezielle Spaltenrollen zuweisen.
6. **Festlegen der Earth-Parameter** — degree, nprune, Teilmengenfilter und weitere Argumente von `earth()`.
7. **Anpassen des Modells** — auf „Fit Earth Model“ klicken und die Ergebnisse im Hauptbereich prüfen.
8. **Export** — Prognosen als Excel herunterladen, Quarto-Berichtspakete erzeugen oder (im Appraisal Mode) RCA-Anpassungen und Sales Comparison Grids berechnen.

Die Konfiguration wird serverseitig in einer Projektdatenbank gespeichert, indiziert nach Projekt und Purpose, über die Schaltflächen „Save current as default“ in den Sections 3 und 4 (siehe Kapitel 3, „Settings Persistence“).

2 Anforderungen an MLS-Eingabedaten

Für Arbeitsabläufe der Immobilienwertermittlung und Marktanalyse stammen die Eingabedaten typischerweise aus einem Export eines Multiple Listing Service (MLS). Dieses Kapitel beschreibt die erwartete Dateistruktur sowie die Spalten, die earthUI verwenden kann.

2.1 Dateiformat & Struktur

earthUI akzeptiert **CSV**- und **Excel**-Dateien (**.xlsx**, **.xls**). Beim Import werden die Spaltennamen automatisch in **snake_case** umgewandelt — so wird beispielsweise aus „Living SqFt“ **living_sqft**, aus „Contract Date“ **contract_date** und aus „Sale Price“ **sale_price**. Diese Normalisierung stellt konsistente Spaltenreferenzen über den gesamten Arbeitsablauf hinweg sicher. Das beim Import verwendete CSV-Trennzeichen und das Dezimalzeichen werden durch die Locale-Einstellungen bestimmt (siehe Kapitel 3, „Locale & Regional Settings“).

Die Datendatei sollte eine flache Tabelle mit einer Zeile pro Objekt und einer Spalte pro Attribut sein. Die erste Zeile der Datei muss die Spaltenüberschriften enthalten.

2.2 Erforderliche Spalten für den Appraisal-Modus

Zwar arbeitet earthUI mit einem beliebigen Satz von Spalten, doch profitiert der vollständige Wertermittlungs-Arbeitsablauf (RCA-Anpassungen (Residual Constraint Approach, dt. etwa: Residualrestriktionsansatz) + Sales Comparison Grid) davon, wenn der MLS-Export die folgenden Spalten enthält:

Spalte	Spezialtyp	Zweck
Sale Price	(target)	Zielvariable für das Modell
Contract Date	contract_date	Zur Berechnung von sale_age (Tage ab dem Wertermittlungstichtag)
Listing Date	listing_date	Zusammen mit dem Vertragsdatum zur Berechnung der DOM, falls keine DOM-Spalte vorhanden ist
Days on Market	dom	Vermarktungsdauer in Tagen; im Sales Grid (Vergleichsraster) angezeigt
Concessions	concessions	Verkaufszugeständnisse; Net SP = Sale Price – Concessions im Sales Grid
Living Area (SF)	living_area	Ermöglicht Residuen pro Quadratfuß (residual_sf , cqa_sf)
Lot Size	lot_size	In der Zeile "Site Size Dimensions" des Sales Grid gruppiert
Site Dimensions	site_dimensions	Mit der Grundstücksgröße im Sales Grid gruppiert
Latitude	latitude	Auf 3 Dezimalstellen gerundet; für die Distanzberechnung und die Location-Gruppe verwendet
Longitude	longitude	Auf 3 Dezimalstellen gerundet; für die Distanzberechnung und die Location-Gruppe verwendet
Area ID	area	In der Zeile "Loc: Long Lat Area" des Sales Grid gruppiert
Actual Age	actual_age	In der Zeile "Actual Age Effective Age" des Sales Grid gruppiert
Effective Age	effective_age	Mit dem tatsächlichen Alter im Sales Grid gruppiert
Parcel Number	(none)	In der APN-Zeile des Sales Grid angezeigt
Listing ID	(none)	In der APN-Zeile des Sales Grid angezeigt
Address	display_only	In der Address-Zeile des Sales Grid angezeigt; vom Modell ausgeschlossen

Die Spaltennamen der Tabelle können in einer Fremdsprache vorliegen — die „speziellen“ Namen sind auf Englisch gehalten, damit das R-Programm sie gesondert behandeln kann. Ansonsten erscheinen die vergebenen Spaltennamen in den Regressionsmodellen, den Grafiken und (bei Gutachten) im Intermediate Sales Grid.

Nicht alle Spalten sind erforderlich. earthUI passt sich an — fehlt eine Spalte, wird das entsprechende Merkmal schlicht weggelassen. Wird beispielsweise keine **concessions**-Spalte zugewiesen, zeigt die Net-SP-Zeile im Sales Grid den Verkaufspreis ohne Abzug von Zugeständnissen. Für Preismodelle im Immobilienbereich werden jedoch bestimmte Spalten dringend empfohlen, um eine akzeptable Anpassungsgüte zu erreichen:

1. **Sale Age** (Verkaufsalter) — die Anzahl der Tage zwischen dem Vertragsdatum des Verkaufs und dem Stichtag der Wertermittlung oder Analyse. Wird eine mehrjährige Verkaufshistorie verwendet, insbesondere über Zeiträume von mehr als 5 Jahren, spielt **sale_age** häufig eine zentrale Rolle bei der Schätzung des Verkaufspreises. Tatsächlich ist diese Variable oft so bedeutsam, dass earthUI ohne sie überhaupt kein Modell liefern kann.
2. **Living Area** (Wohnfläche) — auch unter Bezeichnungen wie „Living Sqft“, „GLA“ (gross living area) und dergleichen bekannt. Dies ist eine weitere maßgebliche Bestimmungsgröße des Verkaufspreises.
3. **Total Bath Count** (Gesamtzahl der Bäder) — die Gesamtzahl der Voll-, Viertel-, Halb- und Dreiviertelbäder. Beispielsweise ergäben zwei Vollbäder und ein Halbbad den Wert 2.5.

4. **Garage Bays** (Garagenstellplätze) oder **Garage Area** (Garagenfläche) — die Anzahl der Garagenstellplätze oder die Garagenfläche in Quadratfuß.
5. **Lot Size** (Grundstücksgröße) — die Grundstücksfläche des Objekts, typischerweise in Quadratfuß oder Acres.
6. **Longitude**, **Latitude** und, sofern verfügbar, **Area ID**. Die Anpassungen für diese Variablen werden im Sales Grid zu einer einzigen Location-Anpassung (Lageanpassung) zusammengefasst.

2.3 Besondere Konventionen für Spaltennamen

earthUI identifiziert Spalten anhand ihrer **Spezialtyp-Zuweisung**, nicht anhand ihres Spaltennamens. Die Spalten im MLS-Export können beliebig benannt werden — entscheidend ist, dass in der Variable Configuration table (Variablenkonfigurationstabelle, Kapitel 6) der korrekte Spezialtyp zugewiesen wird.

Beispielsweise kann ein MLS die Wohnfläche als „GLA“, „Living SqFt“, „liv_area“ oder „gross_living_area“ exportieren. Nach dem Import (bei dem der Name in snake_case umgewandelt wird) wird die Spalte einfach im Special-Dropdown als `living_area` ausgewiesen. earthUI verwendet sie dann unabhängig von ihrem ursprünglichen Namen für die Berechnung der Residuen pro Quadratfuß und die Gruppierung im Sales Grid.

2.4 Datenqualität & Vollständigkeit

- **Fehlende Werte (NA)**: Zeilen mit NA-Werten in einer Prädiktor- oder Zielspalte werden vor der Modellanpassung automatisch entfernt. Intern ist `na.action` stets auf `na.fail` gesetzt; die Entfernung der NA-Werte erfolgt vor dem Aufruf von `earth()`.
- **Datumsspalten**: Datumsangaben sollten in einem Format vorliegen, das R interpretieren kann (z. B. 2025-06-15, 06/15/2025, June 15, 2025). earthUI erkennt Datumsspalten automatisch, wenn mindestens 80% der Werte erfolgreich interpretiert werden können. Zweistellige Jahresangaben (z. B. 06/15/25) werden als das nächstliegende passende Jahr gelesen, sodass aktuelle MLS-Exporte korrekt eingelesen werden; vierstellige Jahresangaben werden stets wörtlich übernommen.
- **Numerische Spalten**: Verkaufspreis, Wohnfläche, Grundstücksgröße, Zugeständnisse und ähnliche Felder müssen numerisch sein. Exportiert das MLS Preise mit Währungssymbolen oder Tausendertrennzeichen (z. B. „\$350,000“ oder „350.000,00“), müssen diese gegebenenfalls vor dem Import bereinigt werden.
- **Faktorspalten**: Kategoriale Variablen wie Area ID, Stil oder Zustand sollten eine überschaubare Anzahl eindeutiger Ausprägungen enthalten. earthUI erkennt Faktoren automatisch, die Erkennung kann jedoch übersteuert werden.

Tipp: Nach dem Import empfiehlt sich ein Blick auf die T/F/NA-Spalte in der Prädiktortabelle — sie zeigt die Anzahl der true-/false-/NA-Werte je Spalte. Spalten mit mehr als 30% fehlenden Werten werden rot markiert. Spalten mit hohem NA-Anteil sollten ausgeschlossen oder die Daten vor dem Import bereinigt werden.

2.5 Platzierung der Zeile des Bewertungsobjekts

Im **Appraisal-Modus** (Gutachtenmodus) **muss Zeile 1 das Bewertungsobjekt enthalten**. Alle übrigen Zeilen sind Vergleichsverkäufe. Der Verkaufspreis des Bewertungsobjekts kann leer gelassen (NA) oder auf einen beliebigen Wert gesetzt werden — earthUI behandelt ihn während der Modellanpassung in jedem Fall als NA.

Im **Market Area Analysis-Modus** (Marktgebietsanalyse) ist die Platzierung eines Bewertungsobjekts in Zeile 1 optional, und es wird keine Zeile automatisch ausgeschlossen. Um

die Zeile des Bewertungsobjekts (oder eine beliebige andere Zeile) aus der Modellanpassung herauszuhalten, ist ihr in einer dafür vorgesehenen Gewichtsspalte der Wert null zuzuweisen oder ein Subset-Filter zu verwenden.

Im **General**-Modus (allgemeiner Modus) gibt es keine besondere Zeilenbehandlung — alle Zeilen werden gleich behandelt.

3 General Purpose Mode

3.1 Überblick

Der General Purpose Mode (Allzweckmodus) ist der Standardmodus beim Start von earthUI. Er stellt den vollständigen MARS-Modellierungs-Arbeitsablauf für beliebige Datensätze bereit — nicht nur für Immobiliendaten. earthUI kann für wissenschaftliche Daten, Finanzanalysen, ingenieurwissenschaftliche Studien oder jedes Regressionsproblem eingesetzt werden, bei dem nichtlineare Zusammenhänge und Interaktionen zwischen Variablen untersucht werden sollen.

Im General Mode blendet die Oberfläche die immobilienpezifischen Funktionen aus (Spezialspalten, Verkaufsalter, Koordinatenrundung, RCA). Die Sidebar ist gestrafft und konzentriert sich auf Variablenauswahl, Parameterkonfiguration, Modellanpassung und Export.

3.2 Der Arbeitsablauf in der Sidebar

Die Sidebar ist in nummerierte, einklappbare Abschnitte gegliedert, die von der Projektauswahl bis zum Export führen:

1. Project — Auswahl eines bestehenden Projekts oder Anlage eines neuen Projekts über den New Project Wizard (Assistent für neue Projekte), der Zweck, Land und Verwaltungsstandort erfasst (nach der Anlage sämtlich gesperrt). Das aktive Projekt bestimmt, von wo Eingabedateien gelesen und wohin alle Ausgaben geschrieben werden — ein separates Feld für den Ausgabeordner existiert nicht.

2. Import Data — Auswahl einer CSV- oder Excel-Datei aus dem Ordner `in/` des aktiven Projekts; ein Refresh-Link durchsucht den Ordner erneut nach manuell hinzugefügten Dateien. Bei Excel-Dateien mit mehreren Blättern erscheint eine Blattauswahl. Spaltennamen werden automatisch in `snake_case` umgewandelt.

3. Variable Configuration — Auswahl der Zielvariablen (mehrere Zielvariablen werden unterstützt), Effective Date (Wertermittlungstichtag) sowie die Prädiktortabelle mit Kontrollkästchen für Include, Factor, Linear und Latent. Ausführliche Einzelheiten enthält Kapitel 6.

4. Earth Call Parameters — Alle Argumente der Funktion `earth()`: degree, penalty, nk, Pruning-Methode, Kreuzvalidierung und weitere. Die vollständige Parameterreferenz enthält Kapitel 7.

Execute Prolog Processing — Ein optionaler, unnummerierter Abschnitt zwischen den Abschnitten 4 und 5, der deaktiviert ist, sofern die experimentelle Prolog-Funktion nicht in den Settings aktiviert wurde. Siehe Kapitel 15.

5. Fit Earth Model — Ein Feld für den Zufallsstartwert (random seed) und die Schaltfläche, die die Modellanpassung asynchron ausführt. Einzelheiten zur Anpassung enthält Kapitel 8.

6. Download Estimated Target Variable(s) & Residuals — (in den Modi für Wertermittlung und Marktanalyse mit „Download Estimated Sale Prices & Residuals“ betitelt) Exportiert Vorhersagen, Residuen, CQA-Scores und die Wertbeiträge je g-Funktion als Excel-Datei. Siehe Kapitel 9.

7–8. RCA Adjustments und Sales Grid — (Vergleichsraster) Nur im Appraisal Mode. Siehe Kapitel 10 und 11.

9. Generate Quarto Report — Schreibt ein in sich geschlossenes Quarto-Berichtsbündel (`.qmd`-Quelldatei nebst Assets) in den Ausgabeordner des Projekts. Außerhalb des Appraisal Mode mit 7 nummeriert. Siehe Kapitel 12.

10. Convert Quarto Report — Rendert eine beliebige `.qmd`-Datei nach HTML, Word oder PDF. Siehe Kapitel 12.

3.3 Tabs des Hauptpanels

Nach der Modellanpassung stellt das Hauptpanel zehn Tabs bereit:

Tab	Inhalt
Data	Vorschau der importierten Daten. Im Appraisal Mode aufgeteilt in Tabellen für das Bewertungsobjekt (Subject Property) und die Vergleichsverkäufe (Comparable Sales).
Equation	Die angepasste Modellgleichung, dargestellt in LaTeX/MathJax, mit jeder Basisfunktion und ihrem Koeffizienten.
Summary	Zentrale Kennzahlen (R^2 , GRSq, GCV, RSS, CV R^2) sowie eine Koeffiziententabelle.
Variable Importance	Balkendiagramm und Rangtabelle der Wichtigkeitswerte der Prädiktoren.
Contribution	Diagramme der g-Funktions-Wertbeiträge. Auswahl einer g-Funktion über das Dropdown-Menü; univariate Prädiktoren werden als Liniendiagramme dargestellt, bivariate Interaktionen ($\text{degree} \geq 2$) als 3D-Perspektiv- und Konturdiagramme.
Correlation	Heatmap der Prädiktorkorrelationen.
Diagnostics	Drei Diagramme: Residuals vs Fitted, Normal Q-Q und Actual vs Predicted.
RCA Adjustments	(nur Wertermittlung) Histogramme von Residual Adj %, Net Adj % und Gross Adj % über die Vergleichsobjekte hinweg, befüllt nach Abschluss von Schritt 7 (Calculate RCA Adjustments).
Residual Contribution	Schätzt die Marktwertbeiträge seltener oder ausgeschlossener Merkmale, indem die Modellresiduen gemeinsam auf Spalten regressiert werden, die das Modell nie verwendet hat (siehe subsection 4.4).
ANOVA	ANOVA-Zerlegungstabelle, die den Beitrag jeder Basisfunktion zur RSS ausweist.
Earth Output	Unformatierte Textausgabe von 'earth::summary()', einschließlich der Details des Pruning-Durchlaufs.

3.4 Persistenz der Einstellungen

earthUI speichert die Konfiguration serverseitig in einer Projektdatenbank (`projects.sqlite` im `regProj`-Stammverzeichnis), indiziert nach Projekt und Zweck. Das Speichern erfolgt **schaltflächen-gesteuert**: Die Schaltfläche „Save current as default“ in Abschnitt 3 speichert Zielvariable, Stichtag und Prädiktorkonfiguration, während die Schaltfläche in Abschnitt 4 die Parameter des earth-Aufrufs sowie die Allowed-Interactions-Matrix speichert. Jede Schaltfläche aktualisiert nur ihre eigene Hälfte, sodass sie einander niemals überschreiben, und **die Modellanpassung speichert nicht automatisch** — es kann frei experimentiert werden, ohne die gespeicherten Standardwerte zu überschreiben. Die verschiedenen Eingabedateien eines Projekts (ein kleiner Testauszug, der vollständige Datensatz) teilen sich eine Konfiguration, und da die Einstellungen im Projektbaum liegen, wandern sie mit, wenn der `regProj`-Ordner synchronisiert, gesichert oder weitergegeben wird.

3.5 Dark Mode

Ein Klick auf das Mond-/Sonnensymbol in der oberen rechten Ecke wechselt zwischen hellem und dunklem Design. Die Design-Präferenz wird im lokalen Speicher (Local Storage) abgelegt und bleibt über Sitzungen hinweg erhalten.

3.6 Locale & regionale Einstellungen

earthUI unterstützt internationale Formatierungskonventionen für Zahlen, Datumsangaben und CSV-Dateien über ein länderbasiertes Locale-System. Das Dropdown-Menü **Country** im Menü **Settings** (Zahnradsymbol) der oberen Menüleiste wählt eine Voreinstellung für 31 unterstützte Länder aus. Jede Voreinstellung konfiguriert:

- **CSV-Trennzeichen** — Komma (,) für USA/UK/Japan oder Semikolon (;) für den Großteil Europas, wo das Komma als Dezimalzeichen dient.
- **Dezimalzeichen** — Punkt (.) oder Komma (,).
- **Tausendertrennzeichen** — Komma (USA/UK/Japan), Punkt (Deutschland/Italien/Spanien), Leerzeichen (Finnland/Frankreich/Polen/Baltikum/Ukraine/Russland) oder Apostroph (Schweiz).
- **Datumsformat** — MM/DD/YYYY (USA), DD/MM/YYYY (Großteil Europas) oder YYYY-MM-DD (Schweden/Litauen/Japan/Kanada).
- **Papierformat** — Letter (USA/Kanada/Mexiko) oder A4 (überall sonst).

3.6.1 Unterstützte Länder

Land	CSV-Trennz.	Dezimal	Tausender	Datum	Papier
USA, Kanada (csv)	,	.	,	MDY/YMD	Letter
UK, Irland, AU, NZ	,	.	,	DMY	A4
Deutschland, Österreich, Belgien, Niederlande, Italien, Spanien, Portugal, Dänemark, Brasilien	;	,	.	DMY	A4
Frankreich, Polen, Tschechien, Norwegen, Lettland, Estland	;	,	(Leerzeichen)	DMY	A4
Finnland, Schweden	;	,	(Leerzeichen)	DMY/YMD	A4
Schweiz	;	,	,	DMY	A4
Ukraine, Russland	;	,	(Leerzeichen)	DMY	A4
Türkei	;	,	.	DMY	A4
Japan, Südkorea	,	.	,	YMD	A4
Mexiko	,	.	,	DMY	Letter
Litauen	;	,	(Leerzeichen)	YMD	A4

3.6.2 Übersteuerung des Papierformats

Unterhalb der Länderauswahl ermöglicht ein Dropdown-Menü **Paper** (Letter oder A4), das Seitenformat des Berichts zu übersteuern, ohne das Land zu wechseln. CSV-Trennzeichen, Dezimalzeichen und die Reihenfolge des Datumsformats folgen der Voreinstellung des gewählten Landes; beim Wechsel des Landes wird die Papierformat-Übersteuerung auf den Standardwert dieses Landes zurückgesetzt.

3.6.3 Speichern der Standardeinstellungen

Ein Klick auf **Save** im Settings-Menü speichert die Locale-Präferenzen global. Diese Standardwerte gelten für alle künftigen Sitzungen, unabhängig davon, welches Projekt geöffnet wird. Projektspezifische Einstellungen (Zielvariable, Prädiktoren, Parameter) werden separat in der Projektdatenbank gespeichert; die Locale-Standardwerte bleiben über alle Projekte hinweg in einer benutzerbezogenen SQLite-Datenbank erhalten. Dieser zweistufige Ansatz ist für Organisationen wie Wirtschaftsprüfungsgesellschaften konzipiert, die mit Daten aus mehreren Ländern arbeiten — das am häufigsten benötigte Land wird als Standard festgelegt und bei Bedarf angepasst.

Hinweis: Die Zahlenformatierung auf Diagrammachsen und Steigungsbeschriftungen passt sich automatisch an das eingestellte Locale an. Das deutsche Locale verwendet Punkte als Tausendertrennzeichen (200.000), das finnische Leerzeichen (200 000), das schweizerische Apostrophe (200'000). Währungssymbole werden nicht angezeigt — earthUI ist währungsneutral.

4 Appraisal Mode (Gutachtenmodus)

Wird als Purpose (Zweck) die Option **For Appraisal** gewählt, konfiguriert sich earthUI für die Bewertung eines einzelnen Objekts. Alle in Kapitel 3 beschriebenen Funktionen bleiben verfügbar; das vorliegende Kapitel behandelt ausschließlich die gutachtenspezifischen Ergänzungen.

4.1 Behandlung der Subject-Zeile

Im Appraisal Mode ist **Zeile 1 des Datensatzes das Bewertungsobjekt**, und alle übrigen Zeilen sind Vergleichsverkäufe. Die Eingabedatei muss entsprechend organisiert sein (siehe Kapitel 2). Der Verkaufspreis des Bewertungsobjekts kann leer gelassen oder auf einen beliebigen Wert gesetzt werden — earthUI behandelt ihn während der Anpassung automatisch als NA.

Nach dem Import teilt sich der Data tab in zwei Abschnitte: **Subject Property** (Zeile 1) und **Comparable Sales** (Zeilen 2+). Zeile 1 wird stets von der Modellanpassung ausgeschlossen — die Benachrichtigung „Skipping row 1 (subject). Fitting on N rows.“ bestätigt dies. Nach der Anpassung erzeugt das Modell dennoch Prognosen für die Subject-Zeile, die in der Ausgabe als `est_<target>` ausgewiesen werden.

4.2 Effective Date & Sale Age

Ein Feld **Effective Date** (Wertermittlungstichtag) erscheint im Abschnitt Variable Configuration in jedem Zweckmodus (mit dem heutigen Datum als Voreinstellung) — auch allgemeine Analysen und Marktanalysen verfügen über einen Stichtag, nicht nur Gutachten —, *verwendet* wird es jedoch nur im Appraisal Mode und im Market Mode. Wird dort eine Spalte im Dropdown-Menü Special als `contract_date` ausgewiesen, berechnet earthUI eine Spalte `sale_age` — die Anzahl ganzer Tage zwischen dem Vertragsdatum des jeweiligen Verkaufs und dem Stichtag. Diese Spalte ersetzt die ursprüngliche Datumsspalte als Prädiktor.

Beim ersten Klick auf Fit nach der Ausweisung eines Vertragsdatums erstellt earthUI die Spalte `sale_age` und weist darauf hin, dass erneut auf Fit zu klicken ist, um sie einzubeziehen.

4.3 Special-Spalten-Ausweisungen

Im Appraisal Mode und im Market Mode erscheint für jeden Prädiktor in der Tabelle Variable Configuration ein Dropdown-Menü **Special**. Die vollständige Liste der Special-Typen und ihrer Wirkungen:

Special-Typ	Wirkung
no	Voreinstellung — keine besondere Behandlung
contract_date	Löst die automatische Berechnung von <code>sale_age</code> aus dem Effective Date aus
listing_date	Wird zusammen mit dem Vertragsdatum zur Berechnung von DOM verwendet, wenn keine Spalte <code>dom</code> existiert
dom	Kennzeichnet die Spalte „Days on Market“; wird im Sales Grid angezeigt
concessions	Kennzeichnet Verkaufszugeständnisse; wird in der Net-SP-Formel (Verkaufspreis – Zugeständnisse) im Sales Grid verwendet
latitude	Werte werden auf 3 Dezimalstellen gerundet; verwendet für die Haversine-Näheberechnung und die Gruppe Location im Sales Grid
longitude	Werte werden auf 3 Dezimalstellen gerundet; verwendet für die Haversine-Näheberechnung und die Gruppe Location im Sales Grid
area	Gruppiert in der Zeile LLoc: Long Lat Area im Sales Grid
living_area	Aktiviert Residuen je Quadratfuß (<code>residual_sf</code> , <code>cqa_sf</code>) in der Download-Ausgabe
lot_size	Gruppiert in der Zeile SSite Size Dimensions im Sales Grid
site_dimensions	Gruppiert mit der Grundstücksgröße im Sales Grid
actual_age	Gruppiert in der Zeile Actual Age Effective Age im Sales Grid
effective_age	Gruppiert mit dem tatsächlichen Alter im Sales Grid
display_only	Spalte wird in Exporte einbezogen, jedoch von der Modellanpassung ausgeschlossen (mehrfach zulässig)
remarks	(Experimentell) Freitext-Bemerkungsspalte, die durch den optionalen Prolog-Schritt in Merkmalsspalten zerlegt wird (mehrfach zulässig); siehe Kapitel 15
list	(Experimentell) Kommagetrennte Spalte, die durch den optionalen Prolog-Schritt in One-Hot-Spalten mit TRUE/FALSE aufgeteilt wird (mehrfach zulässig); siehe Kapitel 15

Pro Special-Typ ist nur eine Spalte zulässig (mit Ausnahme von `display_only`, `remarks` und `list`). Wird ein Special-Typ einer zweiten Spalte zugewiesen, wird er automatisch von der ersten entfernt. Neben dem Variablennamen erscheint eine kleine blaue Kennzeichnung, die den zugewiesenen Special-Typ anzeigt.

4.4 Residualer Wertbeitrag (Schätzung seltener Merkmale)

Merkmale, die für den Regressionskanal zu selten sind — etwa eine Werkstatt, die nur in einer Handvoll Verkäufe vorkommt —, gehören zur Residualschicht des RCA. Der Tab **Residual Contribution** (residualer Wertbeitrag) schätzt ihren Marktbeitrag, indem die Residuen des angepassten Modells mittels gewöhnlicher Kleinst-Quadrat-Schätzung auf Spalten regressiert werden, die das Modell nie verwendet hat. Die Auswahl mehrerer Spalten schätzt diese *gemeinsam*, sodass korrelierte Ausstattungsmerkmale (z. B. `outbuilding_sqft` zusammen mit `pr_has_workshop`) den Wertbeitrag untereinander aufteilen, statt ihn doppelt zu zählen. Binäre Indikatoren liefern Pauschalschätzungen und führen das Residuum jedes gekennzeichneten Verkaufs als Beleg auf; numerische Spalten liefern Wertbeiträge je Einheit; der ausgewiesene Standardfehler entspricht einem SE ($t \geq 2$ als üblicher Signifikanzmaßstab). Terme, die auf weniger als zehn Beobachtungen beruhen, erhalten automatisch den Vermerk CAUTION in der erzeugten Arbeitsdatei-Notiz, die Methode, Schätzwerte und Vorbehalte in einem Absatz für das Gutachten dokumentiert. Von der Kandidatenliste ausgeschlossen werden nur Prädiktoren, die das Modell tatsächlich *verwendet* hat — ein seltener Indikator, der der Anpassung angeboten, vom Vorwärtsdurchlauf jedoch nie ausgewählt wurde, bleibt wählbar, da das Modell keinen Teil

seines Wertes absorbiert hat.

4.5 Überblick über die RCA-Anpassungen

Die Schaltfläche **Calculate RCA Adjustments & Download** (Seitenleistenabschnitt 7, nur im Appraisal Mode nach der Anpassung sichtbar) berechnet marktabgeleitete Anpassungen für jedes Vergleichsobjekt relativ zum Bewertungsobjekt. Der vollständige RCA-Arbeitsablauf ist in Kapitel 10 beschrieben.

Nach der Berechnung der RCA-Anpassungen wird die Schaltfläche **Generate Sales Grid & Download** (Seitenleistenabschnitt 8) verfügbar. Der Sales-Grid-Arbeitsablauf ist in Kapitel 11 beschrieben.

***Hinweis:** Der dem Bewertungsobjekt zugewiesene CQA-Wert steuert, welcher Anteil der Residualverteilung dem Bewertungsobjekt zugerechnet wird. Ein Wert von 5.00 platziert das Bewertungsobjekt am Median der Vergleichsobjekte.*

5 Market Area Analysis Mode (Marktgebietsanalysemodus)

Wird als Purpose die Option **Market Area Analysis** gewählt, stellt earthUI dieselben immobilien-spezifischen Funktionen wie der Appraisal Mode bereit (Special-Spalten, Verkaufsalter, Rundung von Koordinaten), ist jedoch auf die Analyse einer Gruppe von Objekten ausgerichtet und nicht auf die Bewertung eines einzelnen Bewertungsobjekts.

5.1 Unterschiede zum Appraisal Mode

- **Kein automatischer Ausschluss des Bewertungsobjekts** — anders als der Appraisal Mode schließt der Market Mode Zeile 1 nicht automatisch von der Anpassung aus. Alle Zeilen werden einbezogen; um eine Subject-Zeile aus der Anpassung herauszuhalten, ist ihr in einer ausgewiesenen Gewichtsspalte der Wert null zuzuweisen (wie im Demonstrationsdatensatz) oder ein Teilmengenfilter zu verwenden.
- **Keine RCA- oder Sales-Grid-Abschnitte** — die Schritte „Calculate RCA Adjustments & Download“ und „Generate Sales Grid & Download“ sind nicht verfügbar. Der Market Mode konzentriert sich auf Modellanpassung und Ausgabe, nicht auf Anpassungen je Vergleichsobjekt.
- **Nummerierung der Seitenleiste** — ohne die RCA- und Sales-Grid-Schritte trägt der Abschnitt Generate Quarto Report die Nummer 7 statt 9 (der Abschnitt Convert Quarto Report behält die Nummer 10).

5.2 Wann der Market Mode zu verwenden ist

Der Market Area Analysis Mode ist angemessen, wenn

- ein Regressionsmodell für ein Wohnviertel oder Marktgebiet erstellt werden soll, um Werttreiber zu verstehen,
- analysiert werden soll, wie Variablen wie Wohnfläche, Alter, Grundstücksgröße und Lage die Verkaufspreise über eine Gruppe von Objekten hinweg beeinflussen,
- Belegmaterial für eine Marktbedingungsanalyse oder eine Abgrenzung von Wohnvierteln vorbereitet werden soll,
- mit einem Datensatz gearbeitet wird, der ein Bewertungsobjekt in Zeile 1 enthält, dessen Einbeziehung in die oder Ausschluss aus der Anpassung jedoch wahlweise möglich bleiben soll.

Hinweis: Der Market Mode eignet sich auch für allgemeine Immobilienregressionen, bei denen die Special-Spalten-Funktionen (Verkaufsalter, Rundung von Koordinaten) benötigt werden, der RCA-Anpassungsablauf jedoch nicht.

6 Variablenauswahl

In Abschnitt 3 der Seitenleiste — **Variable Configuration** (Variablenkonfiguration) — wird festgelegt, welche Spalten in das Modell eingehen und wie sie behandelt werden.

6.1 Zielvariable(n)

Das Dropdown-Menü **Target (response) variable(s)** (Ziel- bzw. Antwortvariable(n)) am Anfang von Abschnitt 3 führt jede Spalte des Datensatzes auf. Für ein gewöhnliches Modell mit einer einzelnen Antwortvariablen wird eine Spalte ausgewählt, für ein Mehrfachantwort-Modell mehrere Spalten. Werden mehrere Zielvariablen ausgewählt, passt das Modell alle Antwortvariablen simultan mittels `earth(cbind(y1, y2, ...) ~ .)` an.

Als Zielvariablen ausgewählte Spalten werden automatisch aus der Liste der Prädiktoren ausgeschlossen.

6.2 Die Prädiktortabelle

Unterhalb der Zielvariablenauswahl führt eine Tabelle jede verbleibende Spalte mit den folgenden Feldern auf:

Spalte	Beschreibung
Variable	Spaltenname (bei Abschneidung wird der vollständige Name im Tooltip angezeigt), dem ein $\times$ / $+$ -Umschalter vorangestellt ist (siehe unten). In den Modi Wertermittlung/Markt zeigt ein blaues Abzeichen den zugewiesenen Spezialtyp an.
Type	Dropdown-Menü für den Datentyp: <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>logical</code> , <code>factor</code> , <code>Date</code> , <code>POSIXct</code>
Include	Kontrollkästchen — diese Spalte als Prädiktor in das Modell aufnehmen
Factor	Kontrollkästchen — diese Spalte als kategoriale Variable behandeln
Linear	Kontrollkästchen — ausschließlich linearen Eingang erzwingen (keine Hinge-Funktionen)
Latent	Kontrollkästchen — diese Spalte vollständig aus dem Modell heraushalten (ausgeschlossen aus <code>earth</code> und aus den nach <code>glmnetUI</code> / <code>mgcvUI</code> exportierten Daten) und für Arbeiten mit latenten Variablen vorbehalten
Special	Dropdown-Menü (nur Wertermittlung/Markt) — siehe die nachstehende Referenz der speziellen Spaltentypen
T/F/NA	Drei Zählwerte je Spalte: bei logischen Spalten TRUE/FALSE/NA; bei numerischen Spalten $\neq 0$ / $= 0$ /NA. Rot dargestellt, wenn mehr als 30 % der Werte fehlen

Eine Hinweiszeile oberhalb der Tabelle erläutert die Abkürzungen: „Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`).“

Weiches Deaktivieren statt Löschen. Der $\times$ -Umschalter neben jedem Variablennamen graut die Zeile aus und hält die Spalte aus dem Modell heraus, ohne sie aus den Daten zu entfernen; er wechselt zu einem grünen $+$, das die Spalte wieder aktiviert. Spalten werden niemals zerstört.

6.3 Datentyperkennung und Überschreibungen

earthUI erkennt Datentypen beim Import automatisch. Numerische, ganzzahlige, logische, faktorielle und Datumsspalten werden erkannt. Zeichenspalten, die wie Datumsangaben aussehen (mindestens 80 % der Werte lassen sich gegen gängige Datumsformate parsen), werden als **Date** klassifiziert.

Jede Erkennung kann durch Änderung des Dropdown-Menüs **Type** überschrieben werden. Wird eine Spalte auf **character** oder **factor** geändert, wird das Kontrollkästchen Factor automatisch aktiviert. Typänderungen wirken sich darauf aus, wie die Spalte an die Funktion `earth()` übergeben wird.

6.4 Die Markierungen Factor und Linear

- **Factor** — Ist das Kontrollkästchen aktiviert, wird die Spalte als kategoriale Variable behandelt. Dies ist angemessen für Spalten wie **style**, **area_id** oder **grade**, die diskrete Gruppen und keine stetigen Messgrößen darstellen. Faktorspalten gehen als Indikatorvariablen (Dummy-Variablen) in das Modell ein.
- **Linear** — Ist das Kontrollkästchen aktiviert, wird der Spalte ein streng linearer Eingang in das Modell auferlegt — es werden für sie keine Hinge-Funktionen (stückweise lineare Funktionen) erzeugt. Dies ist nützlich, wenn bekannt ist, dass eine Variable in einem streng linearen Zusammenhang mit der Zielvariablen steht.

6.5 Referenz der speziellen Spaltentypen

In den Modi Wertermittlung und Markt stellt das Dropdown-Menü **Special** die folgenden Optionen bereit. Jeder Typ kann höchstens einer Spalte zugewiesen werden (mit Ausnahme von **display_only**, **remarks** und **list**, die mehrfach zulässig sind):

Datums- und Zeittypen:

- **contract_date** — Löst die automatische Berechnung von **sale_age** aus. Die ursprüngliche Datumsspalte wird durch eine ganzzahlige Spalte ersetzt, welche die Anzahl der Tage zwischen dem Verkaufsdatum und dem Wertermittlungstichtag (Effective Date) misst.
- **listing_date** — Dient als Rückfalloption zur Berechnung der Vermarktungsdauer (Days on Market, DOM = Vertragsdatum – Listing-Datum), wenn keine explizite **dom**-Spalte bestimmt ist.
- **dom** — Kennzeichnet die Spalte der Vermarktungsdauer (Days on Market). Wird in der APN-Zeile und der Verkaufsdatumszeile des Sales Grid (Vergleichsraster) angezeigt.

Monetäre Typen:

- **concessions** — Kennzeichnet Verkaufszugeständnisse (Verkäufergutschriften, Käuferanreize usw.). Wird im Sales Grid zur Berechnung des Nettoverkaufspreises verwendet: $\text{Net SP} = \text{Verkaufspreis} - \text{Zugeständnisse}$.

Größen- und Lagetypen:

- **latitude** — Die Werte werden automatisch auf 3 Dezimalstellen gerundet, um Überanpassung zu vermeiden. Wird für Haversine-Näherungsberechnungen (Entfernung vom Bewertungsobjekt zu jedem Vergleichsobjekt) verwendet und in der Lagezeile des Sales Grid gruppiert.
- **longitude** — Gleiche Rundungsbehandlung wie die geographische Breite. Wird für die Entfernungsberechnung und die Lagegruppe verwendet.
- **area** — Typischerweise eine Nachbarschafts- oder Gebietskennung. Wird gemeinsam mit geographischer Breite und Länge in der Zeile „Loc: Long | Lat | Area“ des Sales Grid gruppiert.

- **living_area** — Ermöglicht Residualberechnungen je Quadratfuß (**residual_sf** und **cqa_sf**) in der Download-Ausgabe.
- **lot_size** — Wird in der Zeile „Site Size | Dimensions“ des Sales Grid gruppiert.
- **site_dimensions** — Wird gemeinsam mit der Grundstücksgröße im Sales Grid gruppiert (z. B. „75x120“).

Alterstypen:

- **actual_age** — Wird in der Zeile „Actual Age | Effective Age“ des Sales Grid gruppiert.
- **effective_age** — Wird gemeinsam mit dem tatsächlichen Alter im Sales Grid gruppiert.

Anzeigetypen:

- **display_only** — Die Spalte wird in Excel-Exporte aufgenommen, jedoch vollständig von der Modellanpassung ausgeschlossen. Diese Kennzeichnung eignet sich für Adressfelder, MLS-Nummern, Flurstückskennungen oder sonstige Referenzdaten, die nicht als Prädiktor dienen sollen. Mehrere Spalten können diese Kennzeichnung tragen.

Typen zur Textextraktion (experimentell):

- **remarks** — Eine Freitextspalte (z. B. **public_remarks**), die durch den optionalen Prolog-Schritt in Merkmalspalten zerlegt wird. Ohne Aktivierung dieser Funktion in den Settings bleibt sie wirkungslos. Mehrere Spalten zulässig. Siehe Kapitel 15.
- **list** — Eine kommasetrennte Spalte (z. B. **existing_structures**), die durch denselben Schritt in One-Hot-Spalten der Form `<col>_<value>` mit TRUE/FALSE-Werten aufgespalten wird. Mehrere Spalten zulässig. Siehe Kapitel 15.

6.6 Mehrere Zielvariablen

Die Auswahl mehrerer Zielvariablen führt zur Anpassung eines Earth-Modells mit mehreren Antwortvariablen. Bei Auswahl mehrerer Zielvariablen gilt:

- Das Modell prognostiziert alle Antwortvariablen simultan, wobei die Basisfunktionen über die Zielvariablen hinweg geteilt werden
- Die Schaltfläche **wp (response weights)** (Antwortgewichte) wird aktiv und erlaubt die Zuweisung eines numerischen Gewichts an jede Zielvariable
- Varianzmodelle (**varmod.method**) sind deaktiviert (für Modelle mit mehreren Antwortvariablen nicht unterstützt)
- Die Ergebnisreiter zeigen Kennzahlen, Gleichungen und Diagnosedigramme je Antwortvariable an

Hinweis: Als **display_only** gekennzeichnete Spalten verbleiben im Datensatz und erscheinen im Excel-Export, werden jedoch vollständig aus der Prädiktortabelle ausgeschlossen. Diese Kennzeichnung eignet sich für ID-Spalten, Adressen oder sonstige Referenzfelder.

7 Parameterauswahl

Abschnitt 4 der Seitenleiste — **Earth Call Parameters** (Parameter des Earth-Aufrufs) — bietet Zugriff auf alle Argumente, die von der Funktion `earth()` akzeptiert werden. Jeder Parameter verfügt über ein blaues Hilfesymbol (?) mit einer Tooltip-Erläuterung. Die Parameter sind in einklappbare Unterabschnitte gegliedert.

7.1 Grundparameter

Parameter	Standardwert	Beschreibung
subset	(empty)	Ausdruck zur Zeilenfilterung. Siehe „Subset-Filterung“ weiter unten.
weights	NULL	Spaltenauswahl für Fall-(Zeilen-)Gewichte. Es werden nur numerische Spalten aufgeführt.
wp	NULL	Antwortgewichte für Modelle mit mehreren Zielvariablen. Die Schaltfläche öffnet einen Dialog mit einem numerischen Eingabefeld je Zielvariable (Standard jeweils 1.0).
keepxy	off	Behält x, y, subset und weights im Modellobjekt.
trace	0	Trace-Stufe für die Ausgabe während der Anpassung (0 bis 5).
glm	none	Optionale GLM-Familie: gaussian , binomial oder poisson .
degree	1	Maximale Interaktionsordnung. Ein Wert $\text{degree} \geq 2$ aktiviert automatisch die Kreuzvalidierung und blendet die Matrix Allowed Interactions (zulässige Interaktionen) ein.
penalty	2	GCV-Strafterm je Knoten. Höhere Werte führen zu einfacheren Modellen.

7.2 Vorwärtsthroughlauf (Forward Pass)

Parameter	Standardwert	Beschreibung
nk	auto	Maximale Anzahl an Termen vor dem Pruning. Wird automatisch mit dem für die geladenen Daten empfohlenen Wert vorbelegt (siehe „Empfohlene Parameterwerte“ weiter unten).
thresh	0.001	Schwellenwert des Vorwärtsschritts. Kleinere Werte lassen mehr Terme zu.

Parameter	Standardwert	Beschreibung
minspan	auto	Minimale Spannweite zwischen Knoten. Wird automatisch mit dem empfohlenen Wert vorgelegt. Negative Werte legen die maximale Knotenzahl je Prädiktor fest.
endspan	auto	Randspannweite — minimaler Abstand eines Knotens vom Rand der Daten. Wird automatisch mit dem empfohlenen Wert vorgelegt.
newvar.penalty	0	Strafterm für die Einführung einer neuen Variablen (begünstigt die Wiederverwendung bereits vorhandener Prädiktoren).
fast.k	20	Anzahl der im Fast-MARS-Algorithmus berücksichtigten Elternterme.
fast.beta	1	Steuert den Alterungsfaktor des Fast-MARS-Algorithmus.

7.3 Allowed Interactions (zulässige Interaktionen)

Wird **degree** auf 2 oder höher gesetzt, erscheint unterhalb der Grundparameter eine Interaktionsmatrix. Dabei handelt es sich um ein oberes Dreiecksraster aus Kontrollkästchen, eines für jedes Prädiktorpaar. Ein gesetztes Häkchen bedeutet, dass die beiden Prädiktoren interagieren dürfen; wird es entfernt, ist diese spezifische Interaktion untersagt.

Ein Klick auf einen Prädiktornamen (Zeilen- oder Spaltenbeschriftung) schaltet sämtliche Interaktionen dieser Variablen um. Die Kontrollkästchen **Allow All** und **Clear All** am oberen Rand ermöglichen eine Sammelsteuerung. Die Matrix verwendet fixierte Kopfzeilen (Sticky Headers), sodass Spalten- und Zeilenbeschriftungen beim Scrollen sichtbar bleiben.

Ein Hinweissfeld erinnert daran: „Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).“ (Interaktionsterme erhöhen das Risiko der Überanpassung; die Kreuzvalidierung wurde aktiviert, 10-fach.)

7.4 Pruning

Parameter	Standardwert	Beschreibung
pmethod	backward	Pruning-Methode: backward , none , exhaustive , forward , seqrep oder cv .
nprune	NULL	Maximale Anzahl an Termen nach dem Pruning. Für die automatische Auswahl leer lassen.

7.5 Kreuzvalidierung

Parameter	Standardwert	Beschreibung
nfold	10	Anzahl der Folds (Teilstichproben) der Kreuzvalidierung. Ein Wert von 0 deaktiviert die CV.
ncross	20	Anzahl der Wiederholungen der Kreuzvalidierung.
stratify	on	Stratifiziert die CV-Stichproben, sodass jeder Fold eine ähnliche Verteilung der Zielvariablen aufweist.

Bei $\text{degree} \geq 2$ wird die Kreuzvalidierung automatisch aktiviert (nfold auf 10 gesetzt), um einer Überanpassung durch Interaktionen vorzubeugen.

7.6 Subset-Filterung

Das Texteingabefeld **subset** akzeptiert einen R-Ausdruck, der festlegt, welche Zeilen für die Modellanpassung verwendet werden. Der Ausdruck kann direkt eingegeben werden (z. B. `sale_age < 365 & area_id == 460`) oder mithilfe der Schaltfläche **Build filter...** visuell erstellt werden.

Build Filter Dialog — Ein Klick auf „Build filter.“ öffnet einen geführten Dialog. Jede Bedingungszeile besitzt eine Spaltenauswahlliste, einen Operator (`<`, `>`, `<=`, `>=`, `==`, `!=`) sowie ein Werteingabefeld, das sich dem Spaltentyp anpasst: numerische Eingabe für Zahlen, Datumsauswahl für Datumsangaben und eine Auswahlliste der eindeutigen Werte für Zeichen- bzw. Faktorspalten. Die Bedingungen werden mit AND- (&) oder OR-Verknüpfungen (|) kombiniert. Eine Vorschau am unteren Rand zeigt den Ausdruck sowie die Anzahl der übereinstimmenden Zeilen. Ein Klick auf **Apply** fügt den Ausdruck in das Texteingabefeld ein.

Datumsspalten müssen in manuell erstellten Ausdrücken mit `as.Date("...")` oder `as.POSIXct("...")` umschlossen werden. Der Build-Filter-Dialog erledigt dies automatisch.

Die Subset-Filterung ist nicht-destruktiv — ausgeschlossene Zeilen verbleiben im Datensatz und erhalten im Excel-Export Prognosewerte.

***Tipp:** Beim Kombinieren mit AND sollten OR-Bedingungen durch Klammern gruppiert werden. Ohne Klammern wird $A \& B \mid C$ als $(A \& B) \mid C$ ausgewertet, was unter Umständen nicht dem beabsichtigten Ergebnis entspricht.*

7.7 Empfohlene Parameterwerte

earthUI zeigt unterhalb zentraler Parameter in Abschnitt 4 einen empfohlenen Wert an. Diese Empfehlungen werden reaktiv auf Grundlage der Anzahl der für die Anpassung verwendeten Zeilen (n) und der ausgewählten Prädiktoren (p) aktualisiert. Die Formeln beruhen auf Friedmans MARS-Aufsatz, den internen Algorithmen von earth sowie empirischen Tests. Die Felder **nk**, **minspan** und **endspan** werden beim Laden der Daten mit ihren empfohlenen Werten vorbelegt; eine manuelle Änderung bleibt erhalten.

7.7.1 Parameter des Vorwärtsdurchlaufs

nk (maximale Anzahl an Termen vor dem Pruning):

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

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

Der Standardwert von `earth`, $\min(200, \max(20, 2p)) + 1$, berücksichtigt die Größe des Datensatzes nicht und kann für große Datensätze zu klein ausfallen, wodurch der Vorwärtsdurchlauf eingeschränkt wird, bevor er alle Prädiktoren hinreichend untersucht hat. Der Term $\lfloor n/10 \rfloor$ erlaubt näherungsweise einen Term je 10 Beobachtungen — eine gängige Faustregel zur Vermeidung einer Überparametrisierung. Die Obergrenze von 100 beugt abnehmenden Erträgen vor.

minspan (minimale Anzahl an Beobachtungen zwischen Knoten):

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

endspan (minimale Anzahl an Beobachtungen zu den Datenrändern):

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

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

Der von `earth` automatisch berechnete `minspan`-Wert verwendet Friedmans Gleichung 43: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, die mit $\ln(np)$ skaliert und für große Datensätze zu langsam wächst. Bei $n=1500$ ergibt sie lediglich ≈ 7 , was rund 245 Kandidaten-Knotenpositionen je stetigem Prädiktor entspricht — zu viele, sodass der Vorwärtsdurchlauf Rauschen anpassen kann. Die empfohlene Formel zielt für große n auf etwa 100 Kandidaten je Prädiktor ab, während sie sich für kleine n (wo diese gut erprobt ist) an der automatischen Berechnung von `earth` orientiert. `Endspan` wird geringfügig größer als `minspan` gewählt, um zusätzlichen Schutz an den Datenrändern zu bieten; dies ist hilfreich, wenn die Daten dünn besetzte Ränder aufweisen (im Immobilienbereich häufig).

penalty (GCV-Strafterm je Knoten):

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

Dies ist der eigene Standardwert von `earth` — 2 für additive Modelle, 3 für Interaktionsmodelle (der höhere Strafterm kompensiert den größeren Suchraum).

newvar.penalty (Strafterm für die Einführung eines neuen Prädiktors):

Empfohlen: **0.1**, wenn die Prädiktoren korreliert sind (z. B. Wohnfläche mit Schlafzimmerzahl, Badezimmerzahl, Grundstücksgröße). Dies lenkt den Vorwärtsthroughlauf dahin, bereits im Modell enthaltene Prädiktoren wiederzuverwenden, statt korrelierte Alternativen einzuführen. Der Mechanismus: Die RSS-Verbesserung jedes neuen Prädiktors wird mit $1/(1+\text{penalty})$ multipliziert. Bei `newvar.penalty = 0.1` muss eine neue Variable die RSS um mindestens 10% stärker verbessern als ein weiterer Knoten auf einer bereits vorhandenen Variablen. Dies führt zu einfacheren Modellen mit weniger Prädiktoren, ohne die endgültigen Koeffizientenschätzungen zu beeinflussen (der Strafterm wird nach der Auswahl entfernt). Sind die Prädiktoren nicht korreliert, sollte der Wert bei 0 belassen werden. Zu beachten ist, dass `earth` einen von null verschiedenen Wert für `newvar.penalty` nicht in Kombination mit Fallgewichten unterstützt — `earthUI` deaktiviert das Eingabefeld (und erzwingt 0), sobald eine Gewichtsspalte festgelegt ist.

7.7.2 Pruning-Parameter

pmethod: Empfohlen: **backward**. Das standardmäßige GCV-basierte Rückwärts-Pruning ist deterministisch — dieselben Daten erzeugen stets dasselbe Modell, ohne Abhängigkeit von Zufallsstartwerten. Dies ist für die Reproduzierbarkeit wichtig, insbesondere in der Wertermittlung. Die Methode `cv` verwendet für das Pruning eine Kreuzvalidierung, wodurch eine Abhängigkeit vom Startwert (Seed) entsteht.

nprune: Empfohlen: **leer lassen** (NULL). Die optimale Anzahl an Termen sollte durch GCV bestimmt werden. Das Setzen von `nprune` erzwingt eine feste Obergrenze, die das GCV-Kriterium außer Kraft setzt.

7.7.3 Kreuzvalidierungs-Parameter

nfold (CV-Folds): Empfehlung nach Bereich der Stichprobengröße:

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

Mehr Folds trainieren jeden Fold auf mehr Daten (eine weniger verzerrte Schätzung), verkleinern jedoch den zurückgehaltenen Test-Fold. Die Tabelle steigert die Anzahl mit wachsender Stichprobe schrittweise bis zur üblichen 10-fachen Kreuzvalidierung und hält dabei jeden Test-Fold groß genug (rund 50+ Zeilen) für eine stabile Kennzahl je Fold; für kleine Stichproben fällt sie auf die 5-fache Kreuzvalidierung zurück — dort ist die CVR^2 -Schätzung weiterhin veräuscht, weshalb `ncross` erhöht werden sollte (z. B. 15–20), um die Varianz herauszumitteln. Die Obergrenze von 10 trägt dem Umstand Rechnung, dass jeder zusätzliche Fold eine vollständige `earth`-Anpassung bei vernachlässigbarem diagnostischem Zugewinn bedeutet. Mit `pmethod = "backward"` beeinflusst die Kreuzvalidierung das Modell nicht — sie berechnet lediglich das diagnostische CVR^2 und liefert Residuen für das Varianzmodell.

ncross (CV-Wiederholungen): Empfehlung nach Bereich der Stichprobengröße:

n (Anpassungszeilen)	ncross
0–199	20
200–399	16
400–699	12

n (Anpassungszeilen)	ncross
700–1199	10
1200–2999	7
3000+	5

Die Wiederholungen mitteln den Zufall der Fold-Aufteilung heraus und stabilisieren so CVR^2 ; größere Stichproben benötigen weniger Wiederholungen, da eine einzelne Kreuzvalidierung bereits stabiler ist. Die Residuen der wiederholten CV fließen zudem in das Varianzmodell ein.

7.7.4 Varianzmodell

varmod.method: Empfohlen: **lm** — Milborrows Standard: ein robustes lineares Varianzmodell für Prognoseintervalle (die Streuung wird aus verrauschten Residuen geschätzt, weshalb ein einfaches Modell besser generalisiert). Konvergiert **lm** auf den vorliegenden Daten nicht, empfiehlt sich **rlm** oder **const** (homoskedastisch, konvergiert stets), ersatzweise **earth** (erfasst auch eine nichtlineare Streuung). Erforderlich sind **nfold** > 1 und **ncross** > 1.

7.7.5 Earth-Ausgabedatei

Nach jeder erfolgreichen Anpassung schreibt earthUI die Datei `<filename>_earth_output_<timestamp>.txt` in den Ausgabeordner. Diese Datei enthält die Modellterme, zusammenfassende Statistiken (R^2 , GRSq , CVRSq), Einzelheiten zum Varianzmodell sowie das Trace-Protokoll. Je Anpassung wird eine Datei erzeugt, sodass eine fortlaufende Aufzeichnung zum Vergleich von Parameterkonfigurationen entsteht.

7.8 Voreinstellungen (Settings Defaults)

Ein Optionsfeld am oberen Rand von Abschnitt 4 steuert, welche Voreinstellungen geladen werden:

- **Use last settings for input file** — stellt die Einstellungen wieder her, die zuletzt mit genau dieser Datei verwendet wurden
- **Use default settings** — übernimmt die gespeicherten benutzerdefinierten Voreinstellungen
- **Earth defaults** — setzt alle Parameter auf die Werksvorgaben zurück

Über **Save current as default** am unteren Rand von Abschnitt 4 werden die aktuellen earth-Parameter sowie die Matrix Allowed Interactions gespeichert; die entsprechende Schaltfläche in Abschnitt 3 speichert die Konfiguration von Zielvariable, Stichtag und Prädiktoren. Beides wird je Projekt und Verwendungszweck in der Projektdatenbank abgelegt.

8 Anpassung des Earth-Modells

8.1 Die Schaltfläche „Fit“

Abschnitt 5 der Seitenleiste enthält die Schaltfläche **Fit Earth Model** (Earth-Modell anpassen) zusammen mit einem Feld **Random seed** (Zufalls-Seed). Der Seed macht die Zuordnung der Kreuzvalidierungs-Folds reproduzierbar — ein Dropdown-Menü ruft die letzten 10 verwendeten Seeds ab, und der verwendete Seed wird nach der Anpassung im Reiter **Earth Output** protokolliert. Vor der Anpassung validiert earthUI die Konfiguration — eine Zielvariable muss ausgewählt und mindestens ein Prädiktor muss einbezogen sein. In den Modi Appraisal/Market werden die Spalten für geografische Breite und Länge gerundet, und das Verkaufsalter wird aus dem Stichtag und dem Vertragsdatum berechnet (sofern zugewiesen).

8.2 Anpassungs-Modal & Trace-Ausgabe

Beim Klick auf **Fit** erscheint eine dunkle modale Überlagerung mit:

- **Kopfzeile** — „Fitting Earth Model“ mit einem laufenden Zeitzähler in Sekunden sowie einer Schaltfläche **Abort** (Abbrechen), die die im Hintergrund laufende Anpassung abbricht
- **Trace-Log** — ein scrollbarer Textbereich in nichtproportionaler Schrift, der die Echtzeitausgabe der Funktion `earth()` anzeigt: Informationen zum Datensatz, Fortschritt des Forward Pass, Kreuzvalidierungs-Folds und Abschlusszeit
- **Farbcodierung** — Standardzeilen in Grün, Kreuzvalidierungszeilen in Gelb, Fehler in Rot

earthUI passt Modelle asynchron mittels `callr::r_bg()` an, wodurch die Berechnung in einem separaten R-Prozess ausgeführt wird. Dadurch bleibt die Anwendung während lang laufender Anpassungen reaktionsfähig. Ein Polling-Observable mit 300 ms Intervall liest die Ausgabe des Hintergrundprozesses und leitet sie fortlaufend an das Trace-Log weiter.

Nach Abschluss der Anpassung erscheint eine Meldung „Done in X.Xs“, dem Modal wird eine Schließen-Schaltfläche (X) hinzugefügt, und der Hintergrund wird entfernt, sodass die Ergebnis-Reiter bei Bedarf gerendert werden können. Das Modal schließt sich anschließend automatisch, sobald die sichtbaren Ausgaben fertig gerendert sind — alternativ kann es auch früher geschlossen werden.

Bei Erfolg wird der Schaltfläche **Fit** ein grünes Häkchen hinzugefügt. Tritt ein Fehler auf, wird die Fehlermeldung im Trace-Log angezeigt und eine Benachrichtigung erscheint.

8.3 Anpassungsprotokoll

Jede Anpassung (Erfolg oder Fehlschlag) schreibt eine Protokolldatei in den Ausgabeordner:

- **Dateiname:** `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Inhalt:** Zeitstempel, die gesamte Trace-Ausgabe des Anpassungsprozesses sowie etwaige Fehlermeldungen
- **Speicherort:** der Ausgabeordner des aktiven Projekts

Tipp: Ist `callr` nicht installiert, weicht earthUI auf eine synchrone Anpassung mit einem einfachen Fortschrittsbalken aus. Für die vollständige Trace-Ausgabe sollte `callr` installiert werden.

9 Herunterladen der Daten

Nach der Anpassung lädt Abschnitt 6 der Seitenleiste — im allgemeinen Modus mit „Download Estimated Target Variable(s) & Residuals“ und in den Modi Appraisal und Market mit „Download Estimated Sale Prices & Residuals“ betitelt — eine Excel-Datei mit Prognosen und Diagnostiken herunter. Diese Ausgabe wird in Schritt 7 (RCA) verwendet, um dem Bewertungsobjekt eine CQA-Bewertung (Condition-Quality-Appeal, d. h. Zustand-Qualität-Attraktivität) zuzuweisen. Die Ausgabe ist nach `residual_sf` und `cqa_sf` sortiert, um die Einordnung des Bewertungsobjekts im Ranking zu erleichtern. Bei guter Modellqualität sollten die Objekte auf Grundlage der residualen Merkmale, die nicht in die Regression eingegangen sind, von der geringsten bis zur höchsten Attraktivität geordnet sein. Der mittlere Wert sollte annähernd 0 betragen, die untere Hälfte negative und die obere Hälfte positive Werte aufweisen. Die Objekte schlechtester Qualität, sogenannte „Fixers“ (renovierungsbedürftige Objekte), sollten sich am unteren Ende des Rankings finden und die attraktivsten Objekte am oberen Ende. Ausnahmen ergeben sich üblicherweise bei Anomalien wie Zwangsversteigerungen, Short Sales (Verkäufen unter Restschuld), Nachlassverkäufen (erbschaftsbedingt) sowie Schnellverkäufen aufgrund eines Arbeitsplatzwechsels oder aus anderen Gründen. Die Untersuchung von Anomalien fördert in der Regel einen einschlägigen Grund für die Preisanomalie zutage.

Die Beschriftung der Schaltfläche passt sich dem Zweckmodus an:

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

Das Dateinamensformat lautet `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`.

9.1 Ausgabespalten

Für jede Zielvariable werden die folgenden Spalten angefügt:

Spalte	Beschreibung
<code>est_<target></code>	Modellprognose, z. B. <code>est_sale_price</code> (1 Dezimalstelle)
<code>residual</code>	Ist-Wert minus Prognosewert (1 Dezimalstelle)
<code>cqa</code>	Condition-Quality-Appeal-Score (2 Dezimalstellen, Wertebereich 0–10)
<code>residual_sf</code>	Residuum dividiert durch die Wohnfläche (sofern zugewiesen, 1 Dezimalstelle)
<code>cqa_sf</code>	CQA, berechnet aus dem Ranking über <code>residual_sf</code> (2 Dezimalstellen)
<code><variable>_contribution</code>	Wertbeitrag je g-Funktion (1 Dezimalstelle, eine Spalte je g-Funktion)
<code>basis</code>	Wertbeitrag des Achsenabschnitts, für alle Objekte identisch (1 Dezimalstelle)
<code>calc_residual</code>	Ist-Wert minus (basis + alle Wertbeiträge) — Verifikationsspalte (1 Dezimalstelle)

Bei Modellen mit mehreren Zielvariablen wird ein Suffix `_<i>` angefügt, um die Spalten der einzelnen Zielvariablen zu unterscheiden.

9.2 Spaltenanordnung & Excel-Formatierung

Die Ranking-Spalten werden in der Ausgabedatei an der **äußersten linken Position** platziert, und zwar in dieser Reihenfolge: `residual_sf`, `cqa_sf`, `residual`, `cqa`. Dies erleichtert das

Durchsehen der sortierten Liste und die Beurteilung, wo das Bewertungsobjekt in der CQA-Verteilung liegt.

Auf diese Spalten wird eine Excel-Zahlenformatierung angewendet:

Spalte	Format	Beispiel
<code>residual_sf</code>	Numerisch, 2 Dezimalstellen	12.34
<code>cqa_sf</code>	Numerisch, 2 Dezimalstellen	7.25
<code>residual</code>	Numerisch, 0 Dezimalstellen	15,200
<code>cqa</code>	Numerisch, 2 Dezimalstellen	6.80

9.3 CQA-Scores

Der CQA-Score (Condition-Quality-Appeal) stuft das Residuum jeder Zeile gegenüber allen anderen Residuen ein. Für eine gegebene Zeile ist der CQA der Prozentsatz der Zeilen mit einem kleineren vorzeichenbehafteten Residuum, multipliziert mit 10. Daraus ergibt sich eine Skala von 0 bis 10, wobei gilt:

- **Hoher CQA** ($\approx 9-10$) — das Objekt wurde für deutlich mehr verkauft, als das Modell prognostiziert hat (großes positives Residuum)
- **Niedriger CQA** ($\approx 0-1$) — das Objekt wurde für deutlich weniger als prognostiziert verkauft (großes negatives Residuum)
- **CQA ≈ 5** — das Residuum liegt nahe am Median aller Residuen

Ist eine Spalte `living_area` zugewiesen, liefert `cqa_sf` dasselbe Ranking auf Grundlage der Residuen je Quadratfuß.

9.4 Sortierung & Zeile des Bewertungsobjekts

In den Modi Appraisal und Market werden die Zeilen der Vergleichsobjekte absteigend nach `residual_sf` sortiert (bzw. nach `residual`, wenn keine Wohnfläche zugewiesen ist). Die Zeile des Bewertungsobjekts (Zeile 1) verbleibt an Position 1 und wird nicht mitsortiert. Im allgemeinen Modus werden die Zeilen in ihrer ursprünglichen Reihenfolge exportiert.

Diese Sortierung ermöglicht es dem Gutachter in Verbindung mit den ganz links platzierten Ranking-Spalten, die Vergleichsobjekte rasch von der stärksten Überprognose bis zur stärksten Unterprognose durchzusehen und zu beurteilen, wo der dem Bewertungsobjekt zugewiesene CQA-Score in dieser Verteilung liegt.

Die Faktorstufen in den Prognosedaten werden vor dem Aufruf von `predict()` an die Trainingsdaten angeglichen. Zeilen mit unbekannten Faktorstufen erzeugen NA-Prognosen.

Tipp: Die Spalte `calc_residual` sollte stets (im Rahmen der Rundung) gleich `residual` sein. Weichen die Werte erheblich voneinander ab, deutet dies auf ein untersuchenswertes Berechnungsproblem hin.

9.5 Automatischer mgcvUI-Export

Bei jeder erfolgreichen Anpassung mit `degree  $\leq 2$` speichert earthUI das vollständige Ergebnisobjekt automatisch als `.rds`-Datei im Ausgabeordner des Projekts. Der Dateiname folgt dem Muster `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

Diese Datei kann vom mgcv Post Processing Module (mgcv-Nachbearbeitungsmodul) — das im earthUI-Paket enthalten ist und mit `earthUI::launch_mgcv()` gestartet wird — mittels `readRDS()` geladen werden. Die GAM-Routine verwendet die Knotenpositionen und Basisfunktionen des Earth-Modells als Ausgangspunkte für die GAM-Glättungsterme und ermöglicht so einen nahtlosen Übergang von der MARS- zur GAM-Modellierung.

Modelle mit `degree > 2` werden übersprungen, da mgcvUI ausschließlich paarweise Interaktionen unterstützt. Wird earthUI im Trilogy-Modus (aus der Trilogy UI heraus) gestartet, erlaubt ein zusätzlicher Abschnitt **Lock Model Output** (Modellausgabe fixieren), die .rds-Datei einer bestimmten Anpassung — zusammen mit ihrer Prädiktorbasis und den gemeinsamen CQA-Einstellungen — als kanonisches Earth-Modell zu fixieren, das glmnetUI und mgcvUI importieren.

9.6 glmnetUI-Import

Dieselbe für die GAM-Routine gespeicherte .rds-Datei kann auch in das **elastic net Post Processing Module** (Elastic-Net-Nachbearbeitungsmodul) importiert werden — ebenfalls im earthUI-Paket enthalten und mit `earthUI::launch_glmnet()` zu starten. In der glmnet-Anwendung ist zu **Section 2: Import from earthUI** zu navigieren und die .rds-Datei über die Schaltfläche **Browse** auszuwählen.

glmnetUI verwendet den Standardansatz über `earth::model.matrix()`: Die Basisfunktionen des Earth-Modells (Hinges, Interaktionen und lineare Terme) werden zu den Spalten der Designmatrix von glmnet. glmnet führt anschließend eine regularisierte Regression auf diesen Basisspalten durch und selektiert und schrumpft sie mittels Lasso/Elastic Net. Dies verbindet die adaptive Basiskonstruktion von earth mit der Regularisierung von glmnet.

9.6.1 Wesentliche Vorteile

- **Interpretierbare Interaktionen:** Die Hinge-basierten Interaktionen von earth (einschließlich Dreifachtermen) besitzen eine klare geometrische Bedeutung, im Gegensatz zu den nativen Kreuzprodukt-Interaktionen von glmnet, die vor Gericht oder in Prüfungskontexten schwer zu erläutern sind.
- **Automatische Nichtlinearität:** Die Hinge-Funktionen von earth erfassen nichtlineare Zusammenhänge, die ein Standard-glmnet-Modell mit unbearbeiteten Prädiktoren übersehen würde.
- **Variablenselektion:** glmnet kann die Earth-Basis weiter ausdünnen und Hinge-Terme entfernen, die keinen Beitrag leisten.

9.6.2 Wichtig: Die Gewichtungsspalte muss übereinstimmen

Warnung

Verwendet earthUI eine Gewichtungsspalte (z. B. um die Zeile des Bewertungsobjekts mit `weight=0` auszuschließen), muss dieselbe Gewichtungsspalte in der Variablenkonfiguration von glmnetUI zugewiesen werden. Weichen die Gewichtungseinstellungen voneinander ab, stimmen die Zeilenzahlen nicht überein, und die Earth-Basis kann nicht an die Modellmatrix von glmnet angefügt werden.

9.6.3 Arbeitsablauf

1. Anpassung eines Earth-Modells in earthUI (beliebiger Grad).
2. Die .rds-Ergebnisdatei wird automatisch im Ausgabeordner des Projekts gespeichert.
3. In glmnetUI **Section 2: Import from earthUI** öffnen und zur .rds-Datei navigieren.
4. Überprüfen, dass in beiden Anwendungen **dieselbe Datendatei** und **dieselbe Gewichtungsspalte** verwendet werden.
5. Konfiguration der glmnet-Parameter in Abschnitt 5 (die Interaction Matrix wird bei aktivem earthUI-Import automatisch deaktiviert).

6. Auf **Fit Model** klicken. Die Earth-Basispalten erscheinen in den Reitern Equation, Coefficients und Contributions.

10 RCA-Berechnungen & Download

Die Schaltfläche **Calculate RCA Adjustments & Download** (RCA-Anpassungen berechnen und herunterladen) erscheint in Abschnitt 7 der Seitenleiste und ist nur im Appraisal Mode sichtbar, nachdem ein Modell angepasst wurde. RCA (Residual Constraint Approach) erzeugt marktabgeleitete Anpassungen je Vergleichsobjekt relativ zum Bewertungsobjekt.

10.1 Öffnen des RCA-Dialogs

Ein Klick auf die Schaltfläche öffnet einen kleinen modalen Dialog mit:

- **Score type** (Score-Typ) — Optionsfelder zur Auswahl von **CQA** oder **CQA per SF** (CQA je Quadratfuß; die SF-Option ist nur verfügbar, wenn eine Spalte `living_area` festgelegt wurde). Wird „CQA per SF“ gewählt, so ergibt sich der Residual-Score des Bewertungsobjekts auf Grundlage des ihm zugewiesenen CQA-Scores als dessen Wohnfläche multipliziert mit dem `residual_sf`-Wert, der dem gegebenen CQA_SF-Score entspricht.
- **CQA value** (CQA-Wert) — Eingabefeld für den CQA-Score des Bewertungsobjekts (0.00–10.00). Dieser Score gibt an, wo das Bewertungsobjekt nach Einschätzung des Anwenders in der Qualitätsverteilung der Vergleichsobjekte liegt.
- Schaltfläche **Generate** (Erzeugen) — berechnet die RCA-Ausgabe und lädt sie als Excel-Datei herunter. Die Schaltfläche bleibt deaktiviert, bis ein gültiger Score eingegeben wurde.

10.2 Interpolation des CQA-Scores

Beim Klick auf **Generate** interpoliert earthUI das Residuum des Bewertungsobjekts aus den CQA-/Residuen-Paaren der Vergleichsobjekte:

1. Die CQA-Scores und Residuen der Vergleichsobjekte werden nach CQA aufsteigend sortiert
2. Eine lineare Interpolation (`stats::approx()`) bildet den eingegebenen CQA-Wert auf ein Residuum ab
3. Bei Verwendung von CQA per SF wird das Residuum je Quadratfuß durch Multiplikation mit der Wohnfläche des Bewertungsobjekts in ein Gesamtresiduum zurückgerechnet
4. Der geschätzte Wert des Bewertungsobjekts wird berechnet als: Modellprognose + interpoliertes Residuum

10.3 Ausgabespalten

Die RCA-Excel-Datei (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) enthält alle Zwischenausgabespalten sowie zusätzlich:

Spalte	Beschreibung
<code>subject_value</code>	Modellprognose + interpoliertes Residuum (nur Zeile 1)
<code>subject_cqa</code>	Der eingegebene CQA-Score (nur Zeile 1)
<code><variable>_adjustment</code>	Wertbeitrag des Bewertungsobjekts abzüglich des Wertbeitrags des Vergleichsobjekts (je g-Funktion)
<code>residual_adjustment</code>	Residuum des Bewertungsobjekts abzüglich des Residuums des Vergleichsobjekts
<code>net_adjustments</code>	Summe aller Anpassungen (Wertbeitrag + Residuum)
<code>gross_adjustments</code>	Summe der Absolutbeträge aller Anpassungen
<code>residual_pct</code>	Residualanpassung als Anteil am Verkaufspreis des Vergleichsobjekts

Spalte	Beschreibung
<code>net_adj_pct</code>	Nettoanpassungen als Anteil am Verkaufspreis des Vergleichsobjekts
<code>gross_adj_pct</code>	Bruttoanpassungen als Anteil am Verkaufspreis des Vergleichsobjekts
<code>adjusted_sale_price</code>	Verkaufspreis des Vergleichsobjekts + Nettoanpassungen

Die Anpassungsspalten geben für jedes Vergleichsobjekt an, in welchem Umfang das Modell die Differenz den einzelnen Variablen zuschreibt. `adjusted_sale_price` ist der Verkaufspreis des Vergleichsobjekts nach Anwendung aller modellabgeleiteten Anpassungen — eine Menge angepasster Verkaufspreise, die sich eng um den geschätzten Wert des Bewertungsobjekts gruppieren sollten.

10.4 Die Registerkarte RCA Adjustments

Sobald die RCA-Ausgabe berechnet wurde, zeigt die Registerkarte **RCA Adjustments** (RCA-Anpassungen) im Hauptpanel Histogramme der Verteilungen von Residual Adj %, Net Adj % und Gross Adj % über die Vergleichsobjekte an (bei Modellen mit mehreren Zielgrößen ein Satz je Zielgröße) und ermöglicht so eine schnelle visuelle Kontrolle, wie eng die Anpassungen gruppiert sind.

Tipp: Ein CQA-Score von 5.00 platziert das Bewertungsobjekt am Median der Vergleichsobjekte. Scores über 5 zeigen an, dass das Bewertungsobjekt eine überdurchschnittliche Qualität aufweist; Werte unter 5 deuten auf eine unterdurchschnittliche Qualität hin.

11 Sales Comparison Grid

11.1 Überblick & Zweck

Das **Sales Comparison Grid** (Vergleichsraster) ist eine Excel-Arbeitsmappe, die im Appraisal Mode (Sidebar-Abschnitt 8) nach der Berechnung der RCA-Anpassungen erzeugt wird. Es stellt das Bewertungsobjekt zusammen mit ausgewählten Vergleichsobjekten in einem strukturierten Rasterformat dar, wobei Excel-Formeln die Anpassungen sowie einen angepassten Verkaufspreis für jedes Vergleichsobjekt automatisch berechnen.

Das Raster ist für die Arbeitsakte des Gutachters konzipiert — es kombiniert die aus der Regression abgeleiteten Anpassungen des Earth-Modells mit editierbaren Zellen, in denen der Gutachter das CQA-Residuum bestimmten Objektmerkmalen zuordnen kann. Der Name der Ausgabedatei lautet **SalesGrid_<YYYYMMDD_HHMMSS>.xlsx**.

11.2 Modaldialog zur Auswahl der Vergleichsobjekte

Ein Klick auf „Generate Sales Grid & Download“ öffnet einen modalen Dialog, in dem ausgewählt wird, welche Vergleichsobjekte einbezogen werden sollen:

- **Recommended comps** (empfohlene Vergleichsobjekte) sind vorab angehakt. Dabei handelt es sich um Vergleichsobjekte mit einem Brutto-Anpassungsprozentsatz von unter 25% des Verkaufspreises, aufsteigend nach Verkaufsalter sortiert (jüngste Verkäufe zuerst) und auf 30 begrenzt.
- **Additional comps** (weitere Vergleichsobjekte) werden darunter aufgeführt und sind standardmäßig nicht ausgewählt. Diese weisen Brutto-Anpassungen von 25% oder mehr auf, stehen jedoch weiterhin zur Auswahl (angezeigt werden bis zu 50, aufsteigend nach Brutto-Anpassungsprozentsatz sortiert).
- Es können maximal **30 Vergleichsobjekte** ausgewählt werden, woraus sich bis zu **10 Arbeitsblätter** ergeben (3 Vergleichsobjekte pro Blatt).

Nach Bestätigung der Auswahl werden die ausgewählten Vergleichsobjekte aufsteigend nach Brutto-Anpassungsprozentsatz sortiert, sodass die ähnlichsten Vergleichsobjekte auf Blatt 1 erscheinen.

11.3 Rasterlayout

Jedes Arbeitsblatt besitzt ein Layout mit 20 Spalten, das das Bewertungsobjekt sowie 3 Vergleichsobjekte aufnimmt. Die Spalten je Entität sind:

Entität	Spalten (5 je Entität)
Bewertungsobjekt	Label, Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC)
Jedes Vergleichsobjekt	Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC), Adjustment

Die Zeilen von oben nach unten sind:

1. **Titel** — Blattname (z. B. „Intermediate Sales Comparable Grid — Sheet 1 of 3“)
2. **Kopfzeilen** — „Subject“ sowie die Spaltenüberschriften der Vergleichsobjekte mit Adresse
3. **Adresse** — vollständige Objektadresse
4. **APN | MLS# | DOM | Subj.Prox** — Flurstücksnummer, Listing-ID, Vermarktungsdauer (days on market) und Haversine-Distanz (in Meilen) zum Bewertungsobjekt

5. **Sales Price | Concess. | Net SP** — Verkaufspreis, Preiszugeständnisse und Formel für den Nettoverkaufspreis
6. Kopfzeile **Regression Features**
7. **BASE VALUE** — der Achsenabschnitt (Intercept) des Modells
8. **Date of Sale | OffMkt | OnMkt** — Vertragsdatum, Verkaufsalter und DOM
9. **Gruppierte Zeilen** (bedingt) — Gruppen für Location, Site Size und/oder Age
10. **Modellvariablenzeilen** — eine Zeile je Prädiktor (ohne gruppierte Variablen)
11. **Leere Trennzeile**
12. Kopfzeile **Residual Features**
13. **CQA|Residual** — CQA-Score + Formel für das verbleibende Residuum
14. **Zeilen für Residualmerkmale** — benannte + leere Zeilen für Eintragungen des Gutachters
15. **Total VC / Net Adjustment** (Gesamt-Wertbeitrag / Netto-Anpassung)
16. **Net Adjustment %** (Netto-Anpassung in Prozent)
17. **Gross Adjustment %** (Brutto-Anpassung in Prozent)
18. **Adjusted Sale Price** (angepasster Verkaufspreis) — Formelzeile
19. **Copyright-Fußzeile**

11.4 Sale Price, Concessions & Net SP

Zeile 5 zeigt für jedes Vergleichsobjekt drei Werte:

- **Sale Price** — der Verkaufspreis des Vergleichsobjekts aus den Daten
- **Concessions** — der Wert aus der als **concessions** ausgewiesenen Spalte (0, falls keine Spalte ausgewiesen ist)
- **Net SP** — eine Excel-Formel: Verkaufspreis – Preiszugeständnisse

Die Spalte des Bewertungsobjekts zeigt „N/A“ für den Verkaufspreis (da dieser unbekannt ist) sowie gegebenenfalls einen vorhandenen Wert für Preiszugeständnisse.

11.5 Gruppierte Zeilen (Location, Site, Age)

Wenn bestimmte Spezialtypen ausgewiesen sind und die entsprechenden Variablen im Modell erscheinen, erzeugt earthUI gruppierte Zeilen, die verwandte Variablen zusammenfassen:

Loc: Long | Lat | Area — Erscheint, wenn **longitude**, **latitude** oder **area** im Modell enthalten sind. Zeigt die faktischen Werte jeder beteiligten Variablen, einen kombinierten Wertbeitrag (Summe der einzelnen Wertbeiträge) sowie für Vergleichsobjekte eine Anpassung (kombinierter VC des Bewertungsobjekts – kombinierter VC des Vergleichsobjekts). Die VC-Zellen sind mit hellblauem Hintergrund formatiert.

Site Size | Dimensions — Erscheint, wenn **lot_size** oder **site_dimensions** im Modell enthalten sind. Gleiche Struktur wie die Location-Gruppe.

Actual Age | Effective Age — Erscheint, wenn **actual_age** oder **effective_age** im Modell enthalten sind. Gleiche Struktur.

Von gruppierten Zeilen erfasste Variablen werden **aus den darunterliegenden einzelnen Modellvariablenzeilen ausgeschlossen**, um Doppelzählungen in den Anpassungssummen zu vermeiden.

11.6 Modellvariablenzeilen

Unterhalb der gruppierten Zeilen (bzw. direkt unterhalb von BASE VALUE, falls keine gruppierten Zeilen vorhanden sind) zeigt je eine Zeile pro verbleibendem Modellprädiktor:

- **Faktische Werte** für das Bewertungsobjekt und jedes Vergleichsobjekt (die tatsächlichen Datenwerte)

- **Wertbeitrag (Value Contribution, VC)** — der Beitrag der g-Funktion zum prognostizierten Wert in dieser Zeile
- **Anpassung** (nur Vergleichsobjekte) — VC des Bewertungsobjekts minus VC des Vergleichsobjekts

11.7 CQA|Residual & Zeilen für Residualmerkmale

Die Zeile **CQA|Residual** zeigt den CQA-Score jedes Objekts und enthält eine Formel für das **verbleibende Residuum** — den Teil des Residuums, der noch keinen konkreten Merkmalen zugeordnet wurde. Die Formel lautet:

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

Darunter befinden sich **Zeilen für Residualmerkmale** — 5 benannte Zeilen (Location, View/Appeal, Condition, Quality, Other) sowie 6 leere Zeilen. Dies sind Eingabezellen, in die der Gutachter Wertbeiträge für Merkmale einträgt, die das Modell nicht erfasst. Mit jedem eingetragenen Wert verringert sich das verbleibende Residuum gemäß der Formel automatisch.

Für jedes Vergleichsobjekt enthält die Anpassungsspalte eine Formel: Merkmal-VC des Bewertungsobjekts – Merkmal-VC des Vergleichsobjekts.

11.8 Formel für den Adjusted Sale Price

Die Zeile **Adjusted Sale Price** enthält Excel-Formeln:

- **Bewertungsobjekt:** Summe aller Wertbeitragszellen von BASE VALUE bis zur letzten Residualmerkmalzeile. Dies entspricht der Gesamtprognose des Modells für das Bewertungsobjekt zuzüglich der vom Gutachter zugeordneten Residualmerkmale.
- **Vergleichsobjekte:** Net SP + Summe aller Anpassungszellen von der ersten gruppierten Zeile bzw. Variablenzeile bis zur letzten Residualmerkmalzeile. Dies ergibt den Verkaufspreis des Vergleichsobjekts nach allen aus der Regression abgeleiteten und vom Gutachter eingetragenen Anpassungen.

11.9 Blattschutz & editierbare Zellen

Jedes Arbeitsblatt ist **geschützt**, um versehentliche Änderungen an Formeln und Daten zu verhindern. Der Schutz sperrt:

- Alle Formelzellen (Net SP, verbleibendes Residuum, Anpassungen, angepasster Verkaufspreis usw.)
- Alle Datenzellen (Adressen, Verkaufspreise, faktische Werte, Wertbeiträge aus dem Modell)
- Beschriftungen und Kopfzeilen

Die einzigen **entsperrten** (editierbaren) Zellen sind die Eingabezellen für die Wertbeiträge der Residualmerkmale — die Zellen unterhalb der Zeile CQA|Residual, in die der Gutachter die Aufschlüsselung einträgt. Diese sind mit hellgelbem Hintergrund formatiert, um ihre Editierbarkeit anzuzeigen.

11.10 Arbeiten mit dem Raster in Excel

1. **Öffnen der Datei** in Excel oder einer kompatiblen Tabellenkalkulationsanwendung.
2. **Prüfung der Regressionsanpassungen** — die aus dem Modell abgeleiteten Anpassungen sind vorausgefüllt und gesperrt.
3. **Zuordnung des Residuums** — in den Zeilen für Residualmerkmale (gelbe Zellen) wird die Einschätzung eingetragen, welcher Anteil des verbleibenden Residuums dem jeweiligen Merkmal (Location, View, Condition, Quality usw.) zuzurechnen ist. Das verbleibende Residuum verringert sich gemäß der Formel mit jeder Zuordnung.

4. **Kontrolle des Adjusted Sale Price** — diese Formel aktualisiert sich automatisch, sobald Werte für Residualmerkmale eingetragen werden.
5. **Mehrere Arbeitsblätter** — wurden mehr als 3 Vergleichsobjekte ausgewählt, kann zwischen den Blättern navigiert werden. Jedes Blatt enthält dasselbe Bewertungsobjekt in der linken Spalte mit 3 unterschiedlichen Vergleichsobjekten.

Tipp: Ziel ist es, das CQA-Residuum so auf die Merkmalzeilen zu verteilen, dass das verbleibende Residuum nahe null liegt. Damit wird gezeigt, dass die Differenz zwischen dem Regressionsschätzwert und dem tatsächlichen Verkaufspreis jedes Vergleichsobjekts erklärt werden kann.

12 Herunterladen von Berichten

Nach der Anpassung eines Modells erfolgt die Berichtserstellung als zweistufiger Quarto-Arbeitsablauf: **Generate Quarto Report** (Seitenleistenabschnitt 9 im Appraisal Mode, andernfalls 7) schreibt ein in sich geschlossenes Berichtspaket, und **Convert Quarto Report** (Abschnitt 10) rendert dieses — oder jede andere Quarto-Datei — in die endgültigen Formate.

12.1 Generate Quarto Report

Ein Klick auf **Generate Quarto Report** schreibt ein in sich geschlossenes Quarto-Paket in den Ausgabeordner des Projekts unter `<base>_qmd/`, das die befüllte `.qmd`-Quelldatei, alle vorab erzeugten Diagramm-Assets (PNG und PDF), die Berichtsdaten sowie eine Word-Referenzvorlage enthält. In diesem Schritt findet kein Rendern statt — das Paket kann von Hand bearbeitet, über Quarto-Includes mit anderen Quarto-Quellen kombiniert oder eigenständig mit `quarto preview` in der Vorschau betrachtet werden.

12.2 Convert Quarto Report

Abschnitt 10 stellt eine Dateiauswahl bereit (voreingestellt auf das zuletzt erzeugte Paket des aktuellen Projekts — es kann jedoch zu jeder beliebigen `.qmd`-Datei navigiert werden) sowie Kontrollkästchen für die Formate:

- **HTML** — eine `.html`-Datei mit KaTeX-Mathematikdarstellung, eingebetteten Bildern und einem Inhaltsverzeichnis. Verwendet das Flatly-Bootstrap-Thema und die Schriftart Roboto Condensed.
- **Word** — eine `.docx`-Datei, die sich zur Bearbeitung und Verteilung eignet. Enthält ein Inhaltsverzeichnis sowie Seitenzahlen und verwendet eine angepasste Referenzvorlage für eine einheitliche Formatierung.
- **PDF** — mit LuaLaTeX gesetzt für eine Ausgabe in professioneller Qualität. Das Papierformat richtet sich nach der Gebietsschema-Einstellung (Letter oder A4). Enthält Querformatseiten für große Interaktionsmatrizen und verwendet Roboto Condensed mit Latin Modern Math für Gleichungen. Die PDF-Option wird nur angeboten, wenn eine LaTeX-Distribution erkannt wird.

Ein Klick auf **Convert** rendert die ausgewählten Formate. Das Rendern läuft im Hintergrund — ein modaler Dialog zeigt die verstrichene Zeit und den Quarto-Fortschritt an, während die Anwendung weiterhin reaktionsfähig bleibt. Nach Abschluss bestätigt eine Benachrichtigung den Speicherort der Datei.

Alle Operationen (Modellanpassung, Daten-Downloads, RCA-Berechnungen, Sales-Grid-Erzeugung und Berichts-Rendering) werden mit Start-/End-Zeitstempeln und verstrichenen Zeiten in `<datafile>_earthui_log.txt` im Ausgabeordner protokolliert, zur Fehlersuche und Leistungsüberwachung.

12.3 Berichtsinhalte

Jeder erzeugte Bericht enthält die folgenden Abschnitte:

1. **Datensatzbeschreibung** — Anzahl der Beobachtungen, Zielvariable(n), verwendete Prädiktoren, kategoriale Prädiktoren
2. **Modellspezifikation** — Grad, Kreuzvalidierungsstatus, Anzahl der Terme, im Modell enthaltene Prädiktoren
3. **Zugelassene Interaktionen** (nur bei $\text{Grad} \geq 2$) — die Interaktionsmatrix mit Häkchen, formatiert für den jeweiligen Ausgabebetyp

4. **Ergebniszusammenfassung** — R^2 , GRSq, GCV, RSS und CV R^2 (sofern die Kreuzvalidierung aktiviert war)
5. **Modellgleichung** — die vollständige Gleichung des Earth-Modells in LaTeX-Notation
6. **Koeffizienten & Basisfunktionen** — Tabelle aller Terme, ihrer Koeffizienten und der sie definierenden Basisfunktionen
7. **Variablenwichtigkeit** — Balkendiagramm und Rangtabelle der Wichtigkeitswerte der Prädiktoren
8. **g-Funktions-Beiträge** — Diagramme für jede g-Funktion: Liniendiagramme für univariate Terme, perspektivische 3D- und Konturdiagramme für bivariate Interaktionen
9. **Korrelationsmatrix** — Heatmap der Prädiktorkorrelationen
10. **Diagnostik** — Diagramme „Residuals vs Fitted“, „Normal Q-Q“ und „Actual vs Predicted“
11. **ANOVA-Zerlegung** — Tabelle mit dem Beitrag jeder Basisfunktion zur RSS
12. **Earth-Ausgabe** — unformatierte Textausgabe von `earth::summary()`, einschließlich der Details des Pruning-Durchlaufs

Bei Modellen mit mehreren Zielvariablen werden die Abschnitte 4-8 und 10 für jede Zielvariable wiederholt.

Hinweis: Berichte erfordern das R-Paket *quarto* sowie eine Quarto-Installation. Für die PDF-Ausgabe muss zusätzlich eine LaTeX-Distribution installiert sein (z.B. über `tinytex::install_tinytex()`) — andernfalls wird das PDF-Kontrollkästchen ausgeblendet.

13 Die Post-Processing-Module: Elastic Net und GAM

Das earthUI-Paket wird mit zwei Post-Processing-Modulen ausgeliefert — einer Elastic-Net-Routine auf Basis von `glmnet` und einer GAM-Routine auf Basis von `mgcv`. Beide werden jeweils in einem eigenen Beitrag dieser Ausgabe eingehend behandelt; dieser Abschnitt erläutert, *wozu* sie im earthUI-Arbeitsablauf dienen und wie die drei Routinen zusammenwirken.

13.1 Wozu unterstützende Anpassungen?

Das Earth-Modell liefert das Wertschlussfolgerung. Die Post-Processing-Module dienen dazu, dieses zu **bestätigen**: zwei zusätzliche Schätzungen desselben Bewertungsobjekts, erzeugt durch methodisch unterschiedliche Verfahren und angepasst an dieselben Projektdaten. Wenn eine Elastic-Net-Neuschätzung der Earth-Basis und eine unabhängige GAM-Glättung zu einem im Wesentlichen gleichen Wert gelangen, beruht die Schlussfolgerung des Gutachters nicht mehr auf den Annahmen eines einzelnen Modells — ein bedeutsamer Vorteil in Prüfungs- und Gerichtskontexten. Wenn sie hingegen *voneinander abweichen*, ist diese Abweichung selbst diagnostisch aussagekräftig und deutet in der Regel auf dünne Daten, einen zu flexiblen Term oder ein Ausreißer-Vergleichsobjekt hin, das einer erneuten Prüfung wert ist. Keines der Module steht über dem Earth-Modell; sie stellen unterstützende Evidenz dar, die bewusst ohne Gewichtung präsentiert wird.

13.2 Start und gemeinsamer Zustand

Jede Routine ist eine eigenständige lokale Shiny-Anwendung:

- `earthUI::launch_glmnet()` — die Elastic-Net-Anwendung, auf Port 7879.
- `earthUI::launch_mgcv()` — die GAM-Anwendung, auf Port 7880.

Alle drei Anwendungen lesen denselben Projektbaum und dieselbe Einstellungsdatenbank, sodass das in earthUI geöffnete Projekt in den Modulen unmittelbar verfügbar ist — zusammen mit den von earthUI dafür gespeicherten Übernahmeefeldern: der Response (Zielvariable), dem Effective Date (Bewertungstichtag), dem RCA-CQA-Typ und -Wert des Bewertungsobjekts sowie der Wohnflächen-Zuweisung. Diese Angaben müssen nicht erneut eingegeben werden; die Module lesen sie ein, sodass dieselben Werte in alle drei Anpassungen einfließen.

13.3 Die Elastic-Net-Routine

Die `glmnet`-Routine importiert die Basisfunktionen des Earth-Modells — seine Hinge-Funktionen, Interaktionen und linearen Terme — als Spalten einer Designmatrix und schätzt sie unter Lasso-/Elastic-Net-Regularisierung neu (siehe „`glmnetUI` Import“ oben zur Funktionsweise und zum Vorbehalt bezüglich der Gewichtsspalte). Das Ergebnis behält die interpretierbaren, geometrisch aussagekräftigen Terme von earth bei, während der Regularisierer diejenigen Terme schrumpfen oder entfernen kann, die ihren Beitrag nicht rechtfertigen. Eine ausführliche Darstellung findet sich im `glmnetUI`-Beitrag dieser Ausgabe.

13.4 Die GAM-Routine

Die `mgcv`-Routine geht den umgekehrten Weg: Statt die Basis von earth wiederzuverwenden, passt sie neue penalisierte Glättungsterme an und nutzt dabei die Knotenpositionen des Earth-Modells als Ausgangspunkte (siehe „`mgcvUI` Auto-Export“ oben). Da das GAM die funktionalen Formen unter einem anderen Glättungsparadigma neu herleitet, stellt seine Übereinstimmung mit dem Earth-Modell eine tatsächlich unabhängige Bestätigung dar und nicht bloß ein Echo. Unterstützt werden Earth-Modelle mit $\text{degree} \leq 2$. Eine ausführliche Darstellung findet sich im `mgcvUI`-Beitrag dieser Ausgabe.

13.5 Ein praktischer Arbeitsablauf

1. Zunächst wird das **Earth-Modell** in earthUI angepasst und verfeinert, bis es den Anforderungen genügt. Jede erfolgreiche Anpassung ($\text{Grad} \leq 2$) exportiert automatisch die **.rds**-Datei, die von den Modulen eingelesen wird.
2. Anschließend werden die **glmnet**- und die **mgcv**-Routine für dasselbe Projekt geöffnet, das Earth-Ergebnis importiert und die Anpassung durchgeführt. Es empfiehlt sich, jeweils nur in einer Routine zu arbeiten — die Anpassungen sind langlaufend und die Routinen koordinieren sich über das gemeinsame Projekt, sodass ein sequentielles und kein paralleles Vorgehen sinnvoll ist.
3. Es kann frei iteriert werden. Jede Anwendung behält ihre Ergebnisse beim Wechsel zwischen den Anwendungen bei, und es besteht keine Notwendigkeit, unterwegs Berichte zu erzeugen.
4. Erst wenn alle drei Ergebnisse zufriedenstellend sind, wird zu jeder Routine zurückgekehrt und deren Quarto-Bericht erzeugt. Jede Methode registriert ihre Anpassung und ihr Wertschlussfolgerung im Trilogie-Datensatz des Projekts, sodass die drei Berichte auf einen gemeinsamen Vergleichslauf rückführbar bleiben.

Hinweis: Die Module sind auf jeder Ebene optional: Der eigene Arbeitsablauf von earthUI — Anpassung, RCA, Sales Grids, Berichte — ist auch ohne sie vollständig. Sie empfehlen sich, wenn eine Schlussfolgerung einer Bestätigung bedarf: bei einem schwierigen Bewertungsobjekt, einem dünnen Marktsegment oder einem Auftrag, der voraussichtlich einer Prüfung unterzogen wird.

14 Demo-Datensatz: Appraisal_1.csv

earthUI enthält einen MLS-Demo-Datensatz zur Erkundung des Arbeitsablaufs der Wertermittlung. Er lässt sich programmatisch laden mit:

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

Alternativ kann die Datei in der Shiny-App in den Ordner `in/` eines Projekts kopiert (oder beim Anlegen des Projekts als initiale Datendatei angehängt) und in Section 2 (Import Data) ausgewählt werden.

14.1 Beschreibung

Die Datei enthält 1,501 Wohnimmobilienverkäufe sowie 1 Bewertungsobjekt in Zeile 1 (insgesamt 1,502 Zeilen) aus einem simulierten MLS-Export. Die Daten repräsentieren Verkäufe von Einfamilienhäusern in einem Markt mit mehreren Teilgebieten und einer Bandbreite an Objektgrößen, Gebäudealtern und Lagen.

Es handelt sich nicht um reale Daten; sie basieren jedoch auf einer realistischen Nachbarschaft in Nordkalifornien. Sämtliche Identifikationsangaben wurden verändert oder entfernt.

14.2 Spalten

Spalte	Typ	Spezialtyp	Beschreibung
weight	numeric	—	Beobachtungsgewicht (0 = von der Anpassung ausgeschlossen)
id	numeric	display_only	Interne Datensatz-ID
property_id	numeric	display_only	MLS-Objektkennung
listing_id	character	display_only	MLS-Listing-Nummer
parcel_number	character	display_only	Flurstücksnummer des County Assessor (APN)
street_address	character	display_only	Objektadresse
city_name	character	display_only	Stadt
postal_code	character	display_only	Postleitzahl (ZIP)
county_name	character	display_only	County
contract_date	Date	contract_date	Kaufvertragsdatum (dient der Berechnung von sale_age)
sale_age	numeric	—	Tage vom Vertragsdatum bis zum Wertermittlungstichtag (vorberechnet)
coe_date	Date	display_only	Datum des Escrow-Abschlusses
listing_status	character	display_only	Listing-Status (z. B. SSold)
sale_price	numeric	(target)	Verkaufspreis — Zielvariable
rent	numeric	—	Monatsmiete (für Modelle mit mehreren Zielgrößen)
list_price	numeric	display_only	Angebotspreis
original_list_price	numeric	display_only	Ursprünglicher Angebotspreis
living_sqft	numeric	living_area	Bruttowohnfläche in Quadratfuß
beds_total	integer	—	Anzahl der Schlafzimmer
baths_total	numeric	—	Gesamtzahl der Bäder (z. B. 2.5 = 2 volle + 1 halbes)
lot_size	numeric	lot_size	Grundstücksgröße in Quadratfuß
area_id	integer	area	MLS-Gebietskennung
area_text	character	display_only	Gebietsname
age	numeric	actual_age	Gebäudealter in Jahren
year_built	integer	display_only	Baujahr
latitude	numeric	latitude	Geographische Breite (für das Modell auf 3 Nachkommastellen gerundet)
longitude	numeric	longitude	Geographische Länge (für das Modell auf 3 Nachkommastellen gerundet)
latitude6	numeric	display_only	Geographische Breite in voller Präzision
longitude6	numeric	display_only	Geographische Länge in voller Präzision
garage_spaces	integer	—	Anzahl der Garagenstellplätze
fp_count	integer	—	Anzahl der Kamine
no_of_stories	numeric	—	Anzahl der Stockwerke
style	character	—	Architekturstil
view	character	—	Aussichtstyp (z. B. NNeighborhood", "Hills")
days_on_market	integer	dom	Angebotstage am Markt
listing_date	Date	listing_date	Listing-Datum
sale_concessions	numeric	concessions	Verkäuferzugeständnisse

14.3 Empfohlener Schnelleinstieg

1. earthUI starten: `earthUI::launch()`
2. Im New-Project-Assistenten (New Project wizard) ein Projekt mit dem Purpose **For Appraisal** anlegen und dabei `Appraisal_1.csv` als initiale Datendatei anhängen
3. `Appraisal_1.csv` in Section 2 (Import Data) auswählen
4. `sale_price` als Zielvariable auswählen
5. Die Spezialtypen gemäß der obigen Tabelle zuweisen
6. Als Prädiktoren einbeziehen: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (als Faktor), `age`, `latitude`, `longitude`, `garage_spaces`
7. Den Grad (degree) auf 1 setzen und auf **Fit Earth Model** klicken
8. Die Zwischenausgabe herunterladen (Schritt 6) und das CQA-Ranking prüfen
9. Die RCA-Anpassungen (Schritt 7) mit einem CQA-Score von ~5.00 berechnen

10. Das Sales Comparison Grid (Vergleichsraster) generieren (Schritt 8)

Tipp: Der Demo-Datensatz eignet sich gut, um sämtliche Funktionen von earthUI zu erkunden, bevor eigene MLS-Daten importiert werden. Die oben vorgeschlagenen Prädiktoren liefern typischerweise ein Modell mit einem R^2 (Bestimmtheitsmaß) über 0.85.

15 Experimentell: Prolog-Textextraktion & Ableitungsregeln

Version 0.10.0 führt eine **optionale und ausdrücklich experimentelle** Funktionalität zur Extraktion modellfertiger Merkmale aus den beschreibenden Textspalten eines MLS-Exports ein — allen voran der Spalte „Public Remarks“, aber auch Maklerbemerkungen und kommasetrennte Merkmalslisten wie Raumausstattungen oder bestehende Baulichkeiten. Die Extraktions-Engine ist Prolog, insbesondere **Definite Clause Grammars (DCGs)**: Die Regeln, die Wendungen wie „move in ready“, „3 car garage“ oder „owned solar“ im Bemerkungstext erkennen, sind als Prolog-Grammatikregeln formuliert.

Diese Funktion befindet sich noch in der Entwicklung. Sie kann bereits heute *teilweise* zur Extraktion eingesetzt werden — die mitgelieferte Grammatik erkennt bereits einen nützlichen Satz von Merkmalen —, doch ihre Anpassung an das Vokabular des eigenen MLS setzt voraus, Prolog und insbesondere DCGs lesen und schreiben zu können. Wer diese Voraussetzung nicht mitbringt, kann dieses Kapitel vollständig überspringen: Die Funktion ist standardmäßig deaktiviert, und der Rest von earthUI arbeitet ohne sie exakt gleich.

Warnung

Dies ist eine experimentelle Funktion in aktiver Entwicklung. Ihre Extraktionsgrammatik ist weder vollständig noch für den produktiven Einsatz in der Wertermittlung validiert — extrahierte Spalten müssen vor ihrer Verwendung in einem Modell gegen den Quelltext geprüft werden. Ein wirksamer Einsatz der Funktion erfordert praktische Kenntnisse in Prolog und DCGs. Alles, was sie tut, ist additiv und reversibel; bei Fehlverhalten genügt es, den Schalter in den Settings zu deaktivieren und normal fortzufahren.

15.1 Funktionsweise

MLS-Listings enthalten in Freitext- und in listenartigen Feldern eine Fülle wertrelevanter Informationen. Dieser optionale Schritt überführt diesen Text mittels zweier spezieller Spaltenrollen (Kapitel 6) in gewöhnliche Spalten, die von der Regression genutzt werden können:

- **remarks-Spalten** werden von der Prolog-DCG-Grammatik in je eine Merkmalspalte pro extrahiertem Attribut überführt. Das Spaltenpräfix wird aus den Anfangsbuchstaben des Namens der Quellspalte gebildet: `public_remarks` liefert `pr_*`-Spalten, `agent_remark` liefert `ar_*` und so fort.
- **list-Spalten** (z. B. `existing_structures = SSshed, Workshop`) werden in One-Hot-Spalten `<col>_<value>` mit TRUE/FALSE-Werten aufgespalten. Dieser Teil ist reines R und benötigt überhaupt kein Prolog.

Die mitgelieferte Grammatik (`inst/prolog/mls_remarks.pl`) extrahiert derzeit unter anderem: eine abgestufte Zustandsklasse mit einem `condition_score` von 1–5; Aussichts-Flags und Aussichtstypen; Flags für Küchen- und Badmodernisierung; Hartholz- und andere Bodenbeläge; eine Liste und Anzahl von Aufwertungen; Garagenstellplätze und Garagenfläche; Pool, Spa und weitere Außenanlagen; die Anzahl der Kamine; Nebengebäude (Schuppen, Scheune, Werkstatt, Casita, Gästehaus, Zwinger usw.) mit Größen, die aus Maßangaben wie „40x50“ geparkt werden; Stellplätze und Unterstellmöglichkeiten für Wohnmobile (RV); Solareigentum versus Solar-Leasing, Batteriespeicher; Grundstücksgröße (einschließlich der Umrechnung von Acres in Quadratfuß); auf ein Subjekt bezogene Preisangaben aus dem Verkaufskontext (Preisreduktion, HOA-Beiträge, Solar-Leasingrate oder -Ablöse, Mello-Roos); Lagehinweise; sowie einen Sentiment-Score für Vorzüge und Bedenken in $[-1, 1]$.

15.2 Voraussetzungen: vProlog (Trealla), nicht SWI-Prolog

Das Parsen der Remarks läuft über das R-Paket **vProlog** — eine in sich geschlossene Trealla-Prolog-Engine. Es ist nicht auf CRAN verfügbar; die Installation erfolgt aus dem r-universe-Repository des Autors, wie in „Installing earthUI and vProlog“ zu Beginn dieses Artikels gezeigt (vorgefertigte Binärdateien für Windows und macOS, Quelltext unter github.com/wcraytor/vProlog). Eine systemweite Prolog-Installation (SWI-Prolog oder anderes) ist weder erforderlich noch wird sie verwendet. Fehlt vProlog, wird das Parsen der Remarks mit einem Hinweis übersprungen, während die `list`-Aufspaltung weiterhin ausgeführt wird. Die Hilfsfunktion `vProlog_available()` meldet, ob Engine und mitgelieferte Grammatik beide vorhanden sind.

15.3 Aktivierung und Ausführung

1. **Settings** (Zahnrad) öffnen und **Enable Prolog / list column processing** aktivieren. Der Schalter ist standardmäßig deaktiviert, trägt den Hinweis „advanced feature — for users with Prolog expertise“, wird pro Projekt gespeichert und im Project-Information-Dialog angezeigt.
2. In der Prädiktortabelle die Rolle **Special** der betreffenden Spalten auf **remarks** oder **list** setzen (mehrere Spalten können jede der beiden Rollen tragen).
3. Den Abschnitt **Execute Prolog Processing** verwenden, der zwischen den Sections 4 und 5 der Sidebar liegt. Er bietet zwei Schritte:

1. **Expand lists & remarks** — führt die DCG-Grammatik über jede **remarks**-Spalte aus und spaltet jede **list**-Spalte one-hot auf. Die Expansion erfolgt *inkrementell*: Es werden nur die seit dem letzten Lauf neu designierten Spalten verarbeitet, und die Schaltfläche wird deaktiviert, sobald alles Designierte bereits expandiert wurde. Ein Fortschrittsbalken begleitet den Parsevorgang.

2. **Execute derivation rules** — wendet die eigenen Prolog-Regeln `derive/2` auf die expandierten Spalten an (siehe unten). Die Regeln werden vor der Ausführung syntaktisch geprüft und in einem Probelauf validiert.

15.4 Ableitungsregeln

Für jede Datenzeile assertiert earthUI die Spalten der Zeile als Fakten `cell(Name, Value)` und sammelt jede Lösung der eigenen `derive(Col, Val)`-Regeln — jedes distinkte `Col` wird zu einer neuen Spalte, und pro Spalte gewinnt der *erste* gefundene Wert, sodass eine Fallback-Klausel folgen kann. Hilfsprädikate (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) unterstützen Aggregation und NA-Behandlung, und `drop/1`-Fakten verwerfen Spalten nach dem Regellauf. Beispiel:

```
% 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).
```

Die Regeln liegen in einer projektspezifischen Datei `<project>_rules.pl`, die aus einer vollständig kommentierten Vorlage initialisiert wird, und werden über **Edit derivation rules...** in einem frei schwebenden, verschiebbaren Code-Editor bearbeitet, der sich auch in die Sidebar andocken lässt. Eine Schaltfläche **Check** validiert die Regeln (Syntaxprüfung, Ladeprüfung und ein Probelauf gegen Zeile 1 der Daten), ohne sie auszuführen.

15.5 Versionierte aktive Datendatei

Nach jedem der beiden Schritte wird der erweiterte Data Frame als `<original_name>_<YYMMDD_HHMMSS>.xlsx` in den Eingabeordner des Projekts geschrieben und zur **aktiven** Datendatei des Projekts — beim erneuten Öffnen des Projekts werden die erweiterten Daten geladen, sodass Expansionen nie neu berechnet werden, und dieselbe Datei wird auch von `glmnetUI` und `mgcvUI` konsumiert. Die rohen Remarks-Quellspalten werden automatisch aus dem Modell herausgehalten; die erzeugten Spalten werden *nicht* automatisch als Prädiktoren aufgenommen. Da die Textextraktion viele Spalten erzeugen kann, sollten diejenigen, die (noch) nicht in die Regression eingehen sollen, als **Latent** markiert oder per $\times$ deaktiviert werden (Kapitel 6).

15.6 Programmatische Schnittstelle & mitgelieferte Dateien

Die gesamte Pipeline ist für den Batch-Einsatz exportiert: `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` orchestriert die Durchläufe und baut auf `remarks_to_columns()`, `split_list_column()`, `apply_user_rules()` und `validate_rules()` auf. Die mitgelieferten Prolog-Quellen unter `inst/prolog/` sind `mls_remarks.pl` (die DCG-Grammatik und die Merkmalsaggregation), `mls_analyze.pl` (ein Datei-E/A-Treiber für das Batch-Parsing), `mls_rules.pl` (die den Ableitungsregeln zur Verfügung stehenden Hilfsprädikate) und `climate_regions_ca.pl` (siehe unten).

15.7 Verwandtes Experiment: Klimaregionen-Priors

Eine begleitende experimentelle Funktionalität adressiert Merkmale mit *dünnere Datenlage* — solche, die die Regression nicht schätzen kann, weil sie nur in einer Handvoll Vergleichsobjekte vorkommen (etwa ein Verdunstungskühler, engl. „swamp cooler“). Drei R-Hilfsfunktionen — `climate_region_for()`, `market_area_profile()` und `climate_feature_priors()` — ordnen den Standort eines Projekts (County + Stadt) einer von sieben käuferseitig wahrgenommenen kalifornischen Klimaregionen zu (eine praxisorientierte Reduktion der 16 Gebäudeklimazonen der California Energy Commission) und liefern je Region ordinale Priors für den Wertbeitrag von HLK- und Kühlmerkmalen (Heizung, Klimaanlage, Verdunstungskühler, Whole-House-Fan, Filtration). Das Prolog-Modul `climate_regions_ca.pl` spiegelt dieselbe Klassifikation für die Verwendung innerhalb von Ableitungsregeln; beide Seiten werden aus einer einzigen kanonischen R-Quelle generiert, sodass sie nicht auseinanderlaufen können. Die Priors sollen die Merkmalsbewertung *informieren* und als Plausibilitätsprüfung geschätzter Koeffizienten dienen — Regionszuordnungen (insbesondere bei Counties, die sich über mehrere Regionen erstrecken) sollten vom Gutachter überprüft werden.

15.7.1 Über Kalifornien hinaus: Jurisdiction Packs und Market-Area-Module

Kalifornien ist die erste Jurisdiktion, nicht die Grenze. `earthUI` wird international eingesetzt, und das Wissen über Klima-Priors wird derzeit als **Jurisdiction Packs** restrukturiert: in sich geschlossene, datenförmige Faktentabellen (Zonendefinitionen, Zuordnungen von Verwaltungseinheiten, Merkmals-Priors), die nach den Länder- und Verwaltungsebenen-Codes des Projektsystems

geschlüsselt sind, sodass ein Pack einen US-Bundesstaat, ein deutsches Land oder ein kleines Land insgesamt beschreiben kann. Eine begleitende Schicht von **Market-Area-Modulen** erfasst lokales Vokabular und verfeinerte Priors für ein benanntes Marktgebiet.

Beide sind darauf ausgelegt, *im Feld* von Valuation Engineers und Gutachtern in deren eigener Sprache verfasst zu werden — und, realistisch betrachtet, mit LLM-Unterstützung: Die Welt verfügt nicht über viele Prolog-Programmierer, weshalb Feldartefakte bewusst Tabellen und keine Programme sind. Das Pack-Format ist ein dokumentiertes Schema mit einem vollständig ausgearbeiteten Beispiel (Kalifornien selbst), ausdrücklich damit ein Assistent wie Claude aus einer umgangssprachlichen Beschreibung eines Marktes ein gültiges Pack entwerfen kann; earthUI generiert das zugrunde liegende Prolog aus dem Pack, und ein blockierender Validator (Schema-, Abdeckungs- und referenzielle Integritätsprüfungen sowie ein Probelauf) muss bestanden werden, bevor ein Pack aktiviert werden kann. Jedes im Feld verfasste Pack trägt verpflichtende Provenienzangaben — den verantwortlichen Autor, das Datum und die zitierte Autorität —, sodass der verantwortliche Fachmann und nicht das Entwurfswerkzeug für das Wissen entsteht und Berichte das exakt verwendete Pack samt Version zitieren können.

Tipp: Ein bewährtes Muster: die Remarks expandieren, die granularen `pr_*`-Flags mit `derive/2`-Regeln zu einer Handvoll abgeleiteter Zähler oder Boolescher Werte konsolidieren, die konsolidierten Roh-Flags per `drop/1` verwerfen und nur die abgeleiteten Spalten als Kandidaten-Prädiktoren behalten. So bleibt die Prädiktorliste kurz und die Kollinearität beherrschbar.

Literatur

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

glmnetUI: Eine interaktive Benutzeroberfläche für das glmnet-Paket (Lasso und Elastic Net)

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische

Fassung ist maßgeblich.

William Bert Craytor

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

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

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

Dieser Open-Access-Artikel wird unter den Bedingungen der [Creative-Commons-Namensnennung-4.0-Lizenz](#) verbreitet.

Zusammenfassung. glmnetUI ist eine grafische Benutzeroberfläche für das R-Paket `glmnet`, das regularisierte generalisierte lineare Modelle mittels Lasso-, Ridge- und Elastic-Net-Straftermen anpasst. Es bietet drei Verwendungsmodi — allgemeine prädiktive Modellierung, Immobilienbewertung und Marktgebietsanalyse — und führt den Anwender durch Datenimport, Modellkonfiguration, kreuzvalidierte Modellanpassung, Koeffizientenpfade sowie herunterladbare Berichte. Dieser Artikel dokumentiert die Datenformat-Anforderungen, den Modellierungs-Workflow, die Ausgabedarstellungen und die vollständige Funktionsreferenz von glmnetUI.

Schlüsselwörter: Lasso, Ridge-Regression, Elastic Net, regularisierte Regression, glmnet, R, grafische Benutzeroberfläche, Immobilienbewertung

Anmerkung des Verlags (Juli 2026): Kurz vor Drucklegung wurde glmnetUI in earthUI (Version 0.11.0 und höher) zusammengeführt; earthUI stellt den hier beschriebenen Elastic-Net-Workflow nun neben den MARS- und GAM-Workflows bereit. glmnetUI bleibt in seiner letzten veröffentlichten Fassung (0.5.0) unter <https://github.com/wcraytor/glmnetUI> verfügbar; die Weiterentwicklung erfolgt in earthUI. Dieser Artikel bezieht sich auf glmnetUI in der veröffentlichten Form.

Inhaltsverzeichnis

1	Einleitung	3
1.1	Geschichte und Hintergrund	3
1.1.1	Das Lasso	3
1.1.2	Ridge-Regression und Elastic Net	4
1.1.3	Das Paket glmnet	4
1.1.4	Vergleich mit MARS (earth) und GAM (mgcv)	4
1.1.5	Empfohlener Arbeitsablauf	5
1.2	Was ist glmnetUI?	6
1.3	Drei Purpose-Modi	6
1.4	Immobilien-spezifische Funktionen	7
1.5	Erste Schritte	7
2	Anforderungen an die MLS-Eingabedaten	9
2.1	Dateiformat und Struktur	9
2.2	Erforderliche Spalten für den Wertermittlungsmodus	9
2.3	Konventionen für spezielle Spaltenbezeichnungen	10
2.4	Datenqualität und Vollständigkeit	10
2.5	Platzierung der Bewertungsobjektzeile	11

3	Modus General Purpose	12
3.1	Überblick	12
3.1.1	Skip First Row	12
3.1.2	Einstellungen je Purpose	12
3.2	Der Arbeitsablauf in der Seitenleiste	12
3.3	Projekte	13
3.4	Import aus earthUI (optional)	14
3.5	Registerkarten des Hauptbereichs	14
3.6	Persistenz der Einstellungen	14
3.7	Dark Mode	15
3.8	Locale- und regionale Einstellungen	15
3.8.1	Standardwerte speichern	15
4	Modus Appraisal	16
4.1	Behandlung der Bewertungsobjektzeile	16
4.2	Effective Date und Sale Age	16
4.3	Spezielle Spaltenkennzeichnungen	16
4.4	Überblick über die RCA-Anpassungen	16
5	Modus Market Area Analysis	17
5.1	Unterschiede zum Modus Appraisal	17
5.2	Wann der Market-Modus geeignet ist	17
6	Variablenauswahl	18
6.1	Zielvariable	18
6.2	Die Prädiktortabelle	18
6.3	Erkennung und Überschreiben von Datentypen	18
6.4	Erwartete Vorzeichen und deren Durchsetzung	18
6.5	Referenz der Special-Spaltentypen	19
7	Parameterauswahl	20
7.1	Alpha (Mischungsparameter)	20
7.2	Wahl von Lambda	20
7.3	Family	20
7.4	Standardize	21
7.5	Vorzeichenrestriktionen	21
7.6	Relaxed Lasso	21
7.7	Zufallsstartwert	21
7.8	Interaktionsmatrix	22
7.9	Vom Haupteffekt ausschließen (nur Interaktion)	22
7.10	Erweiterte Parameter	22
7.11	Standardwerte der Einstellungen	23
8	Anpassung des Modells	24
8.1	Die Schaltfläche zur Modellanpassung	24
8.2	Was bei der Anpassung geschieht	24
9	Ergebnis-Registerkarten	25
9.1	Data Preview	25
9.2	Equation	25
9.3	Correlation	25
9.4	Summary	25

9.5	Coefficients	26
9.6	Variable Importance	26
9.7	Contributions	26
9.8	ANOVA	26
9.9	Diagnostics	26
9.10	Glmnet Output	27
9.11	Report	27
10	Herunterladen von Daten	28
10.1	Ausgabespalten	28
10.2	CQA-Scores	28
11	RCA-Berechnungen und Herunterladen	29
11.1	Öffnen des RCA-Dialogs	29
11.2	Interpolation des CQA-Werts	29
11.3	Ausgabespalten	29
12	Sales Comparison Grid	30
12.1	Dialog zur Auswahl der Vergleichsobjekte	30
12.2	Aufbau der Arbeitsmappe	30
12.3	Blattschutz und Restmerkmalzellen	30
13	Herunterladen von Berichten	31
13.1	Der zweistufige Quarto-Ablauf	31
13.2	Berichtsformate	31
13.3	Berichtsinhalte	31
14	Vergleich mit earthUI	33
14.1	Wesentliche Unterschiede	33
14.2	Gemeinsame Funktionen	33
14.3	Wann welches Werkzeug	33
15	Demodatensatz: Appraisal_1.csv	35
15.1	Beschreibung	35
15.2	Spalten	35
15.3	Vorgeschlagener Schnelleinstieg	36
16	Systemanforderungen und Fehlerbehebung	37
16.1	Unterstützte Plattformen	37
16.2	Optionale Abhängigkeiten	37
16.3	Plattformhinweise	37
16.4	Verhalten bei fehlenden Komponenten	37
16.5	Fehlerbehebung	37

1 Einleitung

1.1 Geschichte und Hintergrund

1.1.1 Das Lasso

Das **Lasso** (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator) wurde 1996 von Robert Tibshirani eingeführt. Der Name ist ein bewusstes Wortspiel mit dem Lasso des Cowboys — einem Seil,

das einschränkt und auswählt. Tibshirani wählte ihn, weil das Verfahren die Koeffizienten buchstäblich „einfängt“: Es beschränkt ihre Absolutbeträge und zieht einige exakt auf null, wodurch ausgewählt wird, welche Variablen im Modell verbleiben.

Tibshiranis Einsicht bestand darin, der Kleinste-Quadrate-Zielfunktion einen L_1 -Strafterm (Summe der Absolutbeträge der Koeffizienten) hinzuzufügen. Anders als der L_2 -Strafterm der Ridge-Regression (Hoerl & Kennard, 1970), der die Koeffizienten in Richtung null schrumpft, sie aber nie erreicht, erzeugt der L_1 -Strafterm *dünnbesetzte* Lösungen, bei denen viele Koeffizienten exakt null sind. Damit ist das Lasso zugleich ein Regularisierungs- und ein Variablenselektionsverfahren.

1.1.2 Ridge-Regression und Elastic Net

Die **Ridge-Regression** wurde 1970 von Arthur Hoerl und Robert Kennard vorgeschlagen. Sie ergänzt einen L_2 -Strafterm (Summe der quadrierten Koeffizienten), der alle Koeffizienten proportional schrumpft, jedoch sämtliche Variablen im Modell belässt. Ridge ist wirksam, wenn die Prädiktoren korreliert sind (Multikollinearität), vereinfacht das Modell jedoch nicht.

Das **Elastic Net**, 2005 von Hui Zou und Trevor Hastie eingeführt, verbindet beide Strafterme. Der Mischungsparameter α steuert die Gewichtung: $\alpha = 1$ entspricht dem reinen Lasso, $\alpha = 0$ der reinen Ridge-Regression, Werte dazwischen ergeben einen Kompromiss. Das Elastic Net überwindet eine Schwäche des Lasso: Bei stark korrelierten Prädiktoren neigt das Lasso dazu, einen auszuwählen und die übrigen zu ignorieren, während das Elastic Net korrelierte Prädiktoren gruppiert.

1.1.3 Das Paket glmnet

Das R-Paket **glmnet** wurde von Jerome Friedman, Trevor Hastie und Robert Tibshirani an der Stanford University entwickelt. Es implementiert die Elastic-Net-Regularisierung für generalisierte lineare Modelle mithilfe eines äußerst effizienten Koordinatenabstiegs-Algorithmus (Friedman, Hastie und Tibshirani 2010); es wurde seither auf Cox-Modelle proportionaler Hazards (Simon u. a. 2011) und auf sämtliche Familien generalisierter linearer Modelle (Tay, Narasimhan und Hastie 2023) erweitert. Das Paket unterstützt die Familien gaussian (linear), binomial (logistisch), poisson und weitere GLM-Familien.

Zu den wesentlichen Merkmalen von **glmnet** zählen:

- **Regularisierungspfad:** Passt das Modell in einem einzigen Durchlauf über ein Gitter von Lambda-Werten an — weit schneller, als für jedes Lambda einzeln zu schätzen.
- **Kreuzvalidierung:** `cv.glmnet()` wählt die optimale Regularisierungsstärke automatisch über eine k -fache Kreuzvalidierung.
- **Koeffizientenschränken:** Unterstützt obere und untere Grenzen für einzelne Koeffizienten (Vorzeichenrestriktionen).
- **Relaxed Lasso:** Schätzt das Modell auf den ausgewählten Variablen mit verringertem oder ohne Strafterm neu und verringert so die Verzerrung.

1.1.4 Vergleich mit MARS (earth) und GAM (mgcv)

Die earth-Anwendung und ihre beiden Post Processing Modules (Nachverarbeitungsmodule) glmnet und mgcv verwenden unterschiedliche Modellierungsmaschinen. Die folgende Tabelle fasst die wesentlichen Unterschiede zusammen:

Merkmal	glmnet (Elastic Net)	earth (MARS)	mgcv (GAM)
Modelltyp	Linear (mit Regularisierung)	Stückweise linear (adaptive Splines)	Glatt nichtlinear (penalisierte Splines)
Nichtlinearität	Nur über Interaktionen oder Basiserweiterung	Automatisch über Hinge-Funktionen mit datengetriebenen Knoten	Automatisch über Glättungsfunktionen (Thin Plate, kubisch usw.)
Variablenselektion	Eingebaut über L_1 -Strafterm (Lasso)	Eingebaut über Vorwärts-/Rückwärtsschritte mit GCV-Pruning	Optional über Doppelstrafterm-Schrumpfung (<code>select=TRUE</code>)
Interaktionen	Paarweise Kreuzprodukte (schwer interpretierbar)	Hinge-basierte Produkte (interpretierbar, bis Grad 3)	Tensorprodukt-Glättungen (<code>te()</code> , <code>ti()</code>)
Interpretierbarkeit	Koeffizienten sind lineare Gewichte; einzeln leicht erklärbar	g-Funktionen zeigen stückweise lineare partielle Effekte je Variable	Glatte partielle Effektkurven; sehr anschaulich
Geeignet für	Hochdimensionale Daten, Variablenselektion, kleine Stichproben	Automatische Erkennung von Nichtlinearität und Interaktionen	Glatte nichtlineare Zusammenhänge, flexible Modellierung
RCA-Anpassungen	Lineare Anpassungen je Variable	Stückweise lineare Anpassungen über g-Funktionen	Glatte Anpassungen über partielle Effekte
Overfitting-Kontrolle	Kreuzvalidiertes Lambda	GCV-basiertes Pruning	REML-/GCV-Schätzung der Glättungsparameter

MARS (Multivariate Adaptive Regression Splines) wurde 1991 von Jerome Friedman eingeführt. Das Verfahren bildet stückweise lineare Modelle, indem es Hinge-Funktionen $\max(0, x - k)$ und deren Knotenpositionen adaptiv aus den Daten auswählt. Die R-Implementierung ist das Paket **earth** von Stephen Milborrow. MARS eignet sich hervorragend dafür, nichtlineare Zusammenhänge und Interaktionen automatisch aufzudecken, ohne dass der Anwender sie vorgeben muss. Seine g-Funktionen (nach Variablen gruppierte Basisterme) liefern sehr gut interpretierbare partielle Effektkurven.

GAMs (Generalized Additive Models) wurden von Hastie und Tibshirani (1990) formalisiert. Die R-Implementierung **mgcv** von Simon Wood stellt penalisierte Regressionssplines mit automatischer Glättungsauswahl über REML oder GCV bereit. GAMs erzeugen glatte partielle Effektkurven, die anschaulich darzustellen und zu erläutern sind. Wenn **earthUI** Knotenpositionen nach **mgcvUI** exportiert, dienen die aus **earth** abgeleiteten Knoten als Startpunkte für die GAM-Glättungsterme; so verbindet sich die adaptive Knotensetzung von MARS mit der glatten Schätzung des GAM.

1.1.5 Empfohlener Arbeitsablauf

Für die Immobilienbewertung und ähnliche Anwendungen, die interpretierbare und belastbare Modelle erfordern:

1. Mit **earthUI** beginnen, um die wichtigen Variablen, nichtlinearen Zusammenhänge

und Interaktionen in den Daten aufzudecken. Die automatische Basisauswahl und die g-Funktionen von earth liefern ausgezeichnete erste Erkenntnisse.

2. **Mit mgcvUI verfeinern**, wenn glattere partielle Effekte gewünscht sind. Die Knoten aus earthUI lassen sich nach mgcvUI importieren und erlauben so einen nahtlosen Übergang von stückweise linearen zu glatten Modellen.
3. **glmnetUI verwenden**, wenn eine konsequente Variablenselektion (Lasso) benötigt wird, wenn die Zahl der Prädiktoren sich der Zahl der Beobachtungen nähert oder diese übersteigt, oder wenn ein rein lineares Modell mit Regularisierung aus Gründen der Belastbarkeit gewünscht ist.

1.2 Was ist glmnetUI?

glmnetUI ist eine grafische Benutzeroberfläche für das R-Paket **glmnet**. Es läuft als lokale Shiny-Anwendung — es gibt keine Anmeldung, keinen Server und keine Benutzerkonten. Die Anwendung wird aus R heraus gestartet; anschließend wird ein Datensatz (CSV oder Excel) importiert, das Modell konfiguriert und interaktiv angepasst.

Seit der aktuellen Version ist glmnetUI kein separat installiertes Paket mehr: Es wird *innerhalb* des Pakets earthUI als dessen Elastic-Net Post Processing Module ausgeliefert und mit `earthUI::launch_glmnet()` gestartet. Die Anwendung selbst — und alles in diesem Artikel Beschriebene — bleibt unverändert; lediglich Installation und Startbefehl unterscheiden sich. Die drei Routinen (earth, glmnet, mgcv) teilen sich einen Projektbaum, eine Einstellungsdatenbank und eine Kopie der unterstützenden Infrastruktur.

Die Anwendung stellt einen vollständigen Arbeitsablauf bereit: Datenimport, Variablenkonfiguration, Modellanpassung, Diagnosegrafiken, Variablenwichtigkeit, Modellgleichungen sowie herunterladbare Berichte im Word-, PDF- oder HTML-Format.

Die Elastic-Net-Regression verbindet zwei Regularisierungstechniken:

- **Ridge-Regression** ($\alpha = 0$): Schrumpft alle Koeffizienten proportional in Richtung null. Geeignet, wenn alle Variablen als relevant angesehen werden.
- **Lasso-Regression** ($\alpha = 1$): Kann Koeffizienten exakt auf null schrumpfen und führt damit eine automatische Variablenselektion durch. Erzeugt einfachere, besser interpretierbare Modelle.
- **Elastic Net** ($0 < \alpha < 1$): Eine Mischung beider Ansätze. Nützlich bei korrelierten Prädiktoren.

1.3 Drei Purpose-Modi

Jedes glmnetUI-Projekt trägt einen **Purpose** (Verwendungszweck), der bei der Erstellung gewählt wird (Abschnitt 1, *Project*). Der Purpose des aktiven Projekts bestimmt, welche Werkzeuge und Oberflächenelemente erscheinen; um in einem anderen Modus zu arbeiten, wird ein Projekt mit dem entsprechenden Purpose angelegt oder geöffnet. Es gibt drei Modi:

- **General** (Allgemein) — Elastic-Net-Regression für beliebige Populationen oder Datensätze. Dies ist der Standardmodus. Er stellt den vollständigen regularisierten Regressionsablauf ohne domänenspezifische Ergänzungen bereit.
- **For Appraisal** (für die Wertermittlung) — Elastic-Net-Regression, zugeschnitten auf die Immobilienbewertung. Ergänzt Funktionen speziell für die Bewertung einzelner Objekte, darunter die Behandlung des Bewertungsobjekts, spezielle Spaltenkennzeichnungen sowie den Residual Constraint Approach (RCA) (dt. etwa: Residualrestriktionsansatz).
- **Market Area Analysis** (Marktgebietsanalyse) — Elastic-Net-Regression, zugeschnitten auf Marktgebietsstudien. Ergänzt Funktionen zur Analyse von Objektgruppen in einem definierten Markt, mit optionalem Ausschluss der Bewertungsobjektzeile.

In allen drei Modi ist die zentrale Modellierungsmaschine identisch — es wird stets ein Elastic-Net-Modell (glmnet) angepasst. Die Purpose-Einstellung steuert, welche zusätzlichen Werkzeuge und Oberflächenelemente verfügbar sind.

1.4 Immobilienspezifische Funktionen

Wenn entweder **For Appraisal** oder **Market Area Analysis** gewählt ist, aktiviert glmnetUI mehrere Funktionen, die für die Immobilienanalyse konzipiert sind:

- **Spezielle Spaltenkennzeichnungen** — Jeder Prädiktor kann mit einer speziellen Rolle versehen werden, etwa `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` oder `display_only`. Diese Kennzeichnungen steuern, wie die Spalte bei der Modellanpassung und in der Ausgabe behandelt wird.
- **Rundung von Breiten- und Längengrad** — Als `latitude` oder `longitude` gekennzeichnete Spalten werden automatisch auf drei Nachkommastellen gerundet, um Overfitting zu vermeiden.
- **Spalte Sale Age (Verkaufsalter)** — Ist eine Spalte als `contract_date` gekennzeichnet und ein Effective Date (Stichtag) angegeben, berechnet glmnetUI eine Spalte `sale_age` (Tage zwischen Verkaufsdatum und Stichtag) und setzt sie als Prädiktor ein.
- **RCA-Berechnungen (nur Wertermittlung)** — Im Wertermittlungsmodus kann glmnetUI nach der Modellanpassung eine Ausgabe zum Residual Constraint Approach (RCA) erzeugen, die Anpassungen je Vergleichsobjekt, Netto-/Brutto-Anpassungsübersichten sowie einen angepassten Verkaufspreis für das Bewertungsobjekt liefert.

***Hinweis:** Der Purpose-Modus General eignet sich für beliebige Datensätze — aus Finanzwesen, Wissenschaft, Ingenieurwesen oder Immobilienwirtschaft. Die Modi für Wertermittlung und Marktanalyse fügen lediglich Komfortfunktionen für Immobilienfachleute hinzu.*

1.5 Erste Schritte

Zur Verwendung von glmnetUI:

1. **earthUI in R installieren:** `install.packages("earthUI")` — die glmnet-Routine ist enthalten. (Siehe „Installing earthUI and vProlog“ im earthUI-Artikel dieser Ausgabe für die Entwicklungsversion und optionale Komponenten.)
2. **Anwendung starten:** `earthUI::launch_glmnet()` in R ausführen. Die Anwendung öffnet sich im Webbrowser auf Port 7879. Sie lässt sich auch direkt über `http://localhost:7879` im Browser aufrufen. Die Anwendung merkt sich den zuletzt verwendeten Purpose-Modus und stellt ihn automatisch wieder her.
3. **Projekt anlegen oder öffnen** in Abschnitt 1 der Seitenleiste. Ein Projekt legt den Purpose und den Arbeitsort (Land, Bundesstaat, County, Stadt) sowie seine Eingabe- und Ausgabeordner fest. Projekte werden unter einem gemeinsamen `regProj`-Wurzelverzeichnis abgelegt und mit den Schwesteranwendungen earthUI und mgcvUI geteilt. Das Öffnen eines Projekts mit anderem Purpose ist der Weg, den Modus zu wechseln.
4. **Daten importieren** in Abschnitt 2 durch Auswahl einer CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts. Dateien in diesen Ordner legen und anschließend auf **Refresh** (Aktualisieren) klicken, um sie aufzulisten.
5. **Import aus earthUI (optional)** — Abschnitt 3 ermöglicht den Import einer earthUI-Ergebnisdatei (`.rds`), um die Hinge-Basisfunktionen von earth mit der Regularisierung von glmnet zu verbinden. Dieser Schritt entfällt, wenn kein earthUI-Ergebnis vorliegt.
6. **Variablen konfigurieren** — Zielvariable und Prädiktoren wählen, Datentypen und erwartete Vorzeichen festlegen und etwaige spezielle Spaltenrollen zuweisen.

7. **glmnet-Parameter einstellen** — alpha, lambda, family, Vorzeichenrestriktionen, Interaktionen und weitere Optionen.
8. **Modell anpassen** — auf „Fit Glmnet Model“ klicken und die Ergebnisse im Hauptbereich prüfen.
9. **Exportieren** — Prognosen als Excel-Datei herunterladen, einen Quarto-Bericht erzeugen und konvertieren oder (im Wertermittlungsmodus) RCA-Anpassungen und ein Sales Comparison Grid (Vergleichsraster) berechnen.

Die Einstellungen werden automatisch im lokalen Speicher des Browsers gesichert und wiederhergestellt, sobald dieselbe Eingabedatei erneut geladen wird.

2 Anforderungen an die MLS-Eingabedaten

Für Arbeitsabläufe der Immobilienbewertung und Marktanalyse stammen die Eingabedaten üblicherweise aus einem Export eines Multiple Listing Service (MLS). Dieses Kapitel beschreibt die erwartete Dateistruktur und die Spalten, die glmnetUI verwenden kann.

2.1 Dateiformat und Struktur

glmnetUI akzeptiert **CSV**- und **Excel**-Dateien (**.xlsx**, **.xls**). Beim Import werden Spaltennamen automatisch in **snake_case** umgewandelt — aus „Living SqFt“ wird beispielsweise **living_sqft**, aus „Contract Date“ wird **contract_date** und aus „Sale Price“ wird **sale_price**. Diese Normalisierung gewährleistet einheitliche Spaltenbezüge über den gesamten Arbeitsablauf. Das beim Import verwendete CSV-Trennzeichen und Dezimaltrennzeichen ergibt sich aus den Locale-Einstellungen (siehe Kapitel 3, „Locale & Regional Settings“).

Die Datendatei sollte eine flache Tabelle mit einer Zeile je Objekt und einer Spalte je Merkmal sein. Die erste Zeile der Datei muss die Spaltenüberschriften enthalten.

2.2 Erforderliche Spalten für den Wertermittlungsmodus

glmnetUI arbeitet zwar mit beliebigen Spaltensätzen, doch der vollständige Wertermittlungsablauf profitiert von den folgenden Spalten:

Spalte	Special-Typ	Zweck
Sale Price	(Ziel)	Antwortvariable des Modells
Contract Date	contract_date	Dient zur Berechnung von sale_age (Tage ab Stich-tag)
Listing Date	listing_date	Zusammen mit dem Vertragsdatum zur DOM-Berechnung, falls keine DOM-Spalte vorhanden ist
Days on Market	dom	Vermarktungsdauer in Tagen; wird in Exporten angezeigt
Concessions	concessions	Verkaufszugeständnisse; Netto-Verkaufspreis = Verkaufspreis – Zugeständnisse
Living Area (SF)	living_area	Ermöglicht Residuen je Quadratfuß (residual_sf , cqa_sf)
Lot Size	lot_size	Spalte für die Grundstücksgröße
Site Dimensions	site_dimensions	Mit der Grundstücksgröße gruppiert
Latitude	latitude	Auf 3 Nachkommastellen gerundet
Longitude	longitude	Auf 3 Nachkommastellen gerundet
Area ID	area	Kennung des Marktgebiets bzw. der Lage
Actual Age	actual_age	Objektalter
Effective Age	effective_age	Effektives Objektalter
Address	display_only	Wird in Exporten angezeigt; vom Modell ausgeschlossen

Die Spaltennamen der Tabellenkalkulation dürfen fremdsprachig sein — die „speziellen“ Bezeichnungen sind englisch, damit das R-Programm sie gesondert behandeln kann. Im Übrigen erscheinen die vergebenen Spaltennamen in den Regressionsmodellen, den Grafiken und — bei Wertermittlungen — in den Ausgabeberichten.

Nicht alle Spalten sind erforderlich. glmnetUI passt sich an — fehlt eine Spalte, entfällt schlicht die zugehörige Funktion. Für Immobilienpreismodelle sind jedoch bestimmte Spalten dringend zu empfehlen, um eine akzeptable Anpassungsgüte zu erreichen:

1. **Sale Age** (Verkaufsalter) — die Anzahl der Tage zwischen dem Vertragsdatum des Verkaufs und dem Stichtag der Wertermittlung oder Analyse. Wird eine mehrjährige Verkaufshistorie verwendet, besonders über Zeiträume von mehr als fünf Jahren, spielt `sale_age` bei der Schätzung des Verkaufspreises häufig eine zentrale Rolle.
2. **Living Area** (Wohnfläche) — auch unter Bezeichnungen wie „Living Sqft“, „GLA“ (gross living area) und ähnlich geführt. Dies ist eine weitere zentrale Bestimmungsgröße des Verkaufspreises.
3. **Total Bath Count** (Gesamtzahl der Bäder) — die Gesamtzahl der ganzen, Viertel-, halben und Dreiviertelbäder. Zwei Vollbäder und ein Gästebad ergeben beispielsweise den Wert 2,5.
4. **Garage Bays** oder **Garage Area** — die Anzahl der Garagenstellplätze oder die Garagenfläche.
5. **Lot Size** (Grundstücksgröße) — die Grundstücksfläche des Objekts, üblicherweise in Quadratfuß oder Acres.
6. **Longitude**, **Latitude** und, sofern verfügbar, **Area ID**. Diese Lagevariablen helfen dem Modell, geografische Preisunterschiede abzubilden.

2.3 Konventionen für spezielle Spaltenbezeichnungen

glmnetUI erkennt Spalten anhand ihrer **Special-Kennzeichnung**, nicht anhand ihres Spaltennamens. Die Spalten dürfen im MLS-Export beliebig benannt sein — entscheidend ist, dass in der Tabelle zur Variablenkonfiguration (Kapitel 6) der richtige Special-Typ zugewiesen wird.

So könnte das MLS die Wohnfläche etwa als „GLA“, „Living SqFt“, „liv_area“ oder „gross_living_area“ exportieren. Nach dem Import (bei dem der Name in snake_case umgewandelt wird) genügt es, sie im Special-Auswahlfeld als `living_area` zu kennzeichnen. glmnetUI verwendet sie dann unabhängig vom ursprünglichen Namen für die Berechnung der Residuen je Quadratfuß.

2.4 Datenqualität und Vollständigkeit

- **Fehlende Werte (NA)**: Zeilen mit NA-Werten in einer Prädiktor- oder Zielspalte werden vor der Modellanpassung automatisch entfernt.
- **Datumsspalten**: Zeichenspalten, die gängigen Datumsformaten entsprechen, werden beim Import automatisch geparkt. Formate mit zweistelliger Jahreszahl werden bevorzugt, wenn keine vierstellige Jahreszahl erkannt wird.
- **Numerische Spalten**: Verkaufspreis, Wohnfläche, Grundstücksgröße, Zugeständnisse und ähnliche Felder müssen numerisch sein. Exportiert das MLS Preise mit Währungssymbolen oder Tausendertrennzeichen (etwa „\$350,000“ oder „350.000,00“), sind diese gegebenenfalls vor dem Import zu bereinigen.
- **Faktorspalten**: Text-, TRUE/FALSE- und R-Faktorspalten werden automatisch als kategorial behandelt — eine Regression kann sie nur als Faktoren verarbeiten. Eine **numerische** Spalte (etwa ein numerischer Gebiets- oder Stilcode) wird nur dann zum Faktor, wenn das zugehörige Kontrollkästchen **Factor** gesetzt wird; glmnetUI leitet aus dem Wertebereich einer numerischen Spalte niemals einen kategorialen Charakter ab, sodass diskrete numerische Prädiktoren wie `bath_count` stetig bleiben. Die Special-Rolle ist von der Faktorentscheidung unabhängig.

Hinweis: Nach dem Import empfiehlt sich ein Blick auf die NA-Spalte in der Prädiktortabelle. Spalten mit vielen fehlenden Werten können dazu führen, dass Zeilen entfallen. Spalten mit hohem NA-Anteil sollten ausgeschlossen oder die Daten vor dem Import bereinigt werden.

2.5 Platzierung der Bewertungsobjektzeile

Im Modus **Appraisal** muss **Zeile 1** das **Bewertungsobjekt enthalten**. Alle übrigen Zeilen sind Vergleichsverkäufe. Die Bewertungsobjektzeile wird von der Modellanpassung ausgeschlossen. Nach der Anpassung erzeugt das Modell dennoch Prognosen für diese Zeile.

Im Modus **Market Area Analysis** ist die Platzierung des Bewertungsobjekts in Zeile 1 optional.

Im Modus **General** gibt es keine besondere Zeilenbehandlung — alle Zeilen werden gleich behandelt.

3 Modus General Purpose

3.1 Überblick

Der Modus General Purpose ist beim Start von glmnetUI voreingestellt. Er stellt den vollständigen Elastic-Net-Regressionsablauf für beliebige Datensätze bereit — nicht nur für Immobilien. glmnetUI lässt sich für wissenschaftliche Daten, Finanzanalysen, ingenieurtechnische Studien oder jedes andere Regressionsproblem einsetzen.

Im Modus General entfallen in der Oberfläche die immobiliespezifischen Funktionen (Special-Spalten, Verkaufsalter, Koordinatenrundung, RCA). Die Seitenleiste ist verschlankt und konzentriert sich auf Variablenauswahl, Parameterkonfiguration, Modellanpassung und Export.

3.1.1 Skip First Row

Ein Kontrollkästchen **Skip first row** (erste Zeile überspringen) erscheint oben in der Seitenleiste, sobald ein General- oder Market-Projekt geöffnet ist. Ist es gesetzt, wird Zeile 1 von der Modellanpassung ausgeschlossen. Das ist nützlich, wenn Zeile 1 eine Ziel- oder Referenzbeobachtung enthält, die das Modell nicht beeinflussen soll. Im Modus Appraisal wird Zeile 1 (das Bewertungsobjekt) stets automatisch ausgeschlossen.

3.1.2 Einstellungen je Purpose

Einstellungen (Prädiktorauswahl, Modellparameter, Interaktionen) werden für jede Kombination aus Eingabedatei und Purpose-Modus getrennt gespeichert. Beim Wechsel zwischen General, Appraisal und Market bleiben die Einstellungen jedes Modus unabhängig voneinander erhalten.

3.2 Der Arbeitsablauf in der Seitenleiste

Die Seitenleiste ist in nummerierte, ein- und ausklappbare Abschnitte gegliedert, die von der Projektauswahl bis zum Berichtsexport führen. Die meisten Abschnitte erscheinen erst, wenn ein Projekt geöffnet ist, und mehrere sind an den Purpose des Projekts gebunden. Der Ausgabeordner ergibt sich aus dem aktiven Projekt — ein separates Feld für den Ausgabeordner gibt es nicht.

1. Project (Projekt) — Projekt anlegen oder öffnen. Ein Projekt legt den Purpose und den Arbeitsort (Land, Bundesstaat, County, Stadt) sowie seinen Eingabeordner (`<os>_in/`) und seine Ausgabeordner fest. Projekte werden unter einem gemeinsamen **regProj**-Wurzelverzeichnis abgelegt (Standard `~/regProj`, überschreibbar mit der Umgebungsvariablen `REGPROJ_ROOT`) und mit den Schwesteranwendungen earthUI und mgcvUI geteilt, sodass Modelle aller drei Anwendungen nebeneinander liegen. Ein Klick auf **Close Project** wechselt das Projekt.

2. Import Data (Daten importieren) — Eine CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts auswählen. Dateien über Finder/Explorer in den Ordner legen und anschließend auf **Refresh** klicken, um sie aufzulisten. Bei Excel-Dateien mit mehreren Blättern erscheint eine Blattauswahl. Spaltennamen werden automatisch in `snake_case` umgewandelt.

3. Import from earthUI (optional) (Import aus earthUI) — Eine earthUI-Ergebnisdatei (`.rds`) importieren, um die Hinge-Basisfunktionen von earth mit der Elastic-Net-Regularisierung von glmnet zu verbinden. Die Schaltfläche zum Durchsuchen ist optisch an die Dateiauswahl in Abschnitt 2 angeglichen. Dieser Schritt ist optional — er entfällt, wenn kein earthUI-Ergebnis zum Import vorliegt.

4. Variable Configuration (Variablenkonfiguration) — Auswahlfeld für die Zielvariable, Prädiktortabelle mit Kontrollkästchen für Include, Factor und Force sowie Sign-Auswahlfeldern. Die Spalte Special (zur Kennzeichnung von Vertragsdaten, Koordinaten usw.) erscheint nur in den Modi Appraisal und Market. Einzelheiten in Kapitel 6.

5. glmnet Call Parameters (glmnet-Aufrufparameter) — Die gesamte Modellkonfiguration: alpha, lambda, family, standardize, Vorzeichenrestriktionen, Relaxed Lasso, Zufallsstartwert, Interaktionsmatrix sowie erweiterte Parameter (lambda.min.ratio, nlambda, CV-Verlustmaß, Konvergenzschwelle, maximale Iterationszahl, Achsenabschnitt). Die vollständige Parameterreferenz enthält Kapitel 7.

6. Fit Glmnet Model (Modell anpassen) — Die Schaltfläche, die das Modell rechnet. Die Ergebnisse erscheinen in den Registerkarten rechts.

7. Download Output (Ausgabe herunterladen) — Exportiert Prognosen, Residuen, CQA-Scores und Beiträge je Variable als Excel-Datei in den Ausgabeordner des Projekts. Nach der Modellanpassung in allen Purpose-Modi verfügbar.

8. Calculate RCA Adjustments & Download (RCA-Anpassungen berechnen und herunterladen) — Berechnet RCA-Anpassungen je Variable für jedes Vergleichsobjekt und interpoliert das Residuum des Bewertungsobjekts über den CQA-Wert. Erscheint nur in den Modi Appraisal und Market nach der Modellanpassung. Siehe Kapitel 11.

9. Generate Sales Grid & Download (Vergleichsraster erzeugen und herunterladen) — Erzeugt das vorläufige Sales Comparison Grid und ordnet die Vergleichsobjekte nach dem prozentualen Bruttoanpassungssatz. Erscheint nur im Modus Appraisal, nachdem die RCA-Anpassungen berechnet wurden. Siehe Kapitel 12.

10. Generate Quarto Report (Quarto-Bericht erzeugen) — Schreibt ein in sich geschlossenes Quarto-Berichtspaket (.qmd) — mit allen Grafiken und Tabellen — in den Ausgabeordner des Projekts. Nach der Modellanpassung verfügbar.

11. Convert Quarto Report (Quarto-Bericht konvertieren) — Rendert das erzeugte .qmd-Paket nach HTML, Word oder PDF. Für PDF ist eine LaTeX-Installation erforderlich; wird keine erkannt, wird die PDF-Option ausgeblendet. Siehe Kapitel 13.

In den Modi Market und General entfallen die nur für die Wertermittlung vorgesehenen Schritte, und die beiden Berichtsabschnitte werden entsprechend neu nummeriert.

3.3 Projekte

glmnetUI organisiert die Arbeit in **Projekten** unterhalb eines gemeinsamen **regProj**-Wurzelverzeichnisses (desselben Baums, den auch earthUI und mgcvUI verwenden). Jedes Projekt besitzt einen Eingabeordner (<os>_in) und je Modellierungsmaschine eigene Ausgabeordner (z. B. <os>_out_glmnet).

Zum Anlegen eines Projekts auf **+ New...** in Abschnitt 1 klicken und Folgendes angeben:

- **Purpose** — General, Appraisal oder Market Area.
- **Location** — die politisch-administrative Hierarchie des Projekts. Welche Ebenen auszufüllen sind, hängt vom gewählten Land ab, sodass ausländische Benennungskonventionen automatisch unterstützt werden — ein US-Projekt verwendet State/County/City, ein französisches Region/Département/Commune, ein japanisches Präfektur/Stadt-oder-Bezirk/Distrikt und so fort.
- **Project Name** — eine kurze Bezeichnung, begrenzt auf 8 Zeichen (Buchstaben, Ziffern, _ und -).

Die Anwendung übernimmt den eingegebenen Namen nicht wörtlich: Sie bildet einen eindeutigen, mit Zeitstempel versehenen Ordnernamen, indem sie die Ortscodes und den Erstellungszeitpunkt voranstellt, in der Form <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name> — aus einem US-Projekt namens burl wird beispielsweise us_ca_081_burlin_20260604-131500_burl. Die Grenze von 8 Zeichen gilt nur für den selbst eingegebenen Teil. Nur neu angelegte Projekte erhalten das Zeitstempel-Präfix; bereits auf der Festplatte vorhandene Projekte behalten ihre ursprünglichen Namen und bleiben vollständig zugänglich.

3.4 Import aus earthUI (optional)

Abschnitt 3 der Seitenleiste — **Import from earthUI** — ermöglicht den Import einer earthUI-Ergebnisdatei (`.rds`). Dadurch kann glmnetUI die Hinge-Basisfunktionen von earth (stückweise lineare Terme) mit der Elastic-Net-Regularisierung von glmnet verbinden und so die automatische Nichtlinearitätserkennung von earth mit der Variablenselektion und den Koeffizientenrestriktionen von glmnet kombinieren.

Wird ein earthUI-Ergebnis importiert, bestimmen die Prädiktoren des earth-Modells — nicht die Include-Kontrollkästchen — die glmnet-Anpassung, denn glmnet rechnet auf der Basiserweiterung von earth. Die Tabelle zur Variablenkonfiguration macht dies deutlich: **Include** und **Type** sind gesperrt, Zeilen für Prädiktoren, die das earth-Modell nicht verwendet hat, sind deaktiviert, und ein Hinweis erläutert, dass der Prädiktorsatz aus dem earth-Modell stammt. **Force** und **Sign** bleiben für die Prädiktoren von earth bearbeitbar, da sie zum Zeitpunkt der Anpassung weiterhin auf die earth-Basis wirken.

Die Schaltfläche zum Durchsuchen ist optisch an die Dateiauswahl in Abschnitt 2 angeglichen. Dieser Schritt ist vollständig optional — liegt keine earthUI-Ergebnisdatei vor, wird Abschnitt 3 einfach übersprungen und mit Abschnitt 4 fortgefahren.

3.5 Registerkarten des Hauptbereichs

Nach der Modellanpassung stellt der Hauptbereich die folgenden Registerkarten bereit:

Registerkarte	Inhalt
Data Preview	Vorschau der importierten Daten. Im Wertermittlungsmodus aufgeteilt in Tabellen für Bewertungsobjekt und Vergleichsverkäufe.
Equation	Die angepasste Modellgleichung, dargestellt in LaTeX/MathJax, mit allen von null verschiedenen Koeffizienten und Interaktionstermen.
Correlation	Heatmap der Korrelationen zwischen numerischen Prädiktoren (bereits vor der Anpassung verfügbar).
Summary	Kennzahlen (R^2 , adj. R^2 , GR^2 , $CV\text{-}R^2$, RMSE, MAE) sowie Kennzahlen der Wertermittlung (COD, PRD, Median Ratio).
Coefficients	Tabelle der von null verschiedenen Koeffizienten mit Warnhinweisen bei Vorzeichenverletzungen.
Variable Importance	Balkendiagramm und Rangtabelle der Prädiktorwichtigkeit ($ \beta \times sd(x)$). Interaktive plotly-Tooltips, sofern verfügbar.
Contributions	Grafiken der partiellen Effekte je Variable mit Steigungsbeschriftungen. Streudiagramm plus Anpassungsgerade für numerische Prädiktoren, Box-plots für Faktoren, 3D-Fläche/Streudiagramm/Heatmap für Interaktionen.
ANOVA	Tabelle der Varianzzerlegung: SS je Variable, % der Modell-SS, Koeffizient.
Diagnostics	Vier Grafiken: Koeffizientenpfad, CV-Fehler, beobachtet gegen prognostiziert, Residuen gegen angepasste Werte.
Glmnet Output	Rohausgabe des Modells: verwendeter Zufallsstartwert, Modellausdruck, gewähltes Lambda, $\lambda_{min}/1se$, Gamma (bei Relaxed Lasso) und vollständiger Koeffizientenvektor.
Report	Berichtskonfiguration und Exportfelder.
RCA Adjustments	Ergebnisse der RCA-Analyse (Wertermittlungsmodus, nach der Berechnung).

3.6 Persistenz der Einstellungen

glmnetUI speichert die Konfiguration automatisch im lokalen Speicher des Browsers, adressiert über den Namen der Eingabedatei. Wird dieselbe Datei erneut geladen, werden alle Einstellungen

wiederhergestellt: Zielauswahl, Prädiktor-Kontrollkästchen, Datentypen, erwartete Vorzeichen, glmnet-Parameter und Interaktionsmatrix. Auch der zuletzt verwendete Purpose-Modus wird global gesichert und beim Neustart der Anwendung wiederhergestellt. Drei Optionen stehen zur Verfügung:

- **Use last settings for input file** — die für diese konkrete Datei gespeicherten Einstellungen wiederherstellen
- **Use default settings** — die gespeicherten globalen Standardwerte anwenden
- **glmnet defaults** — auf die Werkseinstellungen zurücksetzen (alpha = 1, CV mit lambda.1se, 10 Folds, gaussian, standardize aktiviert)

Eine Schaltfläche **Save current as default** speichert sämtliche aktuellen Einstellungen als globalen Standard.

3.7 Dark Mode

Ein Klick auf das Mond-/Sonnensymbol in der rechten oberen Ecke schaltet zwischen hellem und dunklem Design um. Die Designwahl wird im lokalen Speicher gesichert und bleibt über Sitzungen hinweg erhalten. Sämtliche Oberflächenelemente (Tabellen, Grafiken, Karten, Schaltflächen) passen sich über CSS-Variablen dem gewählten Design an.

3.8 Locale- und regionale Einstellungen

glmnetUI unterstützt internationale Zahlen- und CSV-Konventionen über ein länderbasiertes Locale-System. Das Auswahlmenü **Settings** in der Titelleiste bietet die Selektoren **Country** und **Paper** für 31 unterstützte Länder. Jede Voreinstellung konfiguriert:

- **CSV-Trennzeichen** — Komma (,) für USA/Vereinigtes Königreich/Japan oder Semikolon (;) für den Großteil Europas, wo das Komma als Dezimaltrennzeichen dient.
- **Dezimaltrennzeichen** — Punkt (.) oder Komma (,).
- **Tausendertrennzeichen** — Komma (USA/Vereinigtes Königreich/Japan), Punkt (Deutschland/Italien/Spanien), Leerzeichen (Finnland/Frankreich/Polen/Baltikum/Ukraine/Russland) oder Apostroph (Schweiz).
- **Papierformat** — Letter (USA/Kanada/Mexiko) oder A4 (überall sonst).

3.8.1 Standardwerte speichern

Ein Klick auf **Save as my default** speichert die Locale-Einstellungen global. Diese Standardwerte gelten für alle künftigen Sitzungen, unabhängig davon, welche Datendatei geladen wird.

***Hinweis:** Die Zahlenformatierung auf Diagrammachsen und Steigungsbeschriftungen passt sich automatisch an das eingestellte Locale an. Das deutsche Locale verwendet Punkte als Tausendertrennzeichen (200.000), das finnische Leerzeichen (200 000), das schweizerische Apostrophe (200'000). Währungssymbole werden nicht angezeigt — glmnetUI ist währungsneutral.*

4 Modus Appraisal

Wird **For Appraisal** als Purpose gewählt, richtet sich glmnetUI auf die Bewertung eines einzelnen Objekts aus. Alle in Kapitel 3 beschriebenen Funktionen bleiben verfügbar; dieses Kapitel behandelt nur die wertermittlungsspezifischen Ergänzungen.

4.1 Behandlung der Bewertungsobjektzeile

Im Wertermittlungsmodus ist **Zeile 1 des Datensatzes das Bewertungsobjekt**, und alle übrigen Zeilen sind Vergleichsverkäufe. Die Eingabedatei muss entsprechend aufgebaut sein (siehe Kapitel 2). Der Verkaufspreis des Bewertungsobjekts kann leer bleiben oder einen beliebigen Wert enthalten — glmnetUI behandelt ihn bei der Anpassung automatisch als NA.

Nach dem Import teilt sich die Registerkarte Data Preview in zwei Bereiche: **Subject Property** (Zeile 1) und **Comparable Sales** (Zeilen 2 ff.). Zeile 1 wird stets von der Modellanpassung ausgeschlossen. Nach der Anpassung erzeugt das Modell dennoch Prognosen für diese Zeile.

4.2 Effective Date und Sale Age

In den Modi Appraisal und Market erscheint im Abschnitt zur Variablenkonfiguration ein Feld **Effective Date** (Stichtag), vorbelegt mit dem heutigen Datum. Wird eine Spalte im Special-Auswahlfeld als **contract_date** gekennzeichnet, berechnet glmnetUI eine Spalte **sale_age** — die ganzzahlige Anzahl der Tage zwischen dem Vertragsdatum des jeweiligen Verkaufs und dem Stichtag. Diese Spalte wird als Prädiktor ergänzt.

Ändert sich der Stichtag, wird **sale_age** automatisch neu berechnet.

4.3 Spezielle Spaltenkennzeichnungen

In den Modi Appraisal und Market erscheint für jeden Prädiktor in der Tabelle zur Variablenkonfiguration ein Auswahlfeld **Special**. Die vollständige Referenz der Special-Typen und ihrer Wirkung enthält Kapitel 6.

4.4 Überblick über die RCA-Anpassungen

Die Schaltfläche **Calculate RCA Adjustments & Download** (Seitenleistenabschnitt 8, in den Modi Appraisal und Market nach der Modellanpassung sichtbar) berechnet marktabgeleitete Anpassungen für jedes Vergleichsobjekt relativ zum Bewertungsobjekt. Der vollständige RCA-Ablauf wird in Kapitel 11 beschrieben.

Hinweis: Der dem Bewertungsobjekt zugewiesene CQA-Wert steuert, welcher Anteil der Residualverteilung dem Bewertungsobjekt zugerechnet wird. Ein Wert von 5.00 platziert das Bewertungsobjekt am Median der Vergleichsobjekte.

5 Modus Market Area Analysis

Wird **Market Area Analysis** als Purpose gewählt, stellt glmnetUI dieselben immobilien-spezifischen Funktionen bereit wie der Wertermittlungsmodus (Special-Spalten, Verkaufsalter, Koordinatenrundung), richtet sich dabei aber auf die Analyse einer Objektgruppe statt auf die Bewertung eines einzelnen Bewertungsobjekts aus.

5.1 Unterschiede zum Modus Appraisal

- **Die Behandlung der Bewertungsobjektzeile ist flexibel** — im Market-Modus wird Zeile 1 nicht automatisch als Bewertungsobjekt behandelt.
- **Kein Sales Comparison Grid** — der Schritt „Calculate RCA Adjustments & Download“ (Abschnitt 8) steht im Market-Modus zur Verfügung, der Schritt Sales Grid (Abschnitt 9) jedoch nur im Wertermittlungsmodus.

5.2 Wann der Market-Modus geeignet ist

Der Modus Market Area Analysis empfiehlt sich, wenn Sie:

- ein Regressionsmodell für eine Lage oder ein Marktgebiet aufbauen, um die Werttreiber zu verstehen
- analysieren, wie Variablen wie Wohnfläche, Alter, Grundstücksgröße und Lage die Verkaufspreise einer Objektgruppe beeinflussen
- Unterlagen für eine Marktkonjunkturanalyse oder eine Abgrenzung der Lage vorbereiten
- mit einem Datensatz arbeiten, für den die Special-Spalten-Funktionen (Verkaufsalter, Koordinatenrundung) benötigt werden, der RCA-Anpassungsablauf jedoch nicht

Hinweis: Der Market Mode eignet sich auch für allgemeine Immobilienregressionen, bei denen die Special-Spalten-Funktionen benötigt werden, der RCA-Anpassungsablauf jedoch nicht.

6 Variablenauswahl

In Abschnitt 4 der Seitenleiste — **Variable Configuration** — wird festgelegt, welche Spalten am Modell teilnehmen und wie sie behandelt werden.

6.1 Zielvariable

Das Auswahlfeld **Target (response) variable** am Kopf von Abschnitt 4 listet sämtliche Spalten des Datensatzes auf. Eine Spalte ist als Antwortvariable zu wählen (etwa der Verkaufspreis). Die Zielspalte wird automatisch aus der Prädiktorliste entfernt.

6.2 Die Prädiktortabelle

Unterhalb der Zielauswahl listet eine Tabelle alle verbleibenden Spalten mit den folgenden Feldern auf:

Spalte	Beschreibung
Variable	Spaltenname
Type	Auswahlfeld für den Datentyp: <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>Date</code> , <code>POSIXct</code>
Include	Kontrollkästchen — diese Spalte als Prädiktor in das Modell aufnehmen
Factor	Kontrollkästchen — die Variable als kategorialen Faktor behandeln (erzeugt Dummy-Spalten in der Modellmatrix)
Force	Kontrollkästchen — in das Modell erzwingen (Straffaktor = 0, wird vom Lasso nie entfernt)
Special	Auswahlfeld (nur Appraisal/Market) — siehe die Referenz der Special-Spalentypen weiter unten
Sign	Erwartetes Koeffizientenvorzeichen: positiv, negativ oder beliebig
NAs	Anzahl fehlender Werte

6.3 Erkennung und Überschreiben von Datentypen

glmnetUI erkennt die Datentypen beim Import automatisch. Numerische, ganzzahlige und Datumsspalten werden erkannt. Zeichenspalten, die wie Datumsangaben aussehen (gängige Datumsformatmuster), werden als `Date` eingestuft.

Jede Erkennung lässt sich über das Auswahlfeld **Type** überschreiben. Eine Typänderung wirkt sich darauf aus, wie die Spalte in der Modellmatrix kodiert wird.

Text-, TRUE/FALSE- und R-Faktorspalten werden automatisch als kategorial behandelt. Eine numerische Spalte wird nur dann zum Faktor, wenn ihr Kontrollkästchen **Factor** gesetzt wird — glmnetUI leitet aus dem Wertebereich einer numerischen Spalte niemals einen kategorialen Charakter ab, sodass diskrete numerische Prädiktoren wie eine Badanzahl stetig bleiben, sofern nichts anderes angegeben wird. Die Factor-Einstellung wird über localStorage gesichert.

6.4 Erwartete Vorzeichen und deren Durchsetzung

Für jeden Prädiktor gibt das Auswahlfeld **Sign** an, ob sein Koeffizient positiv, negativ oder in beliebiger Richtung erwartet wird. Nach der Anpassung gilt:

- Koeffizienten mit unerwartetem Vorzeichen werden **in der Coefficients-Tabelle markiert**

- Vorzeichenwarnungen erscheinen in der Registerkarte Summary

Ist in den glmnet-Parametern (Kapitel 7) das Kontrollkästchen **Enforce Sign Constraints** aktiviert, wird das Modell gezwungen, Koeffizienten mit den erwarteten Vorzeichen zu erzeugen. Umgesetzt wird dies über `upper.limits` und `lower.limits` in glmnet.

6.5 Referenz der Special-Spaltentypen

In den Modi Appraisal und Market bietet das Auswahlfeld **Special** die folgenden Optionen:

Gewichtung:

- **weight** — Spalte mit Beobachtungsgewichten (in allen Modi verfügbar; nur eine zulässig; Zeilen mit Gewicht = 0 werden von der Anpassung ausgeschlossen)

Datums- und Zeittypen:

- **contract_date** — löst die automatische Berechnung von **sale_age** aus dem Stichtag aus
- **listing_date** — dient als Rückfalloption zur Berechnung der Vermarktungsdauer
- **dom** — kennzeichnet die Spalte mit der Vermarktungsdauer
- **sale_age** — kennzeichnet eine bereits vorhandene Verkaufsalter-Spalte, sodass eine beliebige Spalte als Verkaufsalter dienen kann, statt dass glmnetUI es aus **contract_date** berechnet

Monetäre Typen:

- **concessions** — kennzeichnet Verkaufszugeständnisse (Verkäufergutschriften, Käuferanreize usw.)

Transaktionstypen:

- **sale_type** — kennzeichnet die Spalte mit der Verkaufs- bzw. Transaktionsart

Größen- und Lagetypen:

- **latitude** — Werte werden automatisch auf drei Nachkommastellen gerundet
- **longitude** — gleiche Rundungsbehandlung wie beim Breitengrad
- **area** — Kennung des Marktgebiets oder der Lage (z. B. `area_id`, `neighborhood`, `market_area`)
- **living_area** — ermöglicht die Berechnung von Residuen je Quadratfuß (`residual_sf` und `cqa_sf`)
- **lot_size** — Spalte für die Grundstücksgröße
- **site_dimensions** — mit der Grundstücksgröße gruppiert (z. B. „75x120“)

Alterstypen:

- **actual_age** — Spalte mit dem Objektalter
- **effective_age** — effektives Objektalter

Anzeigetypen:

- **display_only** — die Spalte wird in Excel-Exporte aufgenommen, aber vollständig von der Modellanpassung ausgeschlossen. Geeignet für Adressfelder, MLS-Nummern, Flurstückskennungen oder sonstige Referenzdaten. Mehrere Spalten können diese Kennzeichnung tragen.

Hinweis: Als *display_only* gekennzeichnete Spalten verbleiben im Datensatz und erscheinen im Excel-Export, werden jedoch vollständig aus der Prädiktortabelle ausgeschlossen. Diese Kennzeichnung eignet sich für ID-Spalten, Adressen oder sonstige Referenzfelder.

7 Parameterauswahl

Abschnitt 5 der Seitenleiste — **glmnet Call Parameters** — eröffnet den Zugriff auf sämtliche Konfigurationsoptionen des Elastic-Net-Modells. Zu jedem Parameter gehört ein blaues Hilfesymbol (?) mit einer erläuternden Kurzinfo.

7.1 Alpha (Mischungsparameter)

Der Parameter alpha steuert die Art der Regularisierung:

Alpha-Wert	Typ	Verhalten
0	Ridge	Schrumpft alle Koeffizienten proportional; setzt keinen exakt auf null
1	Lasso	Kann Koeffizienten exakt auf null setzen (Variablenselektion)
0 bis 1	Elastic Net	Mischung aus Ridge und Lasso

Zwei Verfahren zur Wahl von alpha stehen zur Verfügung:

- **Fixed** — einen einzelnen Alpha-Wert über den Schieberegler festlegen (Standard: 1.0)
- **Grid Search** — mehrere Alpha-Werte prüfen und denjenigen mit dem geringsten CV-Fehler wählen. Bereich und Anzahl der Werte sind konfigurierbar.

7.2 Wahl von Lambda

Lambda steuert die Stärke der Regularisierung:

Verfahren	Beschreibung
Kreuzvalidierung	Empfohlen. Prüft viele Lambda-Werte mittels k -facher Kreuzvalidierung.
Manuell	Erlaubt die Eingabe eines bestimmten Lambda-Werts, sofern es dafür einen Grund gibt.

Bei Verwendung der Kreuzvalidierung stehen zwei Lambda-Varianten zur Wahl:

- **lambda.1se** (Standard) — das größte Lambda innerhalb einer Standardabweichung des Minimums. Erzeugt einfachere, stärker regularisierte Modelle. Für die Wertermittlung empfohlen.
- **lambda.min** — das Lambda, das den Kreuzvalidierungsfehler minimiert. Erzeugt das am besten angepasste, möglicherweise aber komplexere Modell.

Der Parameter **Number of CV Folds** steuert, wie die Daten für die Kreuzvalidierung aufgeteilt werden (Standard: 10 Folds).

7.3 Family

Wählen Sie die Verteilungsfamilie für Ihre Antwortvariable:

Family	Anwendungsfall
gaussian	Stetige Antworten (z. B. Verkaufspreis). Am häufigsten.

Family	Anwendungsfall
binomial	Binäre Ergebnisse (z. B. verkauft/nicht verkauft).
poisson	Zählraten (z. B. Anzahl der Verkäufe).

7.4 Standardize

Ist dieses Kontrollkästchen gesetzt (Standard), werden alle Prädiktoren vor der Anpassung auf Mittelwert 0 und Standardabweichung 1 skaliert. Damit wirkt der Strafterm unabhängig von der ursprünglichen Skala auf alle Prädiktoren gleich. Die Koeffizienten werden auf der ursprünglichen Skala zurückgegeben. Diese Option sollte in der Regel aktiviert bleiben.

7.5 Vorzeichenrestriktionen

Ist **Enforce Sign Constraints** aktiviert, werden die Koeffizienten so beschränkt, dass sie den in der Variablentabelle (Kapitel 6) festgelegten erwarteten Vorzeichen entsprechen:

- Ein Vorzeichen „positive“ erzwingt den Koeffizienten ≥ 0
- Ein Vorzeichen „negative“ erzwingt den Koeffizienten ≤ 0
- Auf „either“ gesetzte Variablen bleiben unbeschränkt

Umgesetzt wird dies über die Parameter `upper.limits` und `lower.limits` in `glmnet()`.

7.6 Relaxed Lasso

Das gewöhnliche Lasso verwendet denselben Strafterm für die Variablenselektion UND die Koeffizientenschätzung, was wichtige Koeffizienten zu stark in Richtung null schrumpfen kann. Das **Relaxed Lasso** trennt diese Schritte:

1. Zunächst werden die Variablen mittels Lasso ausgewählt (der Strafterm bestimmt, welche Koeffizienten überleben)
2. Anschließend werden die verbliebenen Variablen mit geringerem oder ohne Strafterm neu geschätzt, was weniger verzerrte Schätzungen liefert

Der Schieberegler **Gamma** steuert den Grad der Relaxation:

Gamma	Wirkung
0	Vollständig relaxiert (OLS-Neuschätzung, gar keine Schrumpfung)
1	Keine Relaxation (wie das gewöhnliche glmnet)
Dazwischen	Mischung aus relaxierter und penalisierter Anpassung

Bei Verwendung der Kreuzvalidierung wird Gamma automatisch so gewählt, dass sich eine optimale Leistung ergibt.

7.7 Zufallsstartwert

Ein Textfeld **Random seed**, vorgelegt mit einer zufälligen ganzen Zahl, macht die Zuordnung der Kreuzvalidierungs-Folds reproduzierbar — vor der Anpassung wird `set.seed()` aufgerufen, und der verwendete Startwert wird in der Registerkarte Glmnet Output sowie in der Statusmeldung zur Anpassung angezeigt. Nach jeder Anpassung wird ein neuer Startwert erzeugt (auf Basis der Systemzeit, nicht des Zufallszahlengenerators von R). Die letzten fünf für die aktuelle Datei verwendeten Startwerte werden als anklickbare Verweise angezeigt, sodass sie sich leicht wieder aufrufen lassen; die Startwert-Historie wird beim Laden einer neuen Datei zurückgesetzt.

7.8 Interaktionsmatrix

Eine interaktive obere Dreiecksmatrix erlaubt es zu steuern, welche Variablenpaare interagieren dürfen. Jede Zelle enthält ein Kontrollkästchen — ist es gesetzt, wird der Interaktionsterm $x_1:x_2$ in die Modellmatrix aufgenommen.

- **Allow All** — alle Interaktionspaare aktivieren
- **Clear All** — alle Paare deaktivieren (bei einer neuen Datei ist standardmäßig keines gesetzt)
- **Klick auf einen Variablennamen** — schaltet alle Interaktionen dieser Variablen um

Die Interaktionsauswahl wird je Eingabedatei im localStorage des Browsers gesichert und beim erneuten Laden derselben Datei wiederhergestellt.

Ist mindestens eine Interaktion aktiviert, enthält der exportierte Bericht eine Matrix **Enabled Interactions** (Prädiktoren auf beiden Achsen, ein Häkchen dort, wo der Produktterm $x_1:x_2$ aufgenommen wurde), im Querformat dargestellt und der Matrix Allowed Interactions von earthUI nachgebildet. Zu beachten ist, dass das Lasso den Koeffizienten eines aktivierten Terms dennoch auf null geschrumpft haben kann.

***Hinweis:** Bei vielen Prädiktoren setzt das Lasso nicht benötigte Interaktionsterme automatisch auf null. Mit „Allow All“ zu beginnen und die Auswahl dem Modell zu überlassen, ist ein vertretbares Vorgehen.*

Warnung

Werden glmnet-Interaktionen aktiviert, entstehen Anpassungen, die vor Gericht oder in Prüfungssituationen schwer zu interpretieren und zu erläutern sein können. Interaktionen sind mit Bedacht einzusetzen, und die resultierenden Anpassungen sind auf Plausibilität zu prüfen.

7.9 Vom Haupteffekt ausschließen (nur Interaktion)

Ein **Rechtsklick** auf einen Variablennamen in der Interaktionsmatrix schließt die Variable als Haupteffekt aus dem Modell aus. Nach dem Variablennamen erscheint eine fette „1“ als Anzeige. Ist eine Variable ausgeschlossen, gilt:

- Die Variable wird aus den Haupttermen der Formel **entfernt**
- Die Variable **kann weiterhin** in Interaktionstermen ($x_1:x_2$) auftreten
- Ein erneuter Rechtsklick hebt den Ausschluss auf
- Der Zustand wird je Eingabedatei über localStorage gesichert

Ein typischer Anwendungsfall: Sollen Zeitanpassungen mit der Objektgröße skalieren, wird **sale_age** vom Haupteffekt ausgeschlossen und nur in einer Interaktion mit **living_area** zugelassen. So entsteht ein Interaktionsterm **sale_age:living_area** anstelle eines konstanten **sale_age**-Koeffizienten — die zeitliche Wertsteigerungsrate variiert dann mit der Hausgröße.

7.10 Erweiterte Parameter

Ein ausklappbarer Abschnitt „Advanced“ am Ende der glmnet Call Parameters macht weitere Einstellungen zugänglich. Sie verwenden sinnvolle Standardwerte, sind aber sichtbar, damit sämtliche Modelleinstellungen für Gericht oder Prüfung dokumentiert werden können.

- **Lambda Min Ratio** — Verhältnis des kleinsten zum größten Lambda im Regularisierungspfad. Standard: 0.00001. Bei kleinen Datensätzen (<50 Beobachtungen) kann ein größerer Wert (etwa 0.01) die Leistung verbessern.

- **Number of Lambda Values** — Anzahl der Lambda-Werte im Pfad. Standard: 100.
- **CV Loss Metric** — Verlustfunktion für die Kreuzvalidierung: MSE (Standard für gaussian), MAE (robust gegenüber Ausreißern) oder Devianz.
- **Convergence Threshold** — Konvergenzschwelle des Koordinatenabstiegs. Standard: $1e-07$.
- **Max Iterations** — maximale Anzahl der Koordinatenabstiegs-Iterationen. Standard: 100.000. Bei Konvergenzwarnungen erhöhen.
- **Fit Intercept** — einen Achsenabschnitt (konstanten Term) aufnehmen. Standard: aktiviert. Der Achsenabschnitt ist der „Basiswert“ im RCA der Wertermittlung.

7.11 Standardwerte der Einstellungen

Drei Optionen steuern, wie die Parameter beim Laden einer Datei initialisiert werden:

- **Use last settings for input file** — genau die zuletzt mit dieser Datei verwendeten Einstellungen wiederherstellen
- **Use default settings** — die gespeicherten globalen Standardwerte anwenden
- **glmnet defaults** — auf die Werkseinstellungen zurücksetzen ($\alpha = 1$, CV mit lambda.1se, 10 Folds, gaussian, standardize aktiviert)

Ein Klick auf **Save current as default** speichert sämtliche aktuellen Einstellungen als globalen Standard für künftige Dateien.

8 Anpassung des Modells

8.1 Die Schaltfläche zur Modellanpassung

Abschnitt 6 der Seitenleiste enthält die Schaltfläche **Fit Glmnet Model**. Ein Klick darauf rechnet das Modell mit der aktuellen Konfiguration.

8.2 Was bei der Anpassung geschieht

Beim Klick auf Fit führt glmnetUI Folgendes aus:

1. **Daten vorbereiten** — Zeilen mit NA-Werten entfernen, Gewichte anwenden (sofern gekennzeichnet), die Bewertungsobjektzeile ausschließen (im Wertermittlungsmodus) und Faktorvariablen als Dummy-Spalten kodieren.
2. **Modellmatrix aufbauen** — die Designmatrix einschließlich aller in der Interaktionsmatrix zugelassenen Interaktionsterme erzeugen.
3. **Restriktionen anwenden** — ist die Vorzeichendurchsetzung aktiviert, werden `upper.limits` und `lower.limits` für die Koeffizienten gesetzt.
4. **Kreuzvalidierung durchführen** — ist CV gewählt, wird `cv.glmnet()` aufgerufen, um das optimale Lambda zu bestimmen. Ist das Relaxed Lasso aktiviert, erfolgt der Aufruf als `cv.glmnet(..., relax = TRUE)`.
5. **Endgültiges Modell anpassen** — das angepasste Modell, die Koeffizienten, die Prognosen und die Residuen werden gespeichert.
6. **Alle Ergebnis-Registerkarten aktualisieren** — Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA und Diagnostics werden befüllt.

Eine Statusmeldung unterhalb der Fit-Schaltfläche zeigt die Anzahl der Beobachtungen, die ausgeschlossenen Zeilen und das gewählte Lambda an.

9 Ergebnis-Registerkarten

9.1 Data Preview

Zeigt die importierten Daten als interaktive DataTable. In den Modi Appraisal und Market ist die Vorschau in zwei Tabellen aufgeteilt: **Subject Property** (Zeile 1) und **Comparable Sales** (Zeilen 2 ff.).

Jede Zelle wird auf eine Zeile begrenzt und auf die Spaltenbreite gekürzt, sodass breite Freitextfelder (etwa Objektbemerkungen) eine Tabellenzeile nicht über die Seite ziehen. Ein **Doppelklick auf eine beliebige Zelle** öffnet ein Fenster mit ihrem vollständigen, ungekürzten Inhalt.

9.2 Equation

Zeigt die angepasste Modellgleichung, über MathJax in LaTeX gerendert. Dargestellt werden alle von null verschiedenen Koeffizienten in korrekter mathematischer Formatierung:

- Positive Koeffizienten mit $+$ -Zeichen
- Negative Koeffizienten mit $-$ -Zeichen
- Interaktionsterme als $x_1 \times x_2$
- Zahlen mit Tausendertrennzeichen und angemessener Anzahl von Nachkommastellen

9.3 Correlation

Eine Heatmap-Matrix der Pearson-Korrelationen zwischen allen numerischen Prädiktoren und der Antwortvariablen. Unmittelbar nach dem Datenimport verfügbar — eine Modellanpassung ist nicht erforderlich.

- Farbverlauf: blau (-1) $\rightarrow$ weiß (0) $\rightarrow$ rot ($+1$)
- Korrelationswerte in jeder Zelle ausgewiesen
- Schrift- und Achsengröße passen sich der Zahl der Variablen an
- Nützlich, um Multikollinearität vor der Anpassung zu erkennen

9.4 Summary

Kennzahlen zur Anpassungsgüte, als Karten dargestellt:

- **R^2** — Anteil der erklärten Varianz
- **Adj R^2** — um die Zahl der Prädiktoren bereinigt
- **GR^2** — generalisiertes R^2 aus dem Devianzverhältnis von glmnet beim gewählten Lambda
- **CV R^2** — R^2 aus der Kreuzvalidierung (wenn CV verwendet wird)
- **RMSE** — Wurzel des mittleren quadratischen Fehlers
- **MAE** — mittlerer absoluter Fehler

In den Modi Appraisal und Market zeigen zusätzliche Karten:

- **Median Ratio** — Median von (prognostiziert / beobachtet)
- **COD** — Coefficient of Dispersion (Streuungskoeffizient; niedriger ist besser)
- **PRD** — Price-Related Differential

Eine **Overfitting-Warnung** erscheint, wenn das Trainings- R^2 das CV- R^2 um mehr als 0,1 übersteigt.

Unterhalb der Karten dupliziert eine Koeffiziententabelle mit Vorzeichenwarnungen der Bequemlichkeit halber die Registerkarte Coefficients.

9.5 Coefficients

Eine Tabelle aller von null verschiedenen Koeffizienten mit:

- Variablenname
- Koeffizientenwert
- Erwartetes Vorzeichen (aus der Variablenkonfiguration)
- Vorzeichenwarnung, wenn die Richtung des Koeffizienten nicht der Erwartung entspricht

9.6 Variable Importance

Standardisierte Koeffizientengröße: $|\beta| \times \text{sd}(x)$ je Prädiktor, bei Faktorvariablen über die Dummy-Spalten aggregiert. Zwei Darstellungsformen:

- **Interaktives plotly** (wenn das Paket `plotly` installiert ist): waagerechtes Balkendiagramm mit Tooltips, die die genaue Wichtigkeit, den Koeffizienten und den relativen Anteil zeigen.
- **Statisches ggplot2** (Rückfalloption): waagerechtes Balkendiagramm mit Achsenbeschriftungen samt Tausendertrennzeichen.

Unterhalb der Grafik zeigt eine DataTable Variable, Importance, Coefficient und Relative %.

9.7 Contributions

Grafiken der partiellen Effekte je Variable, die den Beitrag jedes Prädiktors zu den Modellprognosen zeigen. Die Variable wird über ein Auswahlfeld gewählt.

Für numerische Prädiktoren: ein Streudiagramm der x-Werte der Variablen gegen ihren Beitrag ($\beta \times x$), mit einem roten Liniensegment und einer Steigungsbeschriftung. Die Steigungsbeschriftung gibt den marginalen Effekt je Einheit an (etwa +1,234.56/unit), mit adaptiven Einheiten für Variablen mit geringer Spannweite wie geografische Breite und Länge (etwa +0.12/0.001). Der Untertitel zeigt den Wert des Achsenabschnitts (Basiswert).

Für Faktorvariablen: ein Boxplot der Verteilung der Beitragswerte je Faktorstufe.

Für Interaktionen zweier Variablen: drei übereinander angeordnete Darstellungen — eine 3D-Flächengrafik des gemeinsamen Beitrags (interaktiv mit `plotly`, sofern verfügbar, sonst statisch über `persp()`), ein Streudiagramm mit nach Beitrag eingefärbten Punkten sowie eine Heatmap des mittleren Beitrags über beide Wertebereiche.

9.8 ANOVA

Tabelle der Varianzzerlegung, die für jeden Prädiktor Folgendes zeigt:

- **SS** — Quadratsumme: $\sum(\text{contribution}_j - \overline{\text{contribution}}_j)^2$
- **% of Model SS** — Anteil an der gesamten Quadratsumme des Modells
- **Coefficient** — der dominierende Koeffizient dieses Prädiktors

Enthält eine Zeile für den Achsenabschnitt sowie eine Zeile TOTAL MODEL. Interaktionsterme (erkennbar an „:“) werden als eigene Gruppen ausgewiesen.

9.9 Diagnostics

Vier Diagnosegrafiken mit großen Schriften (Basisgröße 16 pt), 15 Achsenteilstrichen und Beschriftungen mit Tausendertrennzeichen (keine wissenschaftliche Notation):

1. **Koeffizientenpfad** — zeigt, wie sich alle Koeffizienten mit wachsendem $\log(\lambda)$ verändern. Die Linien sind nach Variablen eingefärbt. Hilft dabei, den Regularisierungspfad und die Variablenselektion zu veranschaulichen.

2. **CV-Fehler** — Kreuzvalidierungsfehler als Funktion von $\log(\lambda)$. Die Punkte zeigen den mittleren CV-Fehler, die Fehlerbalken ± 1 Standardfehler. Gestrichelte senkrechte Linien markieren $\lambda_{\min}$ (blau) und λ_{1se} (grün).
3. **Beobachtet gegen prognostiziert** — Streudiagramm der beobachteten gegen die prognostizierten Werte. Die Punkte sollten sich um die gestrichelte 45-Grad-Referenzlinie gruppieren. Eine breitere Streuung deutet auf einen größeren Prognosefehler hin.
4. **Residuen gegen angepasste Werte** — Streudiagramm der Residuen gegen die angepassten Werte. Es sollte eine zufällige Streuung um die (gestrichelte) Nulllinie zeigen. Muster deuten auf eine Fehlspezifikation des Modells hin (etwa Nichtlinearität oder Heteroskedastizität).

Jede Grafik verfügt über eine Schaltfläche **Download PNG** zum Speichern mit 150 DPI.

9.10 Glmnet Output

Rohausgabe des Modells: der verwendete Zufallsstartwert, der Modellausdruck, das gewählte Lambda, $\lambda_{\min}/\lambda_{1se}$, Gamma (wenn das Relaxed Lasso aktiviert ist) sowie der vollständige Koeffizientenvektor.

9.11 Report

Berichtsfelder konfigurieren (Name des Sachverständigen, Objektadresse, Berichtsdatum, Aktennummer) und unmittelbar aus der Registerkarte heraus nach Word (.docx) oder PDF exportieren. Der vollständige, auf Quarto gestützte Berichtsablauf (HTML, Word oder PDF) ist in Kapitel 13 beschrieben.

10 Herunterladen von Daten

Nach der Modellanpassung exportiert die Schaltfläche **Download Output (Excel)** (Seitenleistenabschnitt 7, in allen Purpose-Modi verfügbar) eine Excel-Datei mit Prognosen und Diagnosegrößen in den Ausgabeordner des Projekts. Die Abschnittsüberschrift lautet in den Modi Appraisal und Market „Download Estimated Sale Prices & Residuals“ und im Modus General „Download Estimated Target Variable(s) & Residuals“.

10.1 Ausgabespalten

Spalte	Beschreibung
<code>est_<target></code>	Modellprognose (z. B. <code>est_sale_price</code>)
<code>residual</code>	Beobachtet – prognostiziert
<code>cqa</code>	Condition-Quality-Appeal-Score (Skala 0–10)
<code>residual_sf</code>	Residuum / Wohnfläche (wenn <code>living_area</code> gekennzeichnet ist)
<code>cqa_sf</code>	CQA, berechnet aus der Rangfolge über <code>residual_sf</code>
<code><var>_contribution</code>	Beitrag der jeweiligen Variablen zur Prognose
<code>basis</code>	Wert des Achsenabschnitts (für alle Zeilen gleich)

Nach erfolgreichem Abschluss erscheint auf der Download-Schaltfläche ein weißes Häkchen.

10.2 CQA-Scores

Der CQA-Wert ordnet das Residuum jeder Zeile im Vergleich zu allen übrigen auf einer Skala von 0 bis 10 ein:

- **Hoher CQA (~9–10):** deutlich über der Prognose verkauft — vermutlich überdurchschnittlicher Zustand bzw. überdurchschnittliche Qualität oder Attraktivität
- **Niedriger CQA (~0–1):** deutlich unter der Prognose verkauft — vermutlich unterdurchschnittlicher Zustand oder ein Notverkauf
- **CQA ~5:** nahe dem Median

In den Modi Appraisal und Market werden die Zeilen absteigend nach `residual_sf` sortiert, sofern eine Spalte `living_area` gekennzeichnet ist.

11 RCA-Berechnungen und Herunterladen

Der RCA-Ablauf (Residual Constraint Approach) steht in den Modi **Appraisal** und **Market Area Analysis** nach der Modellanpassung zur Verfügung. Das darauf aufbauende Sales Comparison Grid (Kapitel 12) ist dem Wertermittlungsmodus vorbehalten.

11.1 Öffnen des RCA-Dialogs

Klicken Sie auf die Schaltfläche **Calculate RCA Adjustments & Download** in Seitenleistenabschnitt 8. Zu wählen sind:

- Score-Typ **CQA** oder **CQA per SF**. Bei „CQA per SF“ wird das Residuum des Bewertungsobjekts mit der Wohnfläche skaliert.
- Eingabe des CQA-Werts für das Bewertungsobjekt (0.00–9.99, Standard 5.00)

11.2 Interpolation des CQA-Werts

1. Die CQA-Werte und Residuen der Vergleichsobjekte werden sortiert
2. Eine lineare Interpolation bildet den gewählten CQA-Wert auf ein Residuum ab
3. Wert des Bewertungsobjekts = Modellprognose + interpoliertes Residuum

11.3 Ausgabespalten

Spalte	Beschreibung
subject_value	Modellprognose + interpoliertes Residuum
<var>_adjustment	Beitrag des Bewertungsobjekts – Beitrag des Vergleichsobjekts
residual_adjustment	Residuum des Bewertungsobjekts – Residuum des Vergleichsobjekts
net_adjustments	Summe aller Anpassungen
gross_adjustments	Summe der absoluten Anpassungen
adjusted_sale_price	Verkaufspreis des Vergleichsobjekts + Nettoanpassungen

Nach erfolgreicher Berechnung erscheint auf der Schaltfläche ein weißes Häkchen.

12 Sales Comparison Grid

Das Sales Comparison Grid (Vergleichsraster) steht ausschließlich im Modus **Appraisal** zur Verfügung, nachdem die RCA-Anpassungen (Schritt 8) berechnet wurden. Es erzeugt eine formatierte Excel-Arbeitsmappe, die sich zur Aufnahme in Wertermittlungsberichte eignet.

12.1 Dialog zur Auswahl der Vergleichsobjekte

Ein Klick auf die Schaltfläche **Generate Sales Grid & Download** in Seitenleistenabschnitt 9 öffnet einen modalen Dialog, der alle Vergleichsverkäufe aus der RCA-Ausgabe auflistet. Die Vergleichsobjekte sind in zwei Gruppen aufgeteilt:

- **Recommended comps** (empfohlene Vergleichsobjekte, vorausgewählt): Bruttoanpassung < 25 % des Verkaufspreises, sortiert nach Verkaufsalter (jüngste zuerst).
- **Other comps** (übrige Vergleichsobjekte, standardmäßig nicht ausgewählt): die verbleibenden Vergleichsobjekte, sortiert nach Verkaufsalter.

Bis zu 30 Vergleichsobjekte auswählen und anschließend auf **Generate Sales Grid** klicken, um die Arbeitsmappe zu erzeugen.

12.2 Aufbau der Arbeitsmappe

Die erzeugte Arbeitsmappe enthält bis zu 10 Blätter mit jeweils drei nebeneinander angeordneten Vergleichsobjekten:

- **Kopfzeilen:** Adressen von Bewertungsobjekt und Vergleichsobjekten, Verkaufspreise, Verkaufsdaten und Vermarktungsdauer
- **Gruppierte Zeilen:** Lage (Breitengrad, Längengrad, Gebiet), Grundstücksgröße (Grundstücksfläche, Grundstücksmaße) und Alter (tatsächliches Alter, effektives Alter) mit Gruppenüberschriften und Unterpunkten
- **Zeilen der Modellvariablen:** eine Zeile je von null verschiedenem Modellkoeffizienten, mit dem Beitrag des Bewertungsobjekts, den Beiträgen der Vergleichsobjekte und den prozentualen Anpassungen
- **Zeilen für Restmerkmale:** leere, entspernte Zellen für Eingaben des Sachverständigen zu Merkmalen, die das Modell nicht erfasst (etwa Zustand, Qualität, Aussicht)
- **Zusammenfassungszeilen:** Nettoanpassung in %, Bruttoanpassung in % und angepasster Verkaufspreis mit aktiven Excel-Formeln

12.3 Blattschutz und Restmerkmalzellen

Jedes Blatt ist mit einem Kennwort gegen versehentliche Änderungen geschützt, die Wertzellen für die Restmerkmale sind jedoch ausdrücklich entspernt. So können Sachverständige Werte für Merkmale eintragen, die nicht im Modell enthalten sind, während die formelgestützten Anpassungsberechnungen erhalten bleiben.

13 Herunterladen von Berichten

13.1 Der zweistufige Quarto-Ablauf

Der Berichtsexport aus der Seitenleiste ist ein zweistufiger Quarto-Ablauf:

1. **Generate Quarto Report** (Seitenleistenabschnitt 10) — schreibt ein in sich geschlossenes Quarto-Paket (die `.qmd`-Quelle nebst Grafiken, `report_data.rds` und `reference.docx`) in den glmnet-Ausgabeordner des Projekts.
2. **Convert Quarto Report** (Seitenleistenabschnitt 11) — rendert eine `.qmd`-Datei in die gewählten Ausgabeformate. Der Pfad ist mit dem zuletzt erzeugten Paket vorbelegt, es lässt sich jedoch zu jeder beliebigen `.qmd`-Datei navigieren.

13.2 Berichtsformate

Im Konvertierungsschritt stehen drei Formate zur Verfügung:

- **HTML** — in sich geschlossenes HTML-Dokument. Funktioniert auf allen Plattformen ohne zusätzliche Software außer pandoc. Empfohlener Standard.
- **Word (.docx)** — formatiertes Dokument. Geeignet zur Bearbeitung und Weitergabe. Funktioniert auf allen Plattformen.
- **PDF** — über Quarto mit LaTeX erzeugt. Das Papierformat richtet sich nach der Locale-Einstellung (Letter oder A4). **Erfordert eine LaTeX-Installation** (siehe den Anhang zu den Systemanforderungen). Wird kein LaTeX erkannt, wird die PDF-Option automatisch aus den Format-Kontrollkästchen ausgeblendet.

Berichte werden nach Möglichkeit mit Quarto erzeugt, ersatzweise mit rmarkdown. Die Registerkarte Report bietet darüber hinaus Schaltflächen für den direkten Export nach Word und PDF.

13.3 Berichtsinhalte

Die Berichte enthalten sämtliche Inhalte der Registerkarten:

- Beschreibung des Datensatzes (Dateiname, Beobachtungen, Antwortvariable, Prädiktoren, Purpose)
- Modellspezifikation (alpha, lambda, family, standardize, relaxed, von null verschiedene Koeffizienten)
- Ergebnisübersicht (R^2 , adjustiertes R^2 , generalisiertes R^2 , CV- R^2 , RMSE, MAE)
- Modellgleichung (in LaTeX gerendert)
- Koeffiziententabelle
- Variablenwichtigkeit (Balkendiagramm und Tabelle)
- Grafiken der Beiträge je Variable, entsprechend der Registerkarte Contributions:
 - **Numerische Prädiktoren:** Streudiagramm mit linearer Anpassungsgeraden
 - **Faktorvariablen:** Boxplot je Stufe
 - **Interaktionen:** nach Beitrag eingefärbtes Streudiagramm, Heatmap des mittleren Beitrags und statische Momentaufnahme der 3D-Fläche (persp)
- Matrix Enabled Interactions (wenn mindestens eine Interaktion aktiviert ist) — Prädiktoren auf beiden Achsen, mit einem Häkchen dort, wo der Produktterm `x1:x2` aufgenommen wurde, im Querformat dargestellt
- Heatmap der Korrelationsmatrix
- Tabelle der ANOVA-Zerlegung
- Diagnosegrafiken:

- Koeffizientenpfad
 - Kreuzvalidierungsfehler
 - Beobachtet gegen prognostiziert
 - Residuen gegen angepasste Werte
 - Normal-Q-Q-Diagramm
- Modellausgabe

Sämtliche Achsenbeschriftungen und Farblegenden verwenden Zahlen mit Tausendertrennzeichen (keine wissenschaftliche Notation).

Nach erfolgreicher Erzeugung erscheint auf der Berichtsschaltfläche ein weißes Häkchen; während der Bericht gerendert wird, läuft eine Zeitanzeige mit.

14 Vergleich mit earthUI

Die earth-Anwendung und ihr glmnet Post Processing Module bilden ein gemeinsames Werkzeugpaket für die Regressionsmodellierung. Sie teilen Datenformat, Special-Spaltentypen, RCA-Ablauf und Demodatensätze, verwenden jedoch unterschiedliche Modellierungsmaschinen.

14.1 Wesentliche Unterschiede

Merkmal	glmnetUI	earthUI
Verfahren	Elastic Net (glmnet)	MARS/Earth (earth)
Variablenselektion	Automatisch über den Lasso-Strafterm	Automatisch über Vorwärts-/Rückwärts-Pruning
Koeffizientenbeziehungen	Nur lineare Effekte	Stückweise linear (Hinge-Funktionen)
Interaktionen	Paarweise über die Interaktionsmatrix	Bis Grad 3
Koeffizientenvorzeichen	Über Restriktionen durchsetzbar	Nicht unmittelbar durchsetzbar
Relaxierte Anpassung	Relaxed Lasso verfügbar	Nicht anwendbar
Alpha-Optimierung	Gittersuche verfügbar	Nicht anwendbar (kein Mischungsparameter)
Sales Comparison Grid	Verfügbar (Schritt 9)	Verfügbar (Schritt 8)
g-Funktionsgrafiken	Nicht anwendbar	Darstellung gruppierteter Terme
Berichtsformate	HTML, Word, PDF	HTML, Word, PDF

14.2 Gemeinsame Funktionen

Beide Werkzeuge bieten:

- CSV- und Excel-Import mit Umwandlung der Spaltennamen in snake_case
- Locale-Unterstützung für 31 Länder mit CSV-Trennzeichen, Dezimaltrennzeichen und Papierformat
- Spezielle Spaltenkennzeichnungen (contract_date, living_area, area, latitude, longitude usw.)
- Automatische Berechnung von sale_age aus dem Stichtag
- Umschalter zwischen dunklem und hellem Design
- Persistenz der Einstellungen über localStorage
- Korrelations-Heatmap, Variablenwichtigkeit, Beiträge, ANOVA, Diagnostik
- CQA-Bewertung und RCA-Anpassungsablauf
- Sales Comparison Grid mit Auswahl der Vergleichsobjekte und formatierter Excel-Ausgabe
- Kompatibilität mit denselben Demodatensätzen (Appraisal_1.csv)

14.3 Wann welches Werkzeug

- **glmnetUI verwenden**, wenn ein einfaches, interpretierbares lineares Modell mit automatischer Variablenselektion gewünscht ist. Das Elastic Net eignet sich besonders, wenn viele Prädiktoren vorliegen und das Modell die wichtigsten auswählen soll.
- **earthUI verwenden**, wenn nichtlineare Zusammenhänge zu erwarten sind (etwa wenn sich die Wirkung der Wohnfläche auf den Preis je nach Größe ändert) oder Hinge-Funktionen benötigt werden, um Schwelleneffekte abzubilden. Earth-Modelle können

zudem Interaktionen höherer Ordnung (Grad 2 und 3) erfassen. earthUI bietet darüber hinaus eine experimentelle, auf Prolog/DCG gestützte Funktion zur Extraktion von Bemerkungen, für die es in glmnetUI keine Entsprechung gibt.

- **Beide verwenden**, um lineare und nichtlineare Modelle auf denselben Daten zu vergleichen. Ergeben beide ähnliche R^2 -Werte, ist das einfachere glmnet-Modell aus Gründen der Interpretierbarkeit möglicherweise vorzuziehen.

15 Demodatensatz: Appraisal_1.csv

15.1 Beschreibung

Der Demodatensatz wird mit glmnetUI ausgeliefert (und mit earthUI geteilt). Er lässt sich wie folgt auffinden:

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

Die Datei enthält 1.502 Wohnimmobilienverkäufe (zuzüglich eines Bewertungsobjekts in Zeile 1) aus einem simulierten MLS-Export. Die Daten bilden Verkäufe von Einfamilienhäusern in einem mehrgebietigen Markt mit einer Bandbreite von Objektgrößen, Baualtern und Lagen ab.

Es handelt sich nicht um reale Daten, sie beruhen jedoch auf einer realistischen Wohnlage in Nordkalifornien. Sämtliche identifizierenden Angaben wurden verändert oder entfernt.

15.2 Spalten

Spalte	Typ	Special-Typ	Beschreibung
weight	numeric	weight	Beobachtungsgewicht (0 = von der Anpassung ausschließen)
id	numeric	display_only	Interne Datensatzkennung
property_id	numeric	display_only	MLS-Objektkennung
listing_id	character	display_only	MLS-Angebotsnummer
parcel_number	character	display_only	Flurstücksnummer der Katasterbehörde (APN)
street_address	character	display_only	Objektadresse
city_name	character	display_only	Stadt
postal_code	character	display_only	Postleitzahl
county_name	character	display_only	County
contract_date	Date	contract_date	Vertragsdatum des Verkaufs (berechnet sale_age)
sale_age	numeric	—	Tage vom Vertragsdatum bis zum Stichtag
sale_price	numeric	(Ziel)	Verkaufspreis — Antwortvariable
living_sqft	numeric	living_area	Bruttowohnfläche in Quadratfuß
beds_total	integer	—	Anzahl der Schlafzimmer
baths_total	numeric	—	Gesamtzahl der Bäder (z. B. 2,5 = 2 Vollbäder + 1 Gästebad)
lot_size	numeric	lot_size	Grundstücksfläche in Quadratfuß
area_id	integer	area	MLS-Gebietskennung
age	numeric	actual_age	Objektalter in Jahren
latitude	numeric	latitude	Breitengrad (für das Modell auf 3 Nachkommastellen gerundet)
longitude	numeric	longitude	Längengrad (für das Modell auf 3 Nachkommastellen gerundet)
garage_spaces	integer	—	Anzahl der Garagenstellplätze
days_on_market	integer	dom	Vermarktungsdauer in Tagen
listing_date	Date	listing_date	Datum des Angebots
sale_concessions	numeric	concessions	Verkäuferzugeständnisse

15.3 Vorgeschlagener Schnelleinstieg

1. Die glmnet-Routine starten: `earthUI::launch_glmnet()`
2. Ein neues Projekt (Abschnitt 1) mit dem Purpose **For Appraisal** anlegen
3. `Appraisal_1.csv` in den Eingabeordner des Projekts legen und in Abschnitt 2 auswählen (bei Bedarf auf **Refresh** klicken)
4. `sale_price` als Zielvariable wählen
5. Die Special-Typen gemäß der obigen Tabelle zuweisen
6. Folgende Prädiktoren aufnehmen: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (als Faktor), `age`, `latitude`, `longitude`, `garage_spaces`
7. $\alpha = 1$ (Lasso) beibehalten, CV mit `lambda.1se`
8. Auf **Fit Glmnet Model** klicken
9. Die Registerkarten Summary, Coefficients und Diagnostics prüfen
10. Die Ausgabe herunterladen (Schritt 7) und die CQA-Rangfolge prüfen
11. RCA-Anpassungen (Schritt 8) mit einem CQA-Wert von ~5.00 berechnen
12. Ein Sales Comparison Grid (Schritt 9) mit den empfohlenen Vergleichsobjekten erzeugen

16 Systemanforderungen und Fehlerbehebung

16.1 Unterstützte Plattformen

glmnetUI läuft unter **macOS**, **Windows** und **Linux**. RStudio Desktop (2023.06+) wird dringend empfohlen — es bringt **pandoc** mit, das für die Berichte benötigt wird.

16.2 Optionale Abhängigkeiten

- **LaTeX** (TinyTeX, MiKTeX oder MacTeX): nur für die Erzeugung von PDF-Berichten. Wird keine Installation erkannt, wird die PDF-Option automatisch ausgeblendet. Installation mit: `tinytex::install_tinytex()`
- **sysfonts + showtext**: Schriftart Roboto Condensed für Grafiken. Ist sie nicht verfügbar oder besteht keine Internetverbindung, wird automatisch auf die serifenlose Systemschrift zurückgegriffen.
- **earth** ($\geq 5.3.0$): nur für die Importstrecke aus earthUI erforderlich.

16.3 Plattformhinweise

macOS: die wenigsten Probleme. Für PDF TinyTeX installieren: `tinytex::install_tinytex()`

Windows: funktioniert gut mit RStudio. Auf Firmenrechnern mit eingeschränkten Rechten kann es Beschränkungen für das temporäre Verzeichnis geben — die Anwendung gibt dann eine Warnung in der Konsole aus, läuft jedoch ohne Persistenz der Einstellungen weiter.

Linux: möglicherweise sind Systembibliotheken erforderlich: `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

16.4 Verhalten bei fehlenden Komponenten

Fehlende Komponente	Verhalten
Kein LaTeX	PDF-Option ausgeblendet. HTML und Word bleiben verfügbar.
Kein Internet / Schriften nicht ladbar	Serifenlose Systemschrift wird verwendet. Konsolenmeldung wird protokolliert.
Kein RSQLite / schreibgeschütztes Dateisystem	Einstellungen bleiben nicht erhalten. Die Anwendung läuft normal.
Temporäres Verzeichnis nicht beschreibbar	Die Berichtserzeugung schlägt mit einer klaren Fehlermeldung fehl. <code>TMPDIR</code> auf einen beschreibbaren Ort setzen.

16.5 Fehlerbehebung

„PDF option not available“: In R `tinytex::install_tinytex()` ausführen und die Anwendung neu starten.

„Settings will not persist“: Die Berechtigungen für das Benutzerdatenverzeichnis prüfen (macOS: `~/Library/Application Support/R/glmnetUI/`, Linux: `~/.local/share/R/glmnetUI/`, Windows: `%APPDATA%/R/data/glmnetUI/`).

Port 7879 bereits belegt: `lsof -ti:7879 | xargs kill` (macOS/Linux) oder `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell) ausführen.

Literatur

- Friedman, Jerome, Trevor Hastie und Robert Tibshirani (2010). „Regularization Paths for Generalized Linear Models via Coordinate Descent“. In: *Journal of Statistical Software* 33.1, S. 1–22. DOI: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01).
- Simon, Noah u. a. (2011). „Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent“. In: *Journal of Statistical Software* 39.5, S. 1–13. DOI: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).
- Tay, J. Kenneth, Balasubramanian Narasimhan und Trevor Hastie (2023). „Elastic Net Regularization Paths for All Generalized Linear Models“. In: *Journal of Statistical Software* 106.1, S. 1–31. DOI: [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).

mgcvUI: Eine interaktive Benutzeroberfläche für das mgcv-Paket (generalisierte additive Modelle)

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische

Fassung ist maßgeblich.

William Bert Craytor

Jahrgang 1, Ausgabe 1 · Sommer 2026

Veröffentlicht am 1. Juli 2026

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

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

Dieser Open-Access-Artikel wird unter den Bedingungen der [Creative-Commons-Namensnennung-4.0-Lizenz](#) verbreitet.

Zusammenfassung. mgcvUI ist eine grafische Benutzeroberfläche für das R-Paket mgcv, das generalisierte additive Modelle (GAMs) mit penalisierten Regressionssplines und automatischer Glättungsparameterwahl anpasst. Es bietet drei Verwendungsmodi — allgemeine prädiktive Modellierung, Immobilienbewertung und Marktgebietsanalyse — und führt den Anwender durch Datenimport, Spezifikation der Glättungsterme, Modellanpassung, Diagnose- und Effektdiagramme sowie herunterladbare Berichte. Dieser Artikel dokumentiert die Datenformat-Anforderungen, den Modellierungs-Workflow, die Ausgabedarstellungen und die vollständige Funktionsreferenz von mgcvUI.

Schlüsselwörter: generalisierte additive Modelle, GAM, penalisierte Splines, Glättung, mgcv, R, grafische Benutzeroberfläche, Immobilienbewertung

Anmerkung des Verlags (Juli 2026): Kurz vor Drucklegung wurde mgcvUI in earthUI (Version 0.11.0 und höher) zusammengeführt; earthUI stellt den hier beschriebenen GAM-Workflow nun neben den MARS- und Elastic-Net-Workflows bereit. mgcvUI bleibt in seiner letzten veröffentlichten Fassung (0.3.0) unter <https://github.com/wcraytor/mgcvUI> verfügbar; die Weiterentwicklung erfolgt in earthUI. Dieser Artikel bezieht sich auf mgcvUI in der veröffentlichten Form.

Inhaltsverzeichnis

1	Einleitung	4
1.1	Geschichte und Hintergrund	4
1.1.1	Generalisierte additive Modelle	4
1.1.2	Das Paket mgcv	4
1.1.3	Vergleich mit MARS (earth) und Elastic Net (glmnet)	5
1.1.4	Die earthUI-mgcvUI-Pipeline	6
1.1.5	Empfohlener Arbeitsablauf	6
1.2	Was ist mgcvUI?	6
1.3	Die drei Purpose-Modi	7
1.4	Immobilien-spezifische Funktionen	7
1.5	Erste Schritte	8
2	Anforderungen an MLS-Eingabedaten	9
2.1	Dateiformat & Struktur	9
2.2	Erforderliche Spalten im Wertermittlungsmodus	9
2.3	Konventionen für besondere Spaltenbezeichnungen	10

2.4	Datenqualität & Vollständigkeit	10
2.5	Platzierung der Bewertungsobjektzeile	10
3	Der Modus General Purpose	11
3.1	Überblick	11
3.2	Der Arbeitsablauf in der Seitenleiste	11
3.3	Import aus earthUI (optional)	12
3.4	Registerkarten des Hauptbereichs	12
3.5	Dauerhafte Speicherung der Einstellungen	13
3.6	Dark Mode	13
3.7	Locale & Regional Settings (Gebietsschema und regionale Einstellungen)	13
4	Der Appraisal-Modus	15
4.1	Behandlung der Bewertungsobjektzeile	15
4.2	Effective Date & Sale Age	15
4.3	Besondere Spaltenkennzeichnungen	15
4.4	Überblick über die RCA-Anpassungen	15
5	Der Modus Market Area Analysis	16
5.1	Unterschiede zum Appraisal-Modus	16
5.2	Wann der Market-Modus sinnvoll ist	16
6	Variablenauswahl	17
6.1	Zielvariable	17
6.2	Transformation der Zielgröße	17
6.3	Die Prädiktortabelle	17
6.4	Erkennung und Überschreiben von Datentypen	18
6.5	Referenz der Special-Spaltentypen	18
7	Parameterwahl	20
7.1	Parameter-Voreinstellungen	20
7.2	Family (Verteilungsfamilie)	20
7.3	Method (Schätzverfahren)	20
7.4	Gamma (Glättungsmultiplikator)	21
7.5	Kreuzvalidierung	21
7.6	Select (Schrumpfung)	21
7.7	Default Basis (Standardbasis)	21
7.8	Default k (Basisdimension)	21
7.9	Tensortyp	22
7.10	Zugelassene Interaktionen	22
7.11	Faktor-Glättungsterm-Interaktionen	22
7.12	Erweiterte Parameter	22
8	Anpassen des Modells	24
8.1	Die Schaltfläche Fit	24
8.2	Was während der Anpassung geschieht	24
8.3	Anpassung im Hintergrund & Fortschritt	24
9	Ergebnis-Registerkarten	25
9.1	Data	25
9.2	Equation	25
9.3	Summary	25
9.4	Variable Importance	25

9.5	Beitragsdiagramme	26
9.5.1	Univariate Glättungsterme — $s(x)$	26
9.5.2	Faktor-Glättungsterme — $s(x, \text{by} = \text{Faktor})$	26
9.5.3	Interaktionsterme — $ti(x_1, x_2)$ und $te(x_1, x_2)$	26
9.5.4	Parametrische Terme — linear und faktoriell	26
9.6	Correlation	26
9.7	Diagnostics	27
9.8	RCA Adjustments	27
9.9	Anova	27
9.10	Mgcv Output	27
9.11	Sign Check	27
9.12	Concurvity	28
9.13	Basis Check	28
10	Herunterladen der Daten	29
10.1	Ausgabespalten	29
10.2	CQA-Werte	29
11	RCA-Berechnungen und Download	30
11.1	Öffnen des RCA-Dialogs	30
11.2	Interpolation des CQA-Werts	30
11.3	Ausgabespalten	30
12	Vergleichswertraster (Sales Comparison Grid)	31
12.1	Dialog zur Auswahl der Vergleichsobjekte	31
12.2	Aufbau der Arbeitsmappe	31
12.3	Blattschutz & Zellen für Restmerkmale	31
13	Herunterladen von Berichten	32
13.1	Schritt 1: Generate Quarto Report	32
13.2	Schritt 2: Convert Quarto Report	32
13.3	Inhalt des Berichts	32
14	Export von Modellfunktionen	34
14.1	R-Funktionen	34
14.2	Python-Funktionen	34
14.3	C++-Header	34
14.4	SQLite-Datenbank	34
15	Vergleich mit earthUI und glmnetUI	36
15.1	Wesentliche Unterschiede	36
15.2	Gemeinsame Funktionen	36
15.3	Empfohlener Arbeitsablauf	36
16	Demodatensatz: Appraisal_1.csv	38
16.1	Beschreibung	38
16.2	Spalten	38
16.3	Vorgeschlagener Schnelleinstieg	39

17 Systemanforderungen & Fehlerbehebung	40
17.1 Unterstützte Plattformen	40
17.2 R-Version	40
17.3 Erforderliche R-Pakete	40
17.4 Optionale Abhängigkeiten	41
17.5 Plattformspezifische Hinweise	41
17.5.1 macOS	41
17.5.2 Windows	41
17.5.3 Linux	41
17.6 Verhalten bei fehlenden Komponenten	42
17.7 Fehlerbehebung	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 Port 7880 bereits belegt	43
17.7.6 Paketinstallation schlägt unter Linux fehl	43

1 Einleitung

1.1 Geschichte und Hintergrund

1.1.1 Generalisierte additive Modelle

Generalisierte additive Modelle (Generalized Additive Models, GAMs) wurden 1990 von Trevor Hastie und Robert Tibshirani in ihrer Monographie formalisiert. GAMs erweitern das lineare Modell, indem sie jeden linearen Term $\beta_j x_j$ durch eine glatte Funktion $f_j(x_j)$ ersetzen; das Modell lautet dann:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \varepsilon$$

Jedes f_j wird mithilfe eines penalisierten Regressionssplines aus den Daten geschätzt. Die glatten Funktionen erlauben es dem Modell, nichtlineare Zusammenhänge abzubilden, ohne dass der Anwender vorab eine funktionale Form festlegen muss. Der Strafterm steuert die Welligkeit jedes Glättungsterms und verhindert eine Überanpassung, lässt aber genügend Flexibilität, um tatsächliche Muster nachzuzeichnen.

1.1.2 Das Paket mgcv

Das R-Paket `mgcv` (Mixed GAM Computation Vehicle) wurde von Simon Wood an der University of Bath entwickelt (Wood 2017). Erstmals 2001 veröffentlicht, ist es heute die R-Standardimplementierung von GAMs und Bestandteil jeder R-Installation. Die zugrunde liegenden Verfahren sind in einer Reihe von Arbeiten beschrieben, die die stabile Schätzung mehrerer Glättungsparameter (Wood 2004), Thin-Plate-Regressionssplines (Wood 2003), die schnelle stabile REML-Schätzung (Wood 2011) sowie die allgemeine Auswahl glatter Modelle (Wood, Pya und Säfken 2016) behandeln. Zu den wesentlichen Merkmalen zählen:

- **Automatische Glättungsparameterwahl:** Der Glättungsparameter jedes Terms wird im Rahmen der Modellanpassung geschätzt — mittels REML (Restricted Maximum Likelihood), GCV (Generalized Cross-Validation, generalisierte Kreuzvalidierung) oder ML (Maximum Likelihood).

- **Verschiedene Basistypen:** Thin-Plate-Regressionssplines (tp), kubische Regressionssplines (cr), P-Splines (ps) und B-Splines (bs) — jeweils mit unterschiedlichen Eigenschaften.
- **Tensorprodukt-Glättungsterme:** `te()` und `ti()` ermöglichen glatte Interaktionen zwischen Variablen, ohne dem Fluch der Dimensionalität zu unterliegen.
- **Variablenselektion:** Die Option `select = TRUE` fügt einen zusätzlichen Strafterm hinzu, der Glättungsterme auf null schrumpfen kann und damit faktisch eine Variablenselektion vornimmt.
- **Concurvity-Diagnostik:** Misst das Analogon zur Multikollinearität für Glättungsterme.

1.1.3 Vergleich mit MARS (earth) und Elastic Net (glmnet)

Die Anwendung `earth` und ihre beiden Post-Processing-Module (`glmnet` und `mgcv`) verwenden unterschiedliche Modellierungs-Engines. Die folgende Tabelle fasst die wesentlichen Unterschiede zusammen:

Merkmal	mgcv (GAM)	earth (MARS)	glmnet (Elastic Net)
Modelltyp	Glatt nichtlinear (penalisierte Splines)	Stückweise linear (adaptive Splines)	Linear (mit Regularisierung)
Nichtlinearität	Automatisch über glatte Funktionen (Thin Plate, kubisch usw.)	Automatisch über Hinge-Funktionen mit datengetriebenen Knoten	Nur über Interaktionen oder Basiserweiterung
Variablenselektion	Optional über doppelten Schrumpfungstrafterm (<code>select=TRUE</code>)	Integriert über schrittweise Vorwärts-/Rückwärtsauswahl mit GCV-Beschneidung	Integriert über L_1 -Strafterm (Lasso)
Interaktionen	Tensorprodukt-Glättungsterme (<code>te()</code> , <code>ti()</code>)	Hinge-basierte Produkte (interpretierbar, bis Grad 3)	Paarweise Kreuzprodukte (schwer interpretierbar)
Interpretierbarkeit	Glatte partielle Effektkurven; sehr anschaulich	g-Funktionen zeigen stückweise lineare partielle Effekte je Variable	Koeffizienten sind lineare Gewichte; einzeln leicht erklärbar
Am besten geeignet für	Glatte nichtlineare Zusammenhänge, flexible Modellierung	Automatische Erkennung von Nichtlinearität und Interaktionen	Hochdimensionale Daten, Variablenselektion, kleine Stichproben
RCA-Anpassungen	Glatte Anpassungen über partielle Effekte	Stückweise lineare Anpassungen über g-Funktionen	Lineare Anpassungen je Variable
Kontrolle der Überanpassung	Schätzung des Glättungsparameters über REML/GCV	GCV-basierte Beschneidung	Kreuzvalidiertes Lambda

MARS (Multivariate Adaptive Regression Splines) wurde 1991 von Jerome Friedman eingeführt. Das Verfahren bildet stückweise lineare Modelle, indem es Hinge-Funktionen und deren Knotenpositionen adaptiv aus den Daten auswählt. Die R-Implementierung ist das Paket `earth` von Stephen Milborrow.

Die **Elastic-Net**-Regression, implementiert im Paket **glmnet** von Friedman, Hastie und Tibshirani (2010), verbindet Lasso- (L_1) und Ridge-Strafterme (L_2) zur gleichzeitigen Variablen-selektion und Koeffizientenschrumpfung.

1.1.4 Die earthUI–mgcvUI-Pipeline

Ein leistungsfähiger Workflow verbindet die automatische Knotenfindung von earthUI mit der Glättungsschätzung von mgcvUI. Wenn earthUI eine Ergebnisdatei im Format `.rds` exportiert und mgcvUI diese importiert, gilt:

1. **Earth findet die Knoten** — MARS ermittelt datengetriebene Hinge-Positionen (Wendepunkte) für jede Variable.
2. **mgcvUI nutzt sie als Ankerpunkte** — Die earth-Knoten bilden die Basis für kubische Regressionssplines (Basis `cr`) im GAM und verschaffen den Glättungstermen einen Vorsprung durch Knotenpositionen, die tatsächliche Strukturen in den Daten widerspiegeln.
3. **mgcv verfeinert die Anpassung** — Die penalisierte Spline-Schätzung glättet die stückweise lineare earth-Anpassung und erzeugt glatte partielle Effektkurven, während die von earth entdeckte Grundform erhalten bleibt.

Diese Pipeline verbindet explorative Modellierung (earth) mit konfirmatorischer Modellierung (GAM) und vereint die Stärken beider Ansätze.

1.1.5 Empfohlener Arbeitsablauf

Für die Immobilienbewertung und vergleichbare Anwendungen, die interpretierbare und begründbare Modelle erfordern:

1. **Beginnen Sie mit earthUI**, um die wichtigen Variablen, nichtlinearen Zusammenhänge und Interaktionen in Ihren Daten aufzuspüren. Die automatische Basisauswahl und die g-Funktionen von earth liefern ausgezeichnete erste Einsichten.
2. **Verfeinern Sie mit mgcvUI**, wenn Sie glattere partielle Effekte wünschen. Importieren Sie die Knoten von earthUI in mgcvUI, um nahtlos von stückweise linearen zu glatten Modellen überzugehen.
3. **Verwenden Sie glmnetUI**, wenn Sie eine konsequente Variablenselektion (Lasso) benötigen, wenn die Zahl der Prädiktoren die Zahl der Beobachtungen erreicht oder übersteigt, oder wenn Sie aus Gründen der Begründbarkeit ein rein lineares Modell mit Regularisierung wünschen.

1.2 Was ist mgcvUI?

mgcvUI ist eine grafische Benutzeroberfläche für das R-Paket **mgcv**. Sie läuft als lokale Shiny-Anwendung — ohne Anmeldung, ohne Server und ohne Benutzerkonten. Sie starten sie aus R heraus, importieren einen Datensatz (CSV oder Excel), konfigurieren Ihr Modell und passen interaktiv ein GAM an.

Seit der aktuellen Version ist mgcvUI kein separat installiertes Paket mehr: Es ist *im* Paket earthUI als dessen GAM-Post-Processing-Modul enthalten und wird mit `earthUI::launch_mgcv()` gestartet. Die Anwendung selbst — und alles, was in diesem Artikel beschrieben wird — ist unverändert; lediglich Installation und Startbefehl unterscheiden sich. Die drei Routinen (earth, glmnet, mgcv) teilen sich einen Projektbaum, eine Einstellungsdatenbank und eine gemeinsame unterstützende Infrastruktur.

Die Anwendung bietet einen vollständigen Arbeitsablauf: Datenimport, Variablenkonfiguration mit Spezifikation der Glättungsterme, Modellanpassung mit Hintergrundverarbeitung, Diagnosedigramme, glatte partielle Effektkurven und herunterladbare Berichte im Word-, PDF- oder HTML-Format.

Generalisierte additive Modelle ersetzen den linearen Zusammenhang βx für jeden Prädiktor durch eine glatte Funktion $f(x)$. Das bedeutet:

- **Keine Notwendigkeit, funktionale Formen vorzugeben** — die Daten bestimmen die Gestalt jedes Zusammenhangs.
- **Glatte partielle Effekte** — der Beitrag jedes Prädiktors ist eine glatte Kurve statt einer Geraden, sodass leicht erkennbar wird, wie eine Variable die Zielgröße bei verschiedenen Werten beeinflusst.
- **Automatische Glättungsparameterwahl** — `mgcv` bestimmt über REML oder GCV, wie wellig jede Kurve sein soll, und wägt dabei Anpassungsgüte gegen Überanpassung ab.

1.3 Die drei Purpose-Modi

Jedes Projekt besitzt einen **Purpose** (Verwendungszweck), der bei der Anlage in Abschnitt 1 der Seitenleiste (**Project**) gewählt und danach festgeschrieben wird. Der Purpose des aktiven Projekts bestimmt, welche Werkzeuge und Oberflächenelemente erscheinen; um in einem anderen Modus zu arbeiten, legen Sie ein Projekt mit dem entsprechenden Purpose an oder öffnen ein solches. Die drei Modi sind:

- **General** (allgemein) — GAM-Regression für beliebige Grundgesamtheiten oder Datensätze. Dies ist der Standardmodus. Er stellt den vollständigen GAM-Workflow ohne fachspezifische Ergänzungen bereit.
- **For Appraisal** (für die Wertermittlung) — GAM-Regression, zugeschnitten auf die Immobilienbewertung. Ergänzt Funktionen für die Einzelobjektbewertung, darunter die Behandlung des Bewertungsobjekts, besondere Spaltenkennzeichnungen und den Residual Constraint Approach (RCA).
- **Market Area Analysis** (Marktgebietsanalyse) — GAM-Regression, zugeschnitten auf Marktgebietsstudien. Ergänzt Funktionen zur Analyse von Objektgruppen in einem definierten Markt, mit optionaler Behandlung einer Bewertungsobjektzeile.

In allen drei Modi ist die eigentliche Modellierungs-Engine identisch — Sie passen stets ein GAM über `mgcv::gam()` an. Die Purpose-Einstellung steuert lediglich, welche zusätzlichen Werkzeuge und Oberflächenelemente verfügbar sind.

1.4 Immobilienspezifische Funktionen

Ist entweder **For Appraisal** oder **Market Area Analysis** gewählt, aktiviert `mgcvUI` mehrere Funktionen, die für die Immobilienanalyse konzipiert sind:

- **Besondere Spaltenkennzeichnungen** — Jeder Prädiktor kann mit einer besonderen Rolle versehen werden, etwa `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` oder `display_only`. Diese Kennzeichnungen steuern, wie die Spalte bei der Anpassung und in der Ausgabe behandelt wird.
- **Spalte Sale Age** — Ist eine Spalte als `contract_date` gekennzeichnet und ein Effective Date (Wertermittlungstichtag) angegeben, berechnet `mgcvUI` eine Spalte `sale_age` (Tage zwischen Verkaufsdatum und Stichtag) und setzt sie als Prädiktor ein.
- **RCA-Berechnungen (nur Wertermittlung)** — Im Wertermittlungsmodus kann `mgcvUI` nach der Modellanpassung die Ausgabe des Residual Constraint Approach (RCA) berechnen. Diese liefert Anpassungen je Vergleichsobjekt, Zusammenfassungen der Netto- und Bruttoanpassungen sowie einen angepassten Verkaufspreis für das Bewertungsobjekt.

Hinweis: Der Purpose-Modus General eignet sich für beliebige Datensätze — aus Finanzwesen, Wissenschaft, Ingenieurwesen oder Immobilienwirtschaft. Die Modi für Wertermittlung und Marktanalyse fügen lediglich Komfortfunktionen für Immobilienfachleute hinzu.

1.5 Erste Schritte

So verwenden Sie mgcvUI:

1. **Installieren Sie earthUI** in R: `install.packages("earthUI")` — die mgcv-Routine ist enthalten. (Zur Entwicklungsversion und zu optionalen Komponenten siehe „Installing earthUI and vProlog“ im earthUI-Artikel dieser Ausgabe.)
2. **Starten Sie die Anwendung:** Führen Sie in R `earthUI::launch_mgcv()` aus. Die Anwendung öffnet sich in Ihrem Webbrowser auf Port 7880.
3. **Legen Sie ein Projekt an oder öffnen Sie eines** in Abschnitt 1 der Seitenleiste. Ein Projekt legt den Purpose (General, For Appraisal oder Market Area Analysis) sowie den Arbeitsort (Land, Bundesland, Kreis, Stadt) und seine Ein- und Ausgabeordner fest. Projekte werden unter einem gemeinsamen `regProj`-Stammordner abgelegt und mit den Schwesteranwendungen earthUI und glmnetUI geteilt. Das Öffnen eines Projekts mit einem anderen Purpose ist der Weg, den Modus zu wechseln.
4. **Importieren Sie Ihre Daten** in Abschnitt 2, indem Sie eine CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts auswählen. Legen Sie Dateien in diesem Ordner ab und klicken Sie dann auf **Refresh**, um sie aufzulisten.
5. **Import aus earthUI (optional)** — Abschnitt 3 erlaubt den Import einer earthUI-Ergebnisdatei (`.rds`), um das GAM mit von earth gefundenen Knoten zu initialisieren. Überspringen Sie diesen Schritt, wenn Sie kein earthUI-Ergebnis haben.
6. **Konfigurieren Sie die Variablen** — wählen Sie Zielgröße und Prädiktoren, legen Sie Datentypen fest, weisen Sie besondere Spaltenrollen zu und markieren Sie Variablen als Faktor oder linear.
7. **Setzen Sie die mgcv-Parameter** — family, method, gamma, Basistyp, k, Tensortyp, Interaktionen und weitere Optionen.
8. **Passen Sie das Modell an** — klicken Sie auf „Fit Mgcv GAM Model“ und prüfen Sie die Ergebnisse im Hauptbereich.
9. **Exportieren Sie** — laden Sie Prognosen als Excel-Datei herunter, erzeugen und konvertieren Sie einen Quarto-Bericht oder berechnen Sie (im Wertermittlungsmodus) RCA-Anpassungen und ein Vergleichswertraster.

Die Einstellungen werden automatisch in einer SQLite-Datenbank gespeichert und beim erneuten Laden derselben Eingabedatei wiederhergestellt.

2 Anforderungen an MLS-Eingabedaten

Bei Arbeitsabläufen der Immobilienbewertung und Marktanalyse stammen Ihre Eingabedaten in der Regel aus einem Export eines Multiple Listing Service (MLS). Dieses Kapitel beschreibt die erwartete Dateistruktur und die Spalten, die mgcvUI verwenden kann.

2.1 Dateiformat & Struktur

mgcvUI akzeptiert **CSV**- und **Excel**-Dateien (**.xlsx**, **.xls**). Beim Import werden Spaltennamen automatisch in **snake_case** umgewandelt — aus „Living SqFt“ wird beispielsweise **living_sqft**, aus „Contract Date“ wird **contract_date** und aus „Sale Price“ wird **sale_price**. Diese Normalisierung sorgt für einheitliche Spaltenbezüge im gesamten Arbeitsablauf. Das beim Import verwendete CSV-Trennzeichen und Dezimaltrennzeichen richtet sich nach den Locale-Einstellungen (siehe Kapitel 3, „Locale & Regional Settings“).

Ihre Datendatei sollte eine flache Tabelle mit einer Zeile je Objekt und einer Spalte je Merkmal sein. Die erste Zeile der Datei muss die Spaltenüberschriften enthalten.

2.2 Erforderliche Spalten im Wertermittlungsmodus

mgcvUI arbeitet zwar mit beliebigen Spaltensätzen, der vollständige Wertermittlungs-Workflow profitiert jedoch von den folgenden Spalten:

Spalte	Special Type	Zweck
Sale Price	(Zielgröße)	Zielvariable des Modells
Contract Date	contract_date	Dient zur Berechnung von sale_age (Tage ab Stichtag)
Listing Date	listing_date	Zusammen mit dem Vertragsdatum zur DOM-Berechnung, falls keine DOM-Spalte vorliegt
Days on Market	dom	Vermarktungsdauer in Tagen; wird in Exporten angezeigt
Concessions	concessions	Verkaufszugeständnisse; Netto-Verkaufspreis = Verkaufspreis – Zugeständnisse
Living Area (SF)	living_area	Ermöglicht Residuen je Flächeneinheit (residual_sf , cqa_sf)
Lot Size	lot_size	Spalte für die Grundstücksgröße
Site Dimensions	site_dimensions	Wird mit der Grundstücksgröße gruppiert
Latitude	latitude	Auf 3 Nachkommastellen gerundet
Longitude	longitude	Auf 3 Nachkommastellen gerundet
Area ID	area	Kennung des Marktgebiets bzw. der Nachbarschaft
Actual Age	actual_age	Tatsächliches Alter des Objekts
Effective Age	effective_age	Effektives Alter des Objekts
Address	display_only	Erscheint in Exporten; vom Modell ausgeschlossen

Die Spaltennamen in der Tabellenkalkulation dürfen in einer beliebigen Sprache verfasst sein — die „Special“-Bezeichnungen sind englisch, damit das R-Programm sie besonders behandeln kann. Ansonsten erscheinen die von Ihnen vergebenen Spaltennamen in den Regressionsmodellen, den Grafiken und — bei Wertermittlungen — in den Ausgabeberichten.

Nicht alle Spalten sind erforderlich. mgcvUI passt sich an: Fehlt eine Spalte, entfällt schlicht die zugehörige Funktion. Für Immobilienpreismodelle sind bestimmte Spalten jedoch dringend zu empfehlen:

1. **Sale Age** — die Anzahl der Tage zwischen dem Kaufvertragsdatum und dem Stichtag. Wird eine mehrjährige Verkaufshistorie verwendet, spielt **sale_age** häufig eine zentrale Rolle.
2. **Living Area** (Wohnfläche) — auch unter Bezeichnungen wie „Living Sqft“ oder „GLA“ (gross living area) geführt.
3. **Total Bath Count** — die Gesamtzahl der ganzen, viertel, halben und dreiviertel Badezimmer.
4. **Garage Bays** oder **Garage Area** — Garagenstellplätze bzw. Garagenfläche.
5. **Lot Size** — die Grundstücksfläche des Objekts.
6. **Longitude, Latitude** und, sofern verfügbar, **Area ID** zur Abbildung geografischer Preisunterschiede.

2.3 Konventionen für besondere Spaltenbezeichnungen

mgcvUI erkennt Spalten anhand ihrer **Special-Type-Kennzeichnung**, nicht anhand ihres Spaltennamens. Sie können Ihre Spalten im MLS-Export beliebig benennen — entscheidend ist, dass Sie in der Tabelle Variable Configuration (Kapitel 6) den richtigen Special Type zuweisen.

2.4 Datenqualität & Vollständigkeit

- **Fehlende Werte (NA)**: Zeilen mit NA-Werten in einer Prädiktor- oder Zielspalte werden vor der Anpassung automatisch entfernt. mgcvUI beschränkt die Daten auf die im Modell verwendeten Spalten, sodass NAs in nicht verwendeten Spalten keinen Zeilenverlust verursachen.
- **Numerische Spalten**: Verkaufspreis, Wohnfläche, Grundstücksgröße und vergleichbare Felder müssen numerisch sein.
- **Faktorspalten**: Text-, TRUE/FALSE- und R-Faktor-Spalten werden automatisch als kategorial behandelt — sie können nur als Faktoren in das Modell eingehen. Eine **numerische** Spalte (etwa ein numerischer Gebiets- oder Stilcode) wird nur dann zum Faktor, wenn Sie das zugehörige Kontrollkästchen **Factor** setzen; mgcvUI leitet niemals aus dem Wertebereich einer numerischen Variable ab, dass sie kategorial sei. Diskrete numerische Prädiktoren wie **bath_count** bleiben daher stetig (geglättet). Die Special-Rolle ist von der Faktorentscheidung unabhängig.

***Hinweis:** Prüfen Sie nach dem Import die NA-Spalte in der Prädiktortabelle. Spalten mit vielen fehlenden Werten können dazu führen, dass Zeilen entfallen. Spalten mit $\geq 50\%$ NAs werden rot markiert und automatisch ausgeschlossen.*

2.5 Platzierung der Bewertungsobjektzeile

Im Modus **Appraisal** muss **Zeile 1** das **Bewertungsobjekt** enthalten. Alle übrigen Zeilen sind Vergleichsverkäufe. Die Zeile des Bewertungsobjekts wird von der Modellanpassung ausgeschlossen. Nach der Anpassung erzeugt das Modell dennoch Prognosen für diese Zeile.

Im Modus **Market Area Analysis** ist die Platzierung des Bewertungsobjekts in Zeile 1 optional.

Im Modus **General** gibt es keine besondere Zeilenbehandlung — alle Zeilen werden gleich behandelt.

3 Der Modus General Purpose

3.1 Überblick

Der Modus General Purpose ist beim Start von mgcvUI voreingestellt. Er stellt den vollständigen GAM-Workflow für beliebige Datensätze bereit — nicht nur für Immobilien. Sie können mgcvUI für wissenschaftliche Daten, Finanzanalysen, ingenieurtechnische Untersuchungen oder jedes Regressionsproblem einsetzen, bei dem glatte nichtlineare Zusammenhänge zu erwarten sind.

Im General-Modus lässt die Oberfläche die immobilienpezifischen Funktionen (Special-Spalten, Sale Age, RCA) weg. Die Seitenleiste ist gestrafft und konzentriert sich auf Variablenauswahl, Parameterkonfiguration, Modellanpassung und Export.

3.2 Der Arbeitsablauf in der Seitenleiste

Die Seitenleiste ist in nummerierte, aufklappbare Abschnitte gegliedert, die Sie von der Projektauswahl bis zum Berichtsexport führen. Die meisten Abschnitte erscheinen erst, wenn ein aktives Projekt geöffnet ist, und einige sind vom Purpose des Projekts abhängig. Der Ausgabeordner wird aus dem aktiven Projekt abgeleitet — ein separates Feld für den Ausgabeordner gibt es nicht.

1. Project — Projekt anlegen oder öffnen. Ein Projekt legt den Purpose und den Arbeitsort (Land, Bundesland, Kreis, Stadt) sowie seinen Eingabeordner (`<os>_in/`) und Ausgabeordner fest. Projekte werden unter einem gemeinsamen `regProj`-Stammordner abgelegt (Vorgabe `~/regProj`, überschreibbar über die Umgebungsvariable `REGPROJ_ROOT` oder das Settings-Menü) und mit den Schwesteranwendungen `earthUI` und `glmnetUI` geteilt, sodass Modelle aus allen dreien nebeneinander liegen. Klicken Sie auf **Close Project**, um zu wechseln.

2. Import Data — Wählen Sie eine CSV- oder Excel-Datei aus dem Eingabeordner des aktiven Projekts. Legen Sie Dateien über den Datei-Explorer bzw. Finder im Ordner ab und klicken Sie dann auf **Refresh**, um sie aufzulisten. Bei Excel-Dateien mit mehreren Blättern erscheint eine Blattauswahl. Spaltennamen werden automatisch in `snake_case` umgewandelt.

3. Import from earthUI (optional) — Importieren Sie eine `earthUI`-Ergebnisdatei (`.rds`), um das GAM mit von `earth` gefundenen Knotenpositionen (und mit Variablenauswahlen) zu initialisieren. Dieser Schritt ist optional — überspringen Sie ihn, wenn Sie kein `earthUI`-Ergebnis zum Importieren haben.

4. Variable Configuration — Auswahl der Zielvariable, Transformation der Zielgröße (none, log, log10), Prädiktortabelle mit Kontrollkästchen für Include, Factor und Linear. Die Spalte Special erscheint nur in den Modi Appraisal und Market. Einzelheiten siehe Kapitel 6.

5. MgcV Call Parameters — Die gesamte Modellkonfiguration: Parameter-Voreinstellungen, family, method, gamma, Kreuzvalidierung, select, Basistyp, k, Tensortyp, Interaktionsmatrix und erweiterte Parameter. Die vollständige Parameterreferenz finden Sie in Kapitel 7.

6. Fit MgcV GAM Model — Die Schaltfläche, die das Modell ausführt. Die Ergebnisse erscheinen in den Registerkarten rechts.

7. Download Output — Exportiert Prognosen, Residuen, CQA-Werte und Beiträge je Variable als Excel-Datei in den Ausgabeordner des Projekts. In allen Purpose-Modi verfügbar, sobald das Modell angepasst ist.

8. Calculate RCA Adjustments & Download — Berechnet die RCA-Anpassungen je Variable für jedes Vergleichsobjekt und interpoliert das Residuum des Bewertungsobjekts über CQA. Erscheint nur in den Modi Appraisal und Market, nach der Anpassung.

9. Generate Sales Grid & Download — Erzeugt das Vergleichswertraster (Sales Comparison Grid) und empfiehlt Vergleichsobjekte nach dem Prozentsatz der Bruttoanpassung. Erscheint nur im Appraisal-Modus.

10. Generate Quarto Report — Schreibt ein in sich geschlossenes Quarto-Berichtspaket (`.qmd`) — mit allen Diagrammen und Tabellen — in den Ausgabeordner des Projekts. Verfügbar, sobald das Modell angepasst ist.

11. Convert Quarto Report — Rendert das erzeugte `.qmd`-Paket nach HTML, Word oder PDF. Für PDF ist eine LaTeX-Installation erforderlich; wird keine erkannt, ist die PDF-Option ausgeblendet. Siehe Kapitel 13.

Die purpose-spezifischen Schritte entfallen (und die übrigen Schritte werden neu nummeriert) in den anderen Modi: Im Market-Modus fehlt der Schritt Sales Grid, im General-Modus fehlen sowohl der RCA- als auch der Sales-Grid-Schritt.

3.3 Import aus earthUI (optional)

Abschnitt 3 der Seitenleiste — **Import from earthUI** — erlaubt den Import einer earthUI-Ergebnisdatei (`.rds`). Damit wird die earth-mgcv-Pipeline möglich: Die datengetriebenen Knotenpositionen von earth werden zu Ankerpunkten für die GAM-Glättungsterme.

Wird ein earthUI-Ergebnis importiert, geschieht in mgcvUI Folgendes:

- Es erscheint eine Zusammenfassung mit Zielvariable, Bestimmtheitsmaß, Anzahl der Terme und Knotenpositionen je Variable.
- Die Parameter-Voreinstellung „Earth Pipeline“ wird automatisch gewählt (Basis kubischer Regressionsspline, $\gamma = 1.4$).
- Die Zielvariable wird auf die Zielgröße von earth gesetzt.
- Include wird für die Prädiktorvariablen von earth vorab aktiviert.
- Factor wird für die kategorialen Variablen von earth und Linear für die linearen Prädiktoren von earth vorab aktiviert.
- Von earth erkannte Interaktionen werden vorbelegt (in der Interaktionsmatrix gelb hervorgehoben und gegen manuelle Änderungen gesperrt).
- Eine Schaltfläche **Export Knots CSV** stellt die Knoten von earth als CSV-Datei zum Download bereit.

Die Tabelle Variable Configuration zeigt daraufhin einen Hinweis, dass die Variablenauswahlen (Include, Linear, Factor) und die Spline-Knoten aus dem earth-Modell *übernommen* wurden. Anders als in glmnetUI, wo die glmnet-Anpassung an die Basis von earth gebunden ist, passt mgcvUI das GAM anhand Ihrer aktuellen Auswahl an — earth verschafft lediglich einen Vorsprung, und jedes Bedienelement bleibt änderbar.

Eine Schaltfläche **Clear** entfernt die importierten earth-Daten und setzt auf den eigenständigen Betrieb zurück.

3.4 Registerkarten des Hauptbereichs

Nach dem Datenimport bietet der Hauptbereich die folgenden Registerkarten (modellabhängige Registerkarten füllen sich erst nach der Anpassung):

Registerkarte	Inhalt
Data	Interaktive DataTable der importierten Daten. In den Modi Appraisal und Market aufgeteilt in Tabellen für Bewertungsobjekt und Vergleichsverkäufe.
Equation	Vollständige GAM-Formel in MathJax-Darstellung, dazu eine Tabelle der Definitionen der glatten Funktionen.
Summary	Kennzahlen (R^2 , $CV\text{-}R^2$, erklärte Devianz, AIC, n, Anzahl der Glättungsterme) sowie Tabellen der glatten und parametrischen Terme.
Variable Importance	Waagerechtes Balkendiagramm der F-Statistiken (glatt) und $ t $ -Werte (parametrisch), dazu eine Rangtabelle.
Contribution	Interaktive Plotly-Kurven der glatten partiellen Effekte mit Konfidenzbändern und Markierungen der earth-Knoten.
Correlation	Heatmap der Korrelationen numerischer Prädiktoren (bereits vor der Anpassung verfügbar).
Diagnostics	Residuen gegen angepasste Werte, Q-Q-Diagramm, Histogramm, Zielgröße gegen angepasste Werte (über gratia). Streudiagramm beobachtet gegen prognostiziert.
RCA Adjustments	Histogramme der RCA-Analyse (Modi Appraisal/Market, nach der Berechnung).
Anova	Kombinierte ANOVA-Tabelle für parametrische und glatte Terme.
MgcV Output	Rohtext: Laufzeiten, Formel, Modellzusammenfassung, Familie, Concurvity.
Sign Check	Prüfung der Richtungskonsistenz zwischen earth und GAM (bei importiertem earth-Ergebnis).
Concurvity	Gesamt- und paarweise Concurvity-Matrizen mit Farbcodierung.
Basis Check	Tests der Angemessenheit der Basisdimension über gam.check() .

3.5 Dauerhafte Speicherung der Einstellungen

mgcvUI speichert Ihre Konfiguration automatisch in einer SQLite-Datenbank, indiziert nach dem Namen der Eingabedatei. Wenn Sie dieselbe Datei erneut laden, werden alle Einstellungen wiederhergestellt: Zielauswahl, Prädiktor-Kontrollkästchen, Datentypen, mgcv-Parameter und Interaktionsmatrix. Der Purpose ist eine Eigenschaft des aktiven Projekts und keine globale Einstellung; die zuletzt verwendete Eingabedatei wird je Projekt gespeichert und beim erneuten Öffnen des Projekts automatisch geladen.

3.6 Dark Mode

Klicken Sie auf das Mond-/Sonnensymbol oben rechts, um zwischen den Themes Nord Light und Nord Dark zu wechseln. Die Theme-Einstellung wird im localStorage gespeichert und bleibt über Sitzungen hinweg erhalten. Alle Oberflächenelemente passen sich dem gewählten Theme an.

3.7 Locale & Regional Settings (Gebietsschema und regionale Einstellungen)

mgcvUI unterstützt internationale Konventionen für Zahlen- und CSV-Formate über ein länderbasiertes Locale-System. Das **Settings**-Menü in der Titelleiste bietet Auswahlfelder für **Country** (Land) und **Paper** (Papierformat) für über 30 unterstützte Länder. Jede Voreinstellung legt fest:

- **CSV-Trennzeichen** — Komma (,) für USA/Großbritannien/Japan oder Semikolon (;) für den größten Teil Europas.
- **Dezimaltrennzeichen** — Punkt (.) oder Komma (,).

- **Tausendertrennzeichen** — Komma (USA/Großbritannien/Japan), Punkt (Deutschland/Italien/Spanien), Leerzeichen (Finnland/Frankreich/Polen/Baltikum) oder Apostroph (Schweiz).
- **Papierformat** — Letter (USA/Kanada/Mexiko) oder A4 (überall sonst).

Klicken Sie auf **Save**, um Ihre Locale-Einstellungen global zu speichern. Der Abschnitt Import Data verfügt zudem über ein dateibezogenes Auswahlfeld **Locale** für einmalige Importe mit abweichenden Konventionen. Dasselbe Settings-Menü enthält das Feld **regProj root folder**, mit dem der gemeinsame Projektbaum von mgcvUI, earthUI und glmnetUI verlegt werden kann.

***Hinweis:** Die Zahlenformatierung auf Diagrammachsen und Steigungsbeschriftungen passt sich automatisch an das eingestellte Locale an. Das deutsche Locale verwendet Punkte als Tausendertrennzeichen (200.000), das finnische Leerzeichen (200 000), das schweizerische Apostrophe (200'000). Währungssymbole werden nicht angezeigt — mgcvUI ist währungsneutral.*

4 Der Appraisal-Modus

Ist der Purpose des aktiven Projekts **For Appraisal**, richtet sich mgcvUI auf die Bewertung eines einzelnen Objekts ein. Alle in Kapitel 3 beschriebenen Funktionen bleiben verfügbar; dieses Kapitel behandelt nur die wertermittlungsspezifischen Ergänzungen.

4.1 Behandlung der Bewertungsobjektzeile

Im Wertermittlungsmodus ist **Zeile 1 Ihres Datensatzes das Bewertungsobjekt**, alle übrigen Zeilen sind Vergleichsverkäufe. Ihre Eingabedatei muss entsprechend aufgebaut sein (siehe Kapitel 2). Der Verkaufspreis des Bewertungsobjekts kann leer bleiben oder einen beliebigen Wert enthalten — mgcvUI behandelt ihn bei der Anpassung automatisch als NA.

Nach der Anpassung erzeugt das Modell Prognosen für die Zeile des Bewertungsobjekts; sie bilden die Grundlage des RCA-Anpassungsablaufs.

4.2 Effective Date & Sale Age

In den Modi Appraisal und Market erscheint im Abschnitt Variable Configuration ein Feld **Effective Date** (Wertermittlungstichtag), das mit dem heutigen Datum vorbelegt ist. Kennzeichnen Sie eine Spalte im Special-Auswahlfeld als **contract_date**, berechnet mgcvUI eine Spalte **sale_age** — die Anzahl ganzer Tage zwischen dem Vertragsdatum des jeweiligen Verkaufs und dem Stichtag. Diese Spalte wird als Prädiktor ergänzt.

Ändert sich das Effective Date, wird **sale_age** automatisch neu berechnet.

4.3 Besondere Spaltenkennzeichnungen

In den Modi Appraisal und Market erscheint in der Tabelle Variable Configuration für jeden Prädiktor ein Auswahlfeld **Special**. Die vollständige Referenz der Special Types und ihrer Wirkungen finden Sie in Kapitel 6.

4.4 Überblick über die RCA-Anpassungen

Die Schaltfläche **Calculate RCA Adjustments & Download** (in den Modi Appraisal und Market nach der Anpassung sichtbar) berechnet marktabgeleitete Anpassungen für jedes Vergleichsobjekt im Verhältnis zum Bewertungsobjekt. Der vollständige RCA-Ablauf ist in Kapitel 11 beschrieben.

***Hinweis:** Der dem Bewertungsobjekt zugewiesene CQA-Wert steuert, welcher Anteil der Residualverteilung dem Bewertungsobjekt zugerechnet wird. Ein Wert von 5.00 platziert das Bewertungsobjekt am Median der Vergleichsobjekte.*

5 Der Modus Market Area Analysis

Ist der Purpose des aktiven Projekts **Market Area Analysis**, stellt mgcvUI dieselben immobilien-spezifischen Funktionen bereit wie im Wertermittlungsmodus (Special-Spalten, Sale Age), ist jedoch auf die Analyse einer Objektgruppe statt auf die Bewertung eines einzelnen Objekts ausgerichtet.

5.1 Unterschiede zum Appraisal-Modus

- **Flexible Behandlung der Bewertungsobjektzeile** — im Market-Modus wird Zeile 1 nicht automatisch als Bewertungsobjekt behandelt.
- **Kein Sales Grid** — der Schritt „Generate Sales Grid & Download“ steht nicht zur Verfügung.

5.2 Wann der Market-Modus sinnvoll ist

Der Modus Market Area Analysis eignet sich, wenn Sie:

- ein GAM-Modell für eine Nachbarschaft oder ein Marktgebiet aufbauen, um nichtlineare Werttreiber zu verstehen
- analysieren, wie Variablen wie Wohnfläche, Alter, Grundstücksgröße und Lage die Verkaufspreise in einer Objektgruppe beeinflussen
- Belege für eine Marktkonditionenanalyse oder eine Abgrenzung der Nachbarschaft vorbereiten
- mit einem Datensatz arbeiten, für den Sie die Special-Spalten-Funktionen (Sale Age) benötigen, den Sales-Grid-Ablauf jedoch nicht

***Hinweis:** Der Market Mode eignet sich auch für allgemeine Immobilienregressionen, bei denen die Special-Spalten-Funktionen benötigt werden, der Ablauf zur Bewertung eines einzelnen Objekts jedoch nicht.*

6 Variablenauswahl

In Abschnitt 4 der Seitenleiste — **Variable Configuration** — legen Sie fest, welche Spalten in das Modell eingehen und wie sie behandelt werden.

6.1 Zielvariable

Das Auswahlfeld **Target (response) variable** am Kopf von Abschnitt 4 listet jede numerische Spalte Ihres Datensatzes auf. Wählen Sie eine Spalte als Zielvariable (etwa den Verkaufspreis). Die Zielspalte wird automatisch aus der Prädiktorliste ausgeschlossen.

6.2 Transformation der Zielgröße

Unterhalb der Zielauswahl bietet ein Auswahlfeld **Response Transform** drei Möglichkeiten:

- **None (raw)** — Die Zielgröße wird unverändert verwendet. Angebracht, wenn der Zusammenhang zwischen Prädiktoren und Zielgröße auf der ursprünglichen Skala näherungsweise additiv ist.
- **Log (natural)** — Vor der Anpassung wird der natürliche Logarithmus angewandt. Nach der Rücktransformation werden die Anpassungen dadurch proportional (%) statt absolut (\$). Besonders nützlich bei Verkaufspreisen, wo eine Anpassung von 10.000 \$ bei unterschiedlichen Preisniveaus sehr Unterschiedliches bedeutet.
- **Log10** — Es wird der Zehnerlogarithmus angewandt.

***Hinweis:** Die Log-Transformation macht Zeitanpassungen proportional (%) statt absolut (\$). Bei Immobilienmodellen über eine breite Preisspanne führen Log-Transformationen häufig zu besser verteilten Residuen.*

Ist eine Log-Transformation gewählt, werden Werte ≤ 0 in der Zielgröße automatisch herausgefiltert. Sämtliche Ausgaben (Prognosen, Beiträge, Residuen) werden automatisch auf die ursprüngliche Skala zurücktransformiert.

6.3 Die Prädiktortabelle

Unterhalb der Zielauswahl listet eine scrollbare Tabelle alle übrigen Spalten auf. Die Kontrollkästchenspalten verwenden aus Platzgründen gedrehte vertikale Kopfzeilen. Die Spaltenreihenfolge lautet:

Spalte	Beschreibung
Variable	Spaltenname (Festbreitenschrift, bei Bedarf mit Auslassungspunkten gekürzt). Umrandete Zelle.
Type	Auswahlfeld für den Datentyp: numeric, integer, character, factor, logical, Date, POSIXct. Umrandetes Auswahlfeld.
Include	Kontrollkästchen — diese Spalte als Prädiktor in das Modell aufnehmen. Vertikale Kopfzeile.
Factor	Kontrollkästchen — als kategorialen Faktor behandeln (erzeugt je Stufe einen eigenen Koeffizienten). Vertikale Kopfzeile.
Linear	Kontrollkästchen — linearen Zusammenhang erzwingen (kein Glättungsterm, nur βx). Vertikale Kopfzeile.
Special	Auswahlfeld (nur Appraisal/Market) — siehe Referenz der Special-Spaltentypen weiter unten. Umrandetes Auswahlfeld.
NAs	Anzahl und Anteil fehlender Werte, farblich codiert: rot ($\geq 50\%$), orange ($\geq 20\%$), grau (>0). Umrandete Zelle.

Factor vs. Smooth vs. Linear: Standardmäßig erhalten einbezogene numerische Variablen einen Glättungsterm $f(x)$. Das Setzen von **Factor** erzeugt einen kategorialen Term (einen Koeffizienten je Stufe). Das Setzen von **Linear** erzeugt statt eines Glättungsterms einen einfachen linearen Term βx . Variablen, die sowohl als Factor als auch als Linear markiert sind, werden als Factor behandelt.

6.4 Erkennung und Überschreiben von Datentypen

mgcvUI erkennt die Datentypen beim Import automatisch. Numerische, ganzzahlige, logische, Faktor- und Datumsspalten werden erkannt. Sie können jede Erkennung über das Auswahlfeld **Type** überschreiben. Eine Typänderung wirkt sich darauf aus, wie die Spalte im Modell behandelt wird.

6.5 Referenz der Special-Spaltentypen

In den Modi Appraisal und Market bietet das Auswahlfeld **Special** die folgenden Optionen:

Gewichtung:

- **weight** — Spalte mit Beobachtungsgewichten (nur eine zulässig; Zeilen mit Gewicht = 0 werden von der Anpassung ausgeschlossen)

Datums- und Zeittypen:

- **contract_date** — Löst die automatische Berechnung von **sale_age** ab dem Effective Date aus
- **sale_age** — Kennzeichnet eine bereits berechnete Sale-Age-Spalte (Tage ab dem Stichtag), die unverändert verwendet wird
- **listing_date** — Wird als Rückfalloption für die Berechnung der Vermarktungsdauer verwendet
- **dom** — Kennzeichnet die Spalte mit der Vermarktungsdauer (Days on Market)

Datensatztyp:

- **sale_type** — Kennzeichnet die Spalte des Datensatztyps, die das Bewertungsobjekt von den Vergleichsverkäufen unterscheidet

Monetäre Typen:

- **concessions** — Kennzeichnet Verkaufszugeständnisse (Verkäuferzuschüsse, Käuferanreize usw.)

Größen- und Lagetypen:

- **latitude** — Werte werden automatisch auf 3 Nachkommastellen gerundet
- **longitude** — Gleiche Rundung wie bei latitude
- **area** — Kennung des Marktgebiets bzw. der Nachbarschaft
- **living_area** — Ermöglicht die Berechnung von Residuen je Flächeneinheit (**residual_sf** und **cqa_sf**)
- **lot_size** — Spalte für die Grundstücksgröße
- **site_dimensions** — Wird mit der Grundstücksgröße gruppiert

Alterstypen:

- **actual_age** — Spalte für das tatsächliche Objektalter
- **effective_age** — Effektives Objektalter

Anzeigetypen:

- **display_only** — Die Spalte wird in die Excel-Exporte aufgenommen, aber vollständig von der Modellanpassung ausgeschlossen. Geeignet für Adressfelder, MLS-Nummern, Flurstückskennungen oder sonstige Referenzdaten.

Hinweis: Als *display_only* gekennzeichnete Spalten verbleiben in Ihrem Datensatz und erscheinen im Excel-Export, werden jedoch vollständig aus der Prädiktortabelle ausgeschlossen. Diese Kennzeichnung eignet sich für ID-Spalten, Adressen oder sonstige Referenzfelder.

7 Parameterwahl

Abschnitt 5 der Seitenleiste — **Mgcv Call Parameters** — eröffnet den Zugang zu sämtlichen Konfigurationsoptionen des GAM. Jeder Parameter besitzt ein blaues Hilfesymbol (?) mit einer Erläuterung als Kurzinfo.

7.1 Parameter-Voreinstellungen

Ein Auswahlfeld am Kopf bietet zwei Voreinstellungen mit sinnvollen Standardwerten für gängige Arbeitsabläufe:

Voreinstellung	Basis	Einstellungen
Standalone (Erkundung)	tp (Thin Plate)	k = auto (10), gamma = 1.2, select = TRUE
Earth Pipeline (Verfeinerung)	cr (kubische Regression)	k = auto (10), gamma = 1.4, select = TRUE

Die Voreinstellung **Earth Pipeline** wird automatisch gewählt, wenn earthUI-Knoten importiert werden. Die Basis des kubischen Regressionssplines (**cr**) ist für die Einbindung der earth-Knoten erforderlich, weil sie die Angabe exakter Knotenpositionen erlaubt.

7.2 Family (Verteilungsfamilie)

Wählen Sie die Verteilungsfamilie für Ihre Zielvariable:

Familie	Anwendungsfall
gaussian	Stetige Zielgrößen (z. B. Verkaufspreis). Am gebräuchlichsten.
Gamma	Positive stetige Zielgrößen mit einer Varianz proportional zum Mittelwert.
poisson	Zählraten (z. B. Anzahl der Verkäufe).
binomial	Binäre Ergebnisse.
inverse.gaussian	Stark rechtsschiefe positive Daten.

7.3 Method (Schätzverfahren)

Das Verfahren zur Schätzung des Glättungsparameters:

Verfahren	Beschreibung
REML	Restricted Maximum Likelihood (Vorgabe, empfohlen). Am robustesten gegen Überanpassung.
GCV.Cp	Generalisierte Kreuzvalidierung. Neigt dazu, etwas komplexere Modelle zu wählen.
ML	Maximum Likelihood. Ähnlich wie REML, kann aber zu wenig glätten.
P-REML	Pearson-Variante von REML.
P-ML	Pearson-Variante von ML.
GACV.Cp	Generalisierte approximative Kreuzvalidierung.

7.4 Gamma (Glättungsmultiplikator)

Der Parameter gamma multipliziert die effektiven Freiheitsgrade im Glättungskriterium und begünstigt damit glattere (weniger wellige) Anpassungen. Der Vorgabewert ist 1.2.

Wert	Wirkung
1.0	Standardglättung
1.2–1.4	Etwas glatter als die Vorgabe (empfohlen)
2.0+	Deutlich glatter — geeignet für verrauschte Daten oder kleine Stichproben

Höhere Gamma-Werte schützen vor Überanpassung, indem sie Komplexität stärker bestrafen. Die Voreinstellung Earth Pipeline verwendet 1.4, um beim Verfeinern der earth-Knoten zusätzlich zu glätten.

7.5 Kreuzvalidierung

Ist das Kästchen gesetzt (Vorgabe), führt mgcvUI nach der Anpassung eine 10-fache Kreuzvalidierung durch und berechnet ein CV-Bestimmtheitsmaß. Dieses liefert eine ehrliche Schätzung der Prognosegüte außerhalb der Stichprobe.

7.6 Select (Schrumpfung)

Ist das Kästchen gesetzt (Vorgabe), erhält jeder Glättungsterm einen zusätzlichen Strafterm, der ihn auf null schrumpfen kann. Damit wird eine automatische Variablenselektion möglich — unwichtige Glättungsterme werden faktisch aus dem Modell entfernt. Umgesetzt über `select = TRUE` in `gam()`.

7.7 Default Basis (Standardbasis)

Der Typ der Spline-Basisfunktion für Glättungsterme:

Basis	Vollständige Bezeichnung	Beschreibung
tp	Thin-Plate-Regressionsspline	Vorgabe. Isotrop (keine Knotensetzung erforderlich). Gute Allzweckwahl.
cr	Kubischer Regressionsspline	Erlaubt die explizite Setzung von Knoten. Für die Einbindung der earth-Knoten erforderlich.
ps	P-Spline	Gleichmäßig verteilte B-Spline-Basis mit Differenzen-Strafterm. Rechnerisch effizient.
bs	B-Spline	Flexible B-Spline-Basis.

7.8 Default k (Basisdimension)

Die Basisdimension k steuert die maximale Welligkeit jedes Glättungsterms. Der Wert 0 (Vorgabe) bedeutet „automatisch“ — mgcvUI verwendet $k = 10$ oder die Anzahl der earth-Knoten, je nachdem, was angemessen ist.

- **Kleines k (3–5):** Sehr glatte, nahezu lineare Kurven.
- **Mittleres k (8–15):** Mäßige Flexibilität (meist ausreichend).
- **Großes k (20+):** Sehr flexibel — Gefahr der Überanpassung ohne starke Regularisierung.

Hinweis: Die effektiven Freiheitsgrade (EDF) in der Registerkarte Summary zeigen an, wie viele Freiheitsgrade jeder Glättungsterm tatsächlich nutzt. Liegt EDF nahe bei $k - 1$, ist die Basis möglicherweise zu klein — erhöhen Sie k . Ist EDF deutlich kleiner als k , ist der Glättungsterm gut penalisiert.

7.9 Tensorotyp

Für Interaktionen zwischen stetigen Variablen stehen zwei Tensorprodukt-Typen zur Verfügung:

Typ	Funktion	Beschreibung
ti	ti(x1, x2)	Tensor-Interaktion — modelliert nur den Interaktionseffekt; die Haupteffekte werden getrennt durch univariate s() -Terme abgebildet. Wegen der Interpretierbarkeit zu bevorzugen.
te	te(x1, x2)	Tensorprodukt — modelliert den gesamten gemeinsamen Effekt einschließlich der Haupteffekte. Für RCA-Anpassungen schwerer zu zerlegen.

7.10 Zugelassene Interaktionen

Ein aufklappbarer Abschnitt **Allowed Interactions** zeigt eine obere Dreiecksmatrix aus Kontrollkästchen für alle einbezogenen glatten (nicht linearen, nicht faktoriellen) Prädiktoren. Jedes Kästchen aktiviert einen Tensorprodukt-Glättungsterm zwischen dem zugehörigen Variablenpaar.

- **Allow All / Clear All** — Schaltflächen zum Umschalten aller Kästchen.
- **Klick auf einen Variablennamen** — schaltet alle Interaktionen dieser Variable um.
- **Rechtsklick auf einen Variablennamen** — „Block from main effect“ — markiert die Variable zur ausschließlichen Verwendung in Interaktionen (kein univariater Glättungsterm).

Werden earthUI-Knoten importiert, sind die von earth erkannten Interaktionen vorab aktiviert und gesperrt (gelb hinterlegt). Hat das earth-Modell $\text{degree} = 1$ verwendet (keine Interaktionen), erscheint ein entsprechender Hinweis.

Warnung

Das Aktivieren von Tensor-Interaktionen erhöht die Modellkomplexität erheblich. Setzen Sie Interaktionen ein, wenn Fachwissen oder von earth erkannte Belege dafür sprechen. Prüfen Sie, ob die resultierenden Anpassungen plausibel und erklärbar sind.

7.11 Faktor-Glättungsterm-Interaktionen

Sind sowohl Faktor- als auch Glättungsvariablen einbezogen, erscheint eine gesonderte Matrix **Factor-by-Smooth Interactions**. Jedes Kästchen erzeugt einen Term **s(x, by = factor)** — eine eigene glatte Kurve je Faktorstufe.

7.12 Erweiterte Parameter

Ein aufklappbarer Abschnitt „Advanced“ macht weitere Einstellungen zugänglich:

- **Optimizer** — Algorithmus zur Schätzung des Glättungsparameters: `outer/newton` (Vorgabe), `outer/bfgs` oder `efs` (Extended Fellner-Schall).
- **Scale** — Skalenparameter. 0 (Vorgabe) = aus den Daten schätzen. Für Modelle mit fester Skala auf einen bekannten Wert setzen.
- **Discrete covariate method** — Schnelle diskretisierte Anpassung für große Datensätze (`discrete = TRUE`).
- **Threads** — Anzahl paralleler Threads (1–32). Nur wirksam, wenn `discrete = TRUE`.

8 Anpassen des Modells

8.1 Die Schaltfläche Fit

Abschnitt 6 der Seitenleiste enthält die Schaltfläche **Fit MgcV GAM Model**. Ein Klick darauf führt das Modell mit Ihrer aktuellen Konfiguration aus.

8.2 Was während der Anpassung geschieht

Wenn Sie auf Fit klicken, geschieht in mgcvUI Folgendes:

1. **Prüfung der Konfiguration** — es wird geprüft, ob mindestens ein Prädiktor und mindestens 10 vollständige Zeilen vorliegen. Sind weniger als 50 % der Zeilen vollständig, erscheint eine Warnung (mit Angabe der Spalten, die die meisten NAs verursachen).
2. **Aufbereitung der Daten** — Beschränkung auf die für das Modell relevanten Spalten, Entfernen von Zeilen mit NAs, Anwenden von Gewichten (sofern gekennzeichnet), Ausschluss der Bewertungsobjektzeile (im Wertermittlungsmodus), Anwenden der Transformation der Zielgröße.
3. **Aufbau der GAM-Formel** — es werden Glättungsterme `s(x, bs=cr", k=10)` für numerische Prädiktoren, Faktorterm für kategoriale Variablen, lineare Terme für als Linear markierte Variablen und Tensorprodukte für die gewählten Interaktionen gebildet. Earth-Knoten werden als explizite Knotenpositionen in Terme der `cr`-Basis eingebettet.
4. **Behandlung von Sonderfällen** — konstante Variablen werden verworfen, binäre Prädiktoren in Faktoren umgewandelt, k auf die Anzahl der eindeutigen Werte minus 1 begrenzt und die earth-Knoten um Datenquantile ergänzt, wenn mehr Knoten benötigt werden.
5. **Anpassung des GAM** — `mgcv::gam()` wird mit der gebildeten Formel sowie mit `family`, `method`, `gamma`, `select` und den weiteren Parametern aufgerufen.
6. **Kreuzvalidierung** (sofern aktiviert) — 10-fache Kreuzvalidierung zur Berechnung des Bestimmtheitsmaßes außerhalb der Stichprobe.
7. **Aktualisierung aller Ergebnis-Registerkarten** — Summary, Equation, Contribution-Diagramme, Diagnostics und alle übrigen Registerkarten werden gefüllt.

8.3 Anpassung im Hintergrund & Fortschritt

Steht das Paket `callr` zur Verfügung, läuft die Modellanpassung in einem Hintergrundprozess, sodass die Anwendung bedienbar bleibt. Eine dunkel gehaltene, terminalartige Fortschrittsanzeige zeigt:

- einen **Timer** mit den verstrichenen Sekunden bzw. Minuten.
- ein **farbcodiertes Trace-Protokoll**: blau für Status, gelb für den Fortschritt der CV-Faltungen, rot für Fehler, grün für Ergebnisse.
- eine Schaltfläche **Abort**, um den Anpassungsprozess bei Bedarf abubrechen.
- eine Schaltfläche **Close**, die nach Abschluss erscheint.

Ist `callr` nicht verfügbar, läuft die Anpassung synchron (die Anwendung ist bis zum Abschluss blockiert).

Nach erfolgreicher Anpassung erscheint ein weißes Häkchen auf der Schaltfläche Fit, und eine Statuszeile zeigt: „R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X“.

9 Ergebnis-Registerkarten

9.1 Data

Zeigt die importierten Daten als interaktive DataTable mit waagrechtem Bildlauf und 15 Zeilen je Seite. In den Modi Appraisal und Market ist die Registerkarte in eine Tabelle **Subject Property** (Zeile 1) und eine Tabelle **Comparable Sales** (die übrigen Zeilen) unterteilt; im General-Modus wird eine einzige Tabelle angezeigt.

Jede Zelle bleibt einzeilig und wird auf die Spaltenbreite gekürzt, damit breite Freitextfelder (etwa Objektbemerkungen) eine Zeile nicht über die Seite ziehen. Ein einfacher Klick wählt eine Zeile aus; ein **Doppelklick auf eine beliebige Zelle** öffnet ein Fenster mit ihrem vollständigen, ungekürzten Inhalt.

9.2 Equation

Zeigt das angepasste GAM-Modell in zwei Teilen:

Modellgleichung (in MathJax dargestellt):

- Die linke Seite zeigt die Zielgröße mit der zugehörigen Transformation (log, log10) und der Linkfunktion.
- Die rechte Seite zeigt den Wert des Achsenabschnitts, die Glättungsterme als $f_i(\text{Variable})$, die nach Variablen gruppierten Faktortermine mit Anzahl der Stufen sowie die linearen Terme mit Koeffizientenwerten.
- Familie und Verfahren werden oberhalb der Gleichung angezeigt.

Tabelle **Smooth Function Definitions**:

- Formelausdruck des Terms (z. B. `s(living_sqft, bs=cr", k=8)`)
- Variablenname
- Typ (s, te, ti oder linear)
- Typ der Basisfunktion
- Wert von k
- Positionen der earth-Knoten (sofern importiert)

9.3 Summary

Sechs Kennzahlenkarten am oberen Rand:

- **R-squared** — Anteil der erklärten Devianz
- **CV R-squared** — kreuzvalidiertes Bestimmtheitsmaß (bei aktivierter Kreuzvalidierung)
- **Dev. Explained %** — Prozentsatz der erklärten Devianz
- **AIC** — Akaike-Informationskriterium
- **n** — Anzahl der verwendeten Beobachtungen
- **Smooths** — Anzahl der Glättungsterme

Unterhalb der Karten:

- **Tabelle Smooth Terms** — Term, EDF (effektive Freiheitsgrade), Ref.df, F-Statistik, p-Wert
- **Tabelle Parametric Terms** — Term, Schätzwert, Standardfehler, t-Wert, p-Wert

9.4 Variable Importance

Ein waagerechtes Balkendiagramm zeigt die relative Bedeutung jedes Modellterms:

- **Glättungsterme**: nach F-Statistik geordnet (blaue Balken)
- **Parametrische Terme**: nach $|t|$ -Wert geordnet (grüne Balken)

Unter dem Diagramm zeigt eine DataTable Term, Typ, EDF, Statistik und p-Wert, absteigend nach Statistik sortiert.

9.5 Beitragsdiagramme

Jeder Modellterm, der zum Prognosewert beiträgt, besitzt ein interaktives **Plotly**-Beitragsdiagramm, das in einer umrandeten Karte dargestellt wird. Die Registerkarte zeigt Diagramme für alle Termtypen:

9.5.1 Univariate Glättungsterme — $s(x)$

- **Blaue Linie** mit 95 %-Konfidenzband
- **Graue Streupunkte** für die partiellen Residuen (bei nicht log-transformierten Modellen)
- **Rote gestrichelte Senkrechte** an den Positionen der earth-Knoten (sofern earth importiert wurde)
- **Kurzinfo beim Überfahren**: Variablenwert, Beitrag (\$), 95 %-Konfidenzintervall, Rate (Steigung je Einheit)

Bei **log-transformierten Modellen** wird die y-Achse mit der Formel $\bar{y} \times (e^{f(x)} - 1)$ in Geldbeträge zurücktransformiert, sodass die Kurven unmittelbar in Geldeinheiten interpretierbar sind.

Für die Variablen **latitude/longitude** wird die Steigung je 0,001 Grad statt je Grad angegeben, da dies der aussagekräftigere geografische Zuwachs ist.

9.5.2 Faktor-Glättungsterme — $s(x, \text{by} = \text{Faktor})$

Eine farbige Linie je Faktorstufe, in einem gemeinsamen Diagramm überlagert. Nützlich, um zu erkennen, wie der Effekt eines Prädiktors zwischen Gruppen variiert (etwa zwischen verschiedenen Nachbarschaften).

9.5.3 Interaktionsterme — $ti(x_1, x_2)$ und $te(x_1, x_2)$

Eine **Heatmap** zeigt die zweidimensionale Beitragsfläche. Farbskala: rot (negativ) → weiß (null) → blau (positiv). Beim Überfahren werden beide Variablenwerte und der Beitragsbetrag angezeigt.

9.5.4 Parametrische Terme — linear und faktoriell

- **Numerische lineare Terme**: Streudiagramm der Datenpunkte mit überlagerter Anpassungsgeraden
- **Faktorterm**: Balkendiagramm des mittleren Beitrags je Faktorstufe

9.6 Correlation

Eine Heatmap-Matrix der Pearson-Korrelationen zwischen allen numerischen Prädiktoren. Unmittelbar nach dem Datenimport verfügbar — eine Modellanpassung ist nicht erforderlich.

- Farbverlauf: blau (−1) → weiß (0) → rot (+1)
- Korrelationswerte werden in jede Zelle geschrieben
- Nützlich, um bereits vor der Anpassung Concurvity (das glatte Analogon zur Multikollinearität) zu erkennen

9.7 Diagnostics

Zwei Gruppen diagnostischer Darstellungen:

Diagnosetafel (über `gratia::appraise`): Vier Diagramme — Residuen gegen angepasste Werte, Q-Q-Diagramm, Histogramm der Residuen sowie Zielgröße gegen angepasste Werte.

Streudiagramm **Actual vs Predicted**: beobachtete gegen prognostizierte Werte mit einer gestrichelten 45-Grad-Referenzlinie. Bei log-transformierten Modellen werden die Werte auf die ursprüngliche Skala zurücktransformiert.

9.8 RCA Adjustments

Drei Histogramme, die nach der RCA-Berechnung angezeigt werden (nur in den Modi Appraisal/Market):

- **Residual Adjustment %** — Verteilung der Residualanpassungen als Prozentsatz des Verkaufspreises
- **Net Adjustment %** — Verteilung der Nettoanpassungen
- **Gross Adjustment %** — Verteilung der Bruttoanpassungen

Jedes Histogramm nennt in der Unterzeile Mittelwert, Median und Standardabweichung und zeigt gestrichelte Referenzlinien für Mittelwert und Median.

9.9 Anova

Eine kombinierte ANOVA-Tabelle, die parametrische und glatte Terme aus der GAM-Zusammenfassung zusammenführt. Sie zeigt Term, Typ (parametrisch oder glatt) und alle zugehörigen Statistiken.

9.10 Mgcvc Output

Rohe Textausgabe mit:

- Laufzeit der Anpassung
- der GAM-Formel
- der Druckausgabe des Modells
- der vollständigen Ausgabe von `summary(model)`
- Familie und Linkfunktion
- Gesamtmaßen der Concurvity

9.11 Sign Check

Werden earthUI-Knoten importiert, vergleicht diese Registerkarte die Wirkungsrichtung jeder Variablen im earth-Modell mit ihrer Richtung im GAM:

- **earth_direction**: steigend, fallend oder gemischt (auf Basis der Vorzeichen der Hinge-Funktionen)
- **gam_direction**: steigend, fallend oder gemischt (aus der Ableitung des Glättungsterms auf einem Gitter mit 100 Punkten geschätzt)
- **consistent**: TRUE/FALSE
- **warning**: Beschreibung etwaiger Abweichungen

Eine Statuszeile zeigt „All directions consistent“ oder die Anzahl der Variablen mit Abweichungen an.

9.12 Concurvity

Concurvity ist das glatte Analogon zur Multikollinearität. Diese Registerkarte zeigt:

- **Tabelle Overall Concurvity:** Zeilen für die Maße `worst/observed/estimate`, Spalten für jeden Glättungsterm. Farblich codiert: weiß ($<0,5$), gelb ($0,5-0,8$), rot ($>0,8$).
- **Tabelle Pairwise Concurvity:** paarweise Matrix der ungünstigsten Fälle mit derselben Farbcodierung.

Werte oberhalb von etwa 0,8 in der Zeile „worst“ deuten darauf hin, dass sich ein Glättungsterm gut durch die übrigen Glättungsterme annähern lässt, was auf Redundanz hinweist.

9.13 Basis Check

Ausgabe von `mgcv::gam.check()`: Tests der Angemessenheit der Basisdimension für jeden Glättungsterm. Ist das k eines Glättungsterms zu klein, weist der Test dies mit einem niedrigen p-Wert aus und legt eine Erhöhung von k nahe.

10 Herunterladen der Daten

Nach der Anpassung exportiert die Schaltfläche **Download Output (Excel)** (Abschnitt 7 der Seitenleiste) eine Excel-Datei mit Prognosen und Diagnosewerten in den mgcv-Ausgabeordner des aktiven Projekts.

10.1 Ausgabespalten

Spalte	Beschreibung
basis	Rücktransformierter Wert des Achsenabschnitts (für alle Zeilen gleich)
<var>_contribution	Beitrag des jeweiligen Glättungsterms zur Prognose. Bei Log-Modellen proportional auf Geldbeträge aufgeteilt.
est_<target>	Modellprognose (z. B. <code>est_sale_price</code>), gegebenenfalls von der Log-Skala zurücktransformiert
residual	Beobachtet – prognostiziert
calc_residual	Kontrolle: beobachtet – Summe der Beiträge (rücktransformiert)
cqa	Condition-Quality-Appeal-Wert (CQA, Skala 0–10)
residual_sf	Residuum / Wohnfläche (sofern <code>living_area</code> gekennzeichnet ist)
cqa_sf	CQA, berechnet aus der Rangfolge über <code>residual_sf</code>

Im Wertermittlungsmodus sind `residual/cqa/cqa_sf` in Zeile 1 (Bewertungsobjekt) auf NA gesetzt. Ist eine `living_area`-Spalte gekennzeichnet, werden die Zeilen absteigend nach `residual_sf` sortiert.

Nach erfolgreichem Abschluss erscheint ein weißes Häkchen auf der Download-Schaltfläche.

10.2 CQA-Werte

CQA stellt das Residuum jeder Zeile auf einer Skala von 0 bis 10 in eine Rangfolge gegenüber allen anderen:

- **Hoher CQA-Wert (etwa 9–10):** deutlich über der Prognose verkauft — vermutlich überdurchschnittlicher Zustand, überdurchschnittliche Qualität oder Attraktivität
- **Niedriger CQA-Wert (etwa 0–1):** deutlich unter der Prognose verkauft — vermutlich schlechterer Zustand oder Notverkauf
- **CQA um 5:** nahe dem Median

11 RCA-Berechnungen und Download

Der RCA-Ablauf (Residual Constraint Approach) steht in den Modi **Appraisal** und **Market Area Analysis** nach der Modellanpassung zur Verfügung. Das Bewertungsobjekt muss in Zeile 1 stehen.

11.1 Öffnen des RCA-Dialogs

Klicken Sie in Abschnitt 8 der Seitenleiste auf die Schaltfläche **Calculate RCA Adjustments & Download**. Es erscheint ein modaler Dialog mit:

- **Score type:** CQA oder CQA per SF (sofern `living_area` gekennzeichnet ist). Die Variante je Flächeneinheit skaliert das Residuum des Bewertungsobjekts mit der Wohnfläche.
- **Subject CQA Score:** Zulässig sind Werte von 0.00 bis 10.00. Das Feld ist zunächst leer; wurde ein earthUI-Modell importiert, werden der in earthUI verwendete Score-Typ und Wert aus den gemeinsamen Projekteinstellungen übernommen.

Klicken Sie auf **Generate**, um zu berechnen und herunterzuladen. Die Excel-Datei wird in den mgcv-Ausgabeordner des aktiven Projekts geschrieben (`<os>_out_mgcv`).

11.2 Interpolation des CQA-Werts

1. Die CQA-Werte und Residuen der Vergleichsobjekte werden sortiert
2. Eine lineare Interpolation (`approx()`) bildet Ihren CQA-Wert auf ein Residuum ab
3. Wert des Bewertungsobjekts = Modellprognose + interpoliertes Residuum
4. Zeilen mit Gewicht 0 werden so behandelt, als hätten sie dasselbe Residuum wie das Bewertungsobjekt

11.3 Ausgabespalten

Spalte	Beschreibung
<code>subject_value</code>	Modellprognose + interpoliertes Residuum
<code>subject_cqa</code>	Vom Anwender eingegebener CQA-Wert
<code><var>_adjustment</code>	Beitrag des Bewertungsobjekts – Beitrag des Vergleichsobjekts (in Geldeinheiten)
<code>residual_adjustment</code>	Residuum des Bewertungsobjekts – Residuum des Vergleichsobjekts
<code>net_adjustments</code>	Summe aller Anpassungen
<code>gross_adjustments</code>	Summe der Absolutbeträge der Anpassungen
<code>residual_adj_pct</code>	Residualanpassung in % des Verkaufspreises
<code>net_adj_pct</code>	Nettoanpassungen in % des Verkaufspreises
<code>gross_adj_pct</code>	Bruttoanpassungen in % des Verkaufspreises
<code>adjusted_sale_price</code>	Tatsächlicher Verkaufspreis + Nettoanpassungen

Nach erfolgreicher Berechnung erscheint ein weißes Häkchen auf der Schaltfläche.

12 Vergleichswertraster (Sales Comparison Grid)

Das Vergleichswertraster steht ausschließlich im Modus **Appraisal** zur Verfügung, und zwar nach der Berechnung der RCA-Anpassungen. Es erzeugt eine formatierte Excel-Arbeitsmappe, die sich zur Aufnahme in Wertermittlungsberichte eignet.

12.1 Dialog zur Auswahl der Vergleichsobjekte

Ein Klick auf die Schaltfläche **Generate Sales Grid & Download** öffnet einen modalen Dialog, der alle Vergleichsverkäufe auflistet. Die Vergleichsobjekte sind in zwei Gruppen unterteilt:

- **Empfohlene Vergleichsobjekte** (vorab ausgewählt): Bruttoanpassung $< 25\%$ des Verkaufspreises, sortiert nach Verkaufsalter (jüngste zuerst). Höchstens 30 Vergleichsobjekte.
- **Weitere Vergleichsobjekte** (standardmäßig nicht ausgewählt): die übrigen Vergleichsobjekte, sortiert nach dem Prozentsatz der Bruttoanpassung.

Wählen Sie Ihre Vergleichsobjekte aus und klicken Sie dann auf **Generate Sales Grid**, um die Arbeitsmappe zu erstellen.

Erforderliche Special Types: `contract_date` und `living_area` müssen gekennzeichnet sein. **Empfohlene Special Types:** `latitude`, `longitude` und `lot_size` — das Raster funktioniert auch ohne sie, einzelne Felder bleiben dann jedoch leer.

12.2 Aufbau der Arbeitsmappe

Die erzeugte Arbeitsmappe enthält mehrere Blätter mit jeweils drei nebeneinander angeordneten Vergleichsobjekten:

- **Kopfzeilen:** Adressen von Bewertungs- und Vergleichsobjekten, Verkaufspreise, Verkaufsdaten und Vermarktungsdauer
- **Gruppierte Zeilen:** Lage (Breitengrad, Längengrad, Gebiet), Grundstücksgröße (Grundstücksfläche, Grundstücksmaße) und Alter (tatsächliches Alter, effektives Alter) mit Gruppenüberschriften und Unterpunkten
- **Zeilen der Modellvariablen:** eine Zeile je Glättungsterm mit dem Beitrag des Bewertungsobjekts, den Beiträgen der Vergleichsobjekte und den Anpassungsbeträgen
- **Zeilen für Restmerkmale:** leere, entspernte Zellen für Eingaben des Sachverständigen zu Merkmalen, die das Modell nicht erfasst (etwa Zustand, Qualität, Aussicht)
- **Zusammenfassungszeilen:** Nettoanpassung in %, Bruttoanpassung in % und angepasster Verkaufspreis mit aktiven Excel-Formeln

12.3 Blattschutz & Zellen für Restmerkmale

Jedes Blatt ist gegen versehentliche Änderungen geschützt; die Zellen für die Werte der Restmerkmale sind jedoch ausdrücklich entspernt. So können Sachverständige Werte für Merkmale eintragen, die nicht im Modell enthalten sind, während die formelgesteuerten Anpassungsberechnungen erhalten bleiben.

13 Herunterladen von Berichten

Die Berichtserstellung erfolgt in einem zweistufigen Quarto-Ablauf, der den einstufigen Bericht früherer Versionen ersetzt.

13.1 Schritt 1: Generate Quarto Report

Die Schaltfläche **Generate Quarto Report** (verfügbar, sobald das Modell angepasst ist) schreibt ein in sich geschlossenes `.qmd`-Paket in den `mgcv`-Ausgabeordner des aktiven Projekts. Das Paket enthält den Quarto-Quelltext, vorab gerenderte Diagrammdateien (PNG und PDF), eine Datei `report_data.rds` mit allen Tabellen und Modellmetadaten sowie eine Formatvorlage `reference.docx` für die Word-Ausgabe. Da das Paket in sich geschlossen ist, können Sie den `.qmd`-Quelltext vor dem Rendern bearbeiten oder mit anderen Quarto-Dokumenten kombinieren.

Die Diagramme werden bei der Erzeugung des Pakets vorab gerendert; schlägt ein einzelnes Diagramm fehl, wird ein Platzhalterbild geschrieben, und das übrige Paket wird normal erzeugt.

13.2 Schritt 2: Convert Quarto Report

Der Abschnitt **Convert Quarto Report** rendert eine `.qmd`-Datei über `quarto::quarto_render()` in ein oder mehrere Ausgabeformate. Voreingestellt ist das zuletzt erzeugte Paket; über die Schaltfläche **Browse** können Sie jedoch eine beliebige `.qmd`-Datei auswählen. Drei Formate stehen zur Verfügung:

- **HTML** — in sich geschlossenes HTML-Dokument. Funktioniert auf allen Plattformen ohne zusätzliche Software. Empfohlene Vorgabe.
- **Word (.docx)** — gerendert mit der mitgelieferten Formatvorlage `reference.docx`. Geeignet zur Bearbeitung und Weitergabe.
- **PDF** — **erfordert eine LaTeX-Installation** (siehe Systemanforderungen weiter unten). Wird kein LaTeX erkannt, ist die PDF-Option in den Formatkästchen automatisch ausgeblendet.

Die einstufigen Renderer früherer Versionen, `render_gam_report()` und `export_gam_docx()`, stehen für die programmgesteuerte Nutzung weiterhin zur Verfügung.

13.3 Inhalt des Berichts

Der erzeugte Bericht enthält:

- **Datensatz & Modell:** Datendatei, Stichprobenumfang, Familie, Verfahren und Angaben zur Anpassung
- **Ergebniszusammenfassung:** R^2 , $CV\text{-}R^2$, erklärte Devianz, AIC und verwandte Gütemaße
- **Modellgleichung:** Modellgleichung mit f_i -Notation für die Glättungsterme und expliziten Koeffizienten für die linearen Terme, dazu die Definitionen der glatten Funktionen. Lange Gleichungen werden im PDF über mehrere Zeilen umbrochen.
- **Zusammenfassungstabellen:** Glättungsterme (EDF, F, p-Wert) und parametrische Terme (Schätzwert, Standardfehler, t, p-Wert)
- **Aktivierte Interaktionen:** eine Matrix der im Modell enthaltenen Tensor- und Faktor-Glättungsterm-Interaktionen (wird nur angezeigt, wenn Interaktionen vorhanden sind)
- **Variable Importance:** Balkendiagramm, geordnet nach F-Statistik (Glättungsterme) oder $|t|$ (parametrisch)
- **Glatte Effekte** (alle Termtypen):
 - Diagramme univariater Glättungsterme (statische ggplot2-Fassungen der interaktiven Diagramme)

- Mehrlinien-Diagramme der Faktor-Glättungsterme
- Heatmaps der Interaktionen (*ti*- und *te*-Terme)
- Diagramme parametrischer Terme (Streudiagramm für numerische, Balkendiagramm für faktorielle Terme)
- Heatmap der **Prädiktorkorrelationen**
- **Diagnostik:** Residualdiagnostik und Streudiagramm beobachtet gegen prognostiziert
- **mgcv-Zusammenfassung:** die vollständige Textausgabe des angepassten Modells
- **Anhang zur Notation:** ein unnummerierter Schlussabschnitt, der die im Bericht verwendete Notation der Wertbeitragsfunktion erläutert

Nach erfolgreicher Erzeugung erscheint ein weißes Häkchen auf der Schaltfläche Generate Quarto Report.

14 Export von Modellfunktionen

mgcvUI kann die glatten Funktionen des angepassten GAM als eigenständigen Code in mehreren Sprachen exportieren. Jeder Glättungsterm wird auf einem Gitter mit 200 Punkten ausgewertet und als Nachschlagetabelle mit linearer Interpolation exportiert. So lässt sich das Modell außerhalb von R nutzen, ohne von `mgcv` abzuhängen.

Der Funktionsexport erfolgt in der aktuellen Version programmgesteuert. Die exportierenden Hilfsfunktionen nehmen ein angepasstes Ergebnisobjekt entgegen (erzeugt von `fit_gam()`) und schreiben die unten beschriebenen Dateien:

- `generate_r_code()`, `generate_python_code()` und `generate_cpp_code()` — Quelldateien je Sprache
- `export_functions_sqlite()` — SQLite-Datenbank mit Nachschlagetabellen
- `export_functions_zip()` — bündelt eine beliebige Kombination der vier Formate in einem ZIP-Archiv

14.1 R-Funktionen

Datei: `gam_functions.R`

- Kopfkomentare mit Zielvariable, R^2 , $CV\text{-}R^2$ und der Formel.
- Eine Konstante `INTERCEPT`.
- Je Glättungsvariable eine `approxfun()` mit vorab berechneten x/y -Vektoren.
- Koeffizienten der linearen Terme als benannte Konstanten.
- Verwendung: Achsenabschnitt, jedes `g_var(value)` und jedes `coef_var * value` aufsummieren.

14.2 Python-Funktionen

Datei: `gam_functions.py`

- `numpy`-Import.
- Konstante `INTERCEPT`.
- Je Variable: Arrays `_x_var` und `_fx_var` sowie eine Funktion `g_var(x)` auf Basis von `np.interp()`.
- Koeffizienten der linearen Terme als Konstanten `COEF_*`.

14.3 C++-Header

Datei: `gam_functions.hpp`

- Reine Header-Datei mit `#pragma once`.
- Namensraum `gam_functions`.
- Hilfsfunktion `interp()` auf Basis von `std::lower_bound`.
- Je Variable: Inline-Funktion mit statischen `const`-Arrays.
- Koeffizienten der linearen Terme als `constexpr double`.

14.4 SQLite-Datenbank

Datei: `gam_functions.sqlite`

- Tabelle `smooth_grids`: Spalten (`variable`, `x`, `fx`) — die Nachschlagetabellen mit 200 Punkten.
- Tabelle `model_info`: `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.

- Tabelle `linear_terms`: variable, coefficient.
- Tabelle `intercept`: Wert des Achsenabschnitts.

15 Vergleich mit earthUI und glmnetUI

Die Anwendung earth und ihre Post-Processing-Module bilden ein zusammengehöriges Werkzeugpaket für die Regressionsmodellierung. Sie teilen dasselbe Datenformat, dieselben Special-Spaltentypen, denselben RCA-Ablauf und dieselben Demodatensätze, verwenden jedoch unterschiedliche Modellierungs-Engines.

15.1 Wesentliche Unterschiede

Merkmal	mgcvUI	earthUI	glmnetUI
Verfahren	GAM (mgcv)	MARS/Earth (earth)	Elastic Net (glmnet)
Zusammenhänge	Glatt nichtlinear	Stückweise linear	Linear
Variablenselektion	Schrumpfung über <code>select=TRUE</code>	Schrittweise Beschneidung	Lasso-Strafterm
Interaktionen	Tensorprodukte (te/ti)	Hinge-Produkte (Grad 2–3)	Paarweise Produkte
Partielle Effekte	Glatte Kurven mit Konfidenzintervall	g-Funktionen (stückweise)	Lineare Steigungen
Earth-Pipeline	Importiert Knoten aus earth	Quelle der Knoten	Importiert Hinge-Basis
Code-Export	R, Python, C++, SQLite	R, Python, C++, SQLite	—
Berichtsformate	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

Ein Alleinstellungsmerkmal von earthUI verdient Erwähnung: earthUI enthält eine experimentelle Prolog/DCG-Funktion, die strukturierte Objektmerkmale aus MLS-Freitextbemerkungen extrahiert.

15.2 Gemeinsame Funktionen

Alle drei Werkzeuge bieten:

- einen gemeinsamen `regProj`-Projektbaum (Vorgabe `~/regProj`) — Projekte, Eingabedateien und verfahrensspezifische Ausgaben werden von den Schwesternanwendungen gemeinsam genutzt
- CSV- und Excel-Import mit Umwandlung der Spaltennamen in `snake_case`
- Locale-Unterstützung für über 30 Länder mit CSV-Trennzeichen, Dezimaltrennzeichen und Papierformat
- besondere Spaltenkennzeichnungen (`contract_date`, `living_area`, `area`, `latitude`, `longitude` usw.)
- automatische Berechnung von `sale_age` aus dem Stichtag
- Umschalter für Dark/Light Mode (Nord-Theme)
- dauerhafte Speicherung der Einstellungen (SQLite und/oder `localStorage`)
- Korrelations-Heatmap, Variable Importance, Diagnostik
- CQA-Bewertung und RCA-Anpassungsablauf
- Vergleichswertraster mit Auswahl der Vergleichsobjekte und formatierter Excel-Ausgabe
- Kompatibilität mit denselben Demodatensätzen (`Appraisal_1.csv`)

15.3 Empfohlener Arbeitsablauf

- **Beginnen Sie mit earthUI**, um die wichtigen Variablen, nichtlinearen Zusammenhänge und Interaktionen aufzuspüren. Der automatische Modellaufbau von earth liefert den

besten Ausgangspunkt.

- **Verfeinern Sie mit mgcvUI**, wenn Sie glattere partielle Effekte wünschen. Importieren Sie die Knoten von earthUI für einen nahtlosen Übergang.
- **Verwenden Sie glmnetUI**, wenn Sie ein rein lineares Modell mit konsequenter Variablenselektion oder mit Vorzeichenbeschränkungen für die Koeffizienten benötigen.
- **Vergleichen Sie alle drei** an denselben Daten, um zu beurteilen, ob die zusätzliche Komplexität nichtlinearer Modelle gerechtfertigt ist.

16 Demodatensatz: Appraisal_1.csv

16.1 Beschreibung

Der Demodatensatz wird gemeinsam mit earthUI und glmnetUI genutzt. Er liegt im Ordner demo_mls/ des mgcvUI-Quelltextes oder, falls earthUI installiert ist:

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

Die Datei enthält rund 1.500 Verkaufsdatensätze von Wohnimmobilien aus einem simulierten MLS-Export, mit dem Bewertungsobjekt in Zeile 1 und den Vergleichsverkäufen in den übrigen Zeilen. Die Daten bilden Verkäufe von Einfamilienhäusern in einem Markt mit mehreren Gebieten ab, mit einer Bandbreite an Objektgrößen, Baualtern und Lagen.

Es handelt sich nicht um echte Daten, sie beruhen jedoch auf einer realistischen Nachbarschaft in Nordkalifornien. Sämtliche identifizierenden Angaben wurden verändert oder entfernt.

16.2 Spalten

Spalte	Typ	Special Type	Beschreibung
sale_type	character	sale_type	Datensatztyp-Code, der das Bewertungsobjekt von den Vergleichsverkäufen unterscheidet
weight	numeric	weight	Beobachtungsgewicht (0 = von der Anpassung ausschließen)
id	numeric	display_only	Interne Datensatzkennung
property_id	numeric	display_only	MLS-Objektkennung
listing_id	character	display_only	MLS-Angebotsnummer
parcel_number	character	display_only	Flurstücksnummer der Bewertungsbehörde (APN)
street_address	character	display_only	Objektadresse
city_name	character	display_only	Stadt
postal_code	character	display_only	Postleitzahl
county_name	character	display_only	Landkreis
contract_date	Date	contract_date	Kaufvertragsdatum (berechnet sale_age)
sale_age	numeric	—	Tage vom Vertragsdatum bis zum Stichtag (vorab berechnet)
coe_date	Date	display_only	Datum des Abschlusses der Treuhandabwicklung
listing_status	character	display_only	Angebotsstatus (z. B. „Sold“)
sale_price	numeric	(Zielgröße)	Verkaufspreis — Zielvariable
rent	numeric	—	Monatsmiete (für Modelle mit mehreren Zielgrößen)
list_price	numeric	display_only	Angebotspreis
original_list_price	numeric	display_only	Ursprünglicher Angebotspreis
living_sqft	numeric	living_area	Bruttowohnfläche in Quadratfuß
ppsf	numeric	display_only	Preis je Quadratfuß
beds_total	integer	—	Anzahl der Schlafzimmer
baths_total	numeric	—	Gesamtzahl der Bäder (z. B. 2.5 = 2 volle + 1 halbes)
lot_size	numeric	lot_size	Grundstücksgröße in Quadratfuß
area_id	integer	area	MLS-Gebietskennung
area_text	character	display_only	Gebietsname
age	numeric	actual_age	Objektalter in Jahren
year_built	integer	display_only	Baujahr
latitude	numeric	latitude	Breitengrad (für das Modell auf 3 Nachkommastellen gerundet)
longitude	numeric	longitude	Längengrad (für das Modell auf 3 Nachkommastellen gerundet)
latitude6	numeric	display_only	Breitengrad in voller Genauigkeit
longitude6	numeric	display_only	Längengrad in voller Genauigkeit
garage_spaces	integer	—	Anzahl der Garagenstellplätze

Spalte	Typ	Special Type	Beschreibung
<code>fp_count</code>	integer	—	Anzahl der Kamine
<code>no_of_stories</code>	numeric	—	Anzahl der Geschosse
<code>style</code>	character	—	Architekturstil
<code>view</code>	character	—	Art der Aussicht (z. B. „Neighborhood“, „Hills“)
<code>days_on_market</code>	integer	<code>dom</code>	Vermarktungsdauer in Tagen
<code>listing_date</code>	Date	<code>listing_date</code>	Angebotsdatum
<code>sale_concessions</code>	numeric	<code>concessions</code>	Verkäuferzugeständnisse

16.3 Vorgeschlagener Schnelleinstieg

1. Starten Sie die mgcv-Routine: `earthUI::launch_mgcv()`
2. Legen Sie ein neues Projekt (Abschnitt 1) mit dem Purpose **For Appraisal** an
3. Legen Sie `Appraisal_1.csv` in den Eingabeordner des Projekts und wählen Sie die Datei in Abschnitt 2 aus (klicken Sie bei Bedarf auf **Refresh**)
4. Wählen Sie `sale_price` als Zielgröße
5. Setzen Sie Response Transform optional auf **Log (natural)**, um proportionale Anpassungen zu erhalten
6. Weisen Sie die Special Types wie in der Tabelle oben zu
7. Beziehen Sie folgende Prädiktoren ein: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (als Faktor), `age`, `latitude`, `longitude`, `garage_spaces`
8. Behalten Sie die Vorgaben bei: REML, $\gamma = 1.2$, tp-Basis, $k = \text{auto}$, `select = TRUE`
9. Klicken Sie auf **Fit Mgcv GAM Model**
10. Prüfen Sie die Registerkarten Summary, Contribution (Diagramme der Glättungsterme) und Diagnostics
11. Laden Sie die Ausgabe herunter (Schritt 7) und prüfen Sie die CQA-Rangfolge
12. Berechnen Sie die RCA-Anpassungen (Schritt 8) mit einem CQA-Wert von etwa 5.00
13. Erzeugen Sie ein Vergleichswertraster (Schritt 9) mit den empfohlenen Vergleichsobjekten

17 Systemanforderungen & Fehlerbehebung

17.1 Unterstützte Plattformen

mgcvUI läuft unter **macOS**, **Windows** und **Linux**. Die Anwendung wird unter macOS mit RStudio entwickelt und vorrangig getestet. Plattformspezifische Hinweise finden Sie weiter unten.

17.2 R-Version

Erforderlich ist $R \geq 4.1.0$. RStudio Desktop (2023.06 oder neuer) wird dringend empfohlen — es bringt **Quarto** und **pandoc** mit (für das Rendern der Berichte erforderlich) und bietet eine bequeme Umgebung zum Starten der Anwendung.

17.3 Erforderliche R-Pakete

Die folgenden Pakete werden automatisch als Abhängigkeiten installiert:

Paket	Zweck
shiny	Framework für Webanwendungen
mgcv	Anpassung der GAM-Modelle
valengrCore	Gemeinsamer Wertermittlungskern (Special-Rollen, Übernahme aus earthUI)
ggplot2, plotly	Statische und interaktive Beitragsdiagramme
gratia	Extraktion der Glättungsterme und Diagnostik
DT	Interaktive Datentabellen
quarto, rmarkdown	Erzeugung und Rendern der Quarto-Berichte
readxl	Import von Excel-Dateien
shinyFiles	Datei- und Ordnerauswahl
RSQLite, DBI	Einstellungs- und Projektdatenbanken (SQLite)
jsonlite	Serialisierung des Variablenzustands

17.4 Optionale Abhängigkeiten

Komponente	Wann erforderlich
LaTeX (TinyTeX, MiKTeX oder MacTeX)	Nur für die Erzeugung von PDF-Berichten. Wird keine LaTeX-Installation erkannt, ist die PDF-Option im Formatauswahlfeld automatisch ausgeblendet. Installation mit: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	Schriftart Roboto Condensed für ggplot2-Diagramme. Ist sie nicht verfügbar oder besteht keine Internetverbindung, weicht die Anwendung automatisch auf die serifenlose System-schrift aus.
earth ($\geq 5.3.0$)	Nur für die Importpipeline aus earthUI erforderlich. Für den eigenständigen Betrieb nicht nötig.
callr	Modellanpassung im Hintergrund (hält die Anwendung bedienbar). Ohne dieses Paket läuft die Anpassung synchron.
writexl, openxlsx	Excel-Downloads (Modellausgabe, RCA, Vergleichswertraster). Die Anwendung weist auf ein Fehlen hin.
bslib	Bootstrap-5-Theming (Nord Light/Dark).
officer	Programmgesteuerter Word-Export früherer Versionen (<code>export_gam_docx()</code>).

17.5 Plattformspezifische Hinweise

17.5.1 macOS

Die wenigsten Probleme. Für Systembibliotheken wird Homebrew empfohlen. Werden PDF-Berichte benötigt:

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

Funktioniert gut mit RStudio Desktop. Wichtige Hinweise:

- **Firmen- oder gesperrte Rechner:** Ist das Temp-Verzeichnis oder das Benutzerdatenverzeichnis eingeschränkt, können die dauerhafte Speicherung der Einstellungen und die Berichtserzeugung fehlschlagen. Die Anwendung zeigt dann eine Warnung in der R-Konsole an, arbeitet aber ohne Speicherung weiter.
- **LaTeX:** Installieren Sie TinyTeX aus R heraus (gleicher Befehl wie oben für macOS) oder MiKTeX von miktex.org.
- **Lange Dateipfade:** Windows begrenzt Pfade auf 260 Zeichen. Halten Sie den Pfad Ihres Arbeitsverzeichnisses kurz.

17.5.3 Linux

Am uneinheitlichsten zwischen den Distributionen. Unter Ubuntu/Debian sind unter Umständen Systembibliotheken erforderlich, bevor sich die R-Pakete kompilieren lassen:


```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

Für PDF-Berichte: `tinytex::install_tinytex()` oder `texlive-full` über die Paketverwaltung des Systems installieren.

Auf Servern ohne Anzeige (headless) werden interaktive Plotly-Diagramme in Berichten als statische Alternativen gerendert. Die Shiny-Anwendung selbst benötigt eine Verbindung über einen Webbrowser.

17.6 Verhalten bei fehlenden Komponenten

mgcvUI ist so ausgelegt, dass es auch bei fehlenden optionalen Komponenten arbeitet:

Fehlende Komponente	Verhalten
Kein LaTeX	Die PDF-Option ist im Auswahlfeld der Berichtsformate ausgeblendet. HTML und Word bleiben verfügbar.
Kein Internet / Schriftarten fehlen	Roboto Condensed wird durch die serifenlose Systemschrift ersetzt. Eine Konsolenmeldung wird protokolliert.
Kein SQLite / schreibgeschütztes Dateisystem	Einstellungen bleiben zwischen Sitzungen nicht erhalten. Eine Warnung erscheint in der Konsole. Ansonsten läuft die Anwendung normal.
gratia::appraise schlägt fehl	Die Registerkarte Diagnostics und die Berichte zeigen eine Platzhaltermeldung. Verwenden Sie ersatzweise <code>mgcv::gam.check()</code> in der Konsole.
Temp-Verzeichnis nicht beschreibbar	Die Berichtserzeugung schlägt mit einer klaren Fehlermeldung fehl. Abhilfe: Setzen Sie die Umgebungsvariable <code>TMPDIR</code> auf einen beschreibbaren Ort.

17.7 Fehlerbehebung

17.7.1 „PDF option is not available“

Es wurde keine LaTeX-Installation erkannt. Führen Sie in R aus:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

Starten Sie danach die Anwendung neu. Die PDF-Option erscheint dann.

17.7.2 „Settings will not persist“

Das Verzeichnis der SQLite-Datenbank kann nicht angelegt oder beschrieben werden. Prüfen Sie die Berechtigungen für:

- **macOS/Linux:** `~/Library/Application Support/R/mgcvUI/` oder `~/.local/share/R/mgcvUI/`
- **Windows:** `%APPDATA%/R/data/mgcvUI/`

17.7.3 „Diagnostics plot unavailable“

Die Funktion `gratia::appraise()` schlägt bei komplexen Modellen (viele Tensor-Interaktionen, nicht standardmäßige Familien) gelegentlich fehl. Dies ist ein bekanntes Problem der Ursprungsbibliothek. Abhilfe: Führen Sie `mgcv::gam.check(model)` in der R-Konsole aus, um die Diagnostikdiagramme unmittelbar anzuzeigen.

17.7.4 „Report error: number of items to replace...“

Dies war ein bekanntes Problem früherer Versionen im Zusammenhang mit dem Paket `gratia`. Stellen Sie sicher, dass Sie die neueste Fassung von `mgcvUI` installiert haben. Die aktuelle Fassung fängt diesen Fehler ab und erzeugt den Bericht mit einem Platzhalter für den Diagnostikabschnitt.

17.7.5 Port 7880 bereits belegt

Startet die Anwendung nicht, weil Port 7880 (aus einer früheren Sitzung) belegt ist, führen Sie aus:

```
# macOS/Linux:
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 Paketinstallation schlägt unter Linux fehl

Es fehlen Systembibliotheken. Installieren Sie die im Linux-Abschnitt oben aufgeführten Entwicklungshheader und wiederholen Sie `install.packages()`.

Literatur

- Wood, S. N. (2003). „Thin-plate regression splines“. In: *Journal of the Royal Statistical Society (B)* 65.1, S. 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, S. 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, S. 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. Aufl. Chapman und Hall/CRC.
- Wood, S. N., N. Pya und B. Säfken (2016). „Smoothing parameter and model selection for general smooth models (with discussion)“. In: *Journal of the American Statistical Association* 111, S. 1548–1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).

Grundlagen

Überblick

Foundations • The Valuation Engineer

Bert Craytor, Herausgeber • 1. August 2026

Version 1.0.0

DOI: [10.67330/82vssv81](https://doi.org/10.67330/82vssv81)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Foundations veröffentlicht präzise, zitierfähige Darstellungen von Konzepten, die der modernen Bewertungsmethodik zugrunde liegen. Die Einträge greifen Begriffe auf, denen praktizierende Gutachter begegnen — manchmal unter anderen Namen, manchmal ohne sie als solche zu erkennen — und stellen sie in ihren formalen Kontext: woher sie stammen, was sie fachlich bedeuten, wo sie in der Praxis auftreten und warum sie für die Vertretbarkeit von Gutachten wichtig sind.

Die Eröffnungsausgabe etabliert die sechs Kernkonzepte des hedonischen Rahmenwerks, so angeordnet, dass jeder Eintrag auf dem Vokabular der vorangehenden aufbaut. Die Abfolge führt vom *Heterogenen Gut* (jener Eigenschaft von Immobilien, die Bewertungsmethodik überhaupt erst notwendig macht) über den *Merkmalsraum* und die *Hedonische Preisfunktion* (den formalen Apparat, der das Bündel von Attributen ordnet, das eine Immobilie darstellt), weiter über den *Impliziten Preis* und die *Hedonische Regression* (wie der marginale Beitrag jedes Attributs aus Marktdaten zurückgewonnen wird) bis hin zur *Latenten Variable* (was das nicht erklärte Restsignal in jedem hedonischen Modell tatsächlich darstellt).

Ein einziges durchgehendes Beispiel — acht Vergleichsobjekte aus einem Wohnviertel in Pacifica, Kalifornien — zieht sich durch alle sechs Einträge und bildet den pädagogischen roten Faden der Ausgabe.

Warum diese sechs, in dieser Reihenfolge

Die Auswahl der sechs Einträge ist nicht willkürlich. Jedes Konzept ist tragend für die moderne hedonische Methodik, jedes begegnet Gutachtern tatsächlich in der alltäglichen Praxis, und jedes wird oft genug missverstanden oder bleibt unbenannt, sodass eine zitierfähige Fassung praktischen Wert hat.

Die Reihenfolge ist durch konzeptionelle Abhängigkeit vorgegeben. Das *heterogene Gut* ist die Wurzel: Es benennt die Eigenschaft von Immobilien (und Gebrauchtwagen, Arbeitsmärkten und einigen anderen Gütern), die das gesamte hedonische Rahmenwerk notwendig macht. Der *Merkmalsraum* ist der formale Apparat, um heterogene Güter als Vektoren darzustellen; ohne ihn lässt sich nichts Weiteres präzise sagen. Die *hedonische Preisfunktion* ist das Gleichgewichtsobjekt, das Punkte im Merkmalsraum auf Preise abbildet. *Implizite Preise* sind ihre partiellen Ableitungen — die marginalen Beiträge, die jede Anpassung in einem Vergleichsraster (Sales Grid) behauptet, ob der Gutachter sie so benennt oder nicht. Die *hedonische Regression* ist der empirische Schätzer, der implizite Preise aus beobachteten Transaktionen zurückgewinnt; dies ist der Eintrag, in dem Theorie zu Datenanalyse wird. *Latente Variablen* setzen dort an, wo die Residuen der Regression beginnen: die unbeobachteten Attribute, deren Effekte im Residuum als strukturiertes Signal überleben statt als verworfbares Rauschen.

Wer in dieser Reihenfolge liest, sollte am Ende einen durchgehenden Faden besitzen — von „Warum gibt es Bewertungsmethodik überhaupt?“ bis zu „Was ist von den Residuen eines ange-

passten Modells zu halten?“ — eine Reise, die die meisten Gutachter nach Erfahrung des Autors nur stückweise über viele Praxisjahre hinweg zurücklegen.

Das durchgehende Beispiel

Ein einziger Datensatz aus acht Wohn-Vergleichsobjekten in Pacifica, Kalifornien, zieht sich durch alle sechs Einträge:

Comp	GLA (sf)	Grundstück (sf)	Aussicht	Zustand	Verkaufspreis
A	1,650	9,000	ja	gut	\$1,425,000
B	1,650	7,200	nein	gut	\$1,280,000
C	1,850	7,500	nein	gut	\$1,407,500
D	1,450	7,000	nein	gut	\$1,155,000
E	1,700	8,000	ja	ausgezeichnet	\$1,470,000
F	1,550	7,100	nein	durchschnittl.	\$1,177,500
G	1,700	10,500	nein	gut	\$1,392,500
H	1,600	7,300	nein	gut	\$1,312,500

Das Beispiel ist bewusst klein genug, um es von Hand durchdenken zu können, und reich genug, um jedes Konzept zu üben. Eintrag 001 untersucht die Differenz von \$145,000 zwischen Comp A und Comp B als gemeinsamen Beitrag zweier Merkmale (Aussicht und Grundstück), die in einer einzigen Paarvergleichsbeobachtung miteinander verflochten sind. Eintrag 002 platziert alle acht Comps als Punkte in einem vierdimensionalen Merkmalsraum und zeigt, dass naive Distanzmaße sie falsch reihen. Eintrag 003 schlägt eine funktionale Form für die hedonische Preisfunktion dieses Viertels vor. Eintrag 004 gewinnt die impliziten Preise analytisch zurück und zerlegt die ursprüngliche A-gegen-B-Differenz in einen Beitrag von \$127,039 aus Aussicht und Grundstück, wobei \$18,000 unerklärt bleiben. Eintrag 005 schätzt das Modell formal in R und berichtet die vollständige `lm()`-Ausgabe einschließlich Standardfehlern und Residuen. Eintrag 006 untersucht das eine Residuum, das das Modell nicht erklären kann — eine Abweichung von +\$46,455 bei Comp H — und verbindet es mit der Latente-Variablen-Sichtweise, die der andernorts vom Autor entwickelten Methodik des Residual Constraint Approach zugrunde liegt.

Der Datensatz ist im Repositorium der Ausgabe unter `shared/data/pacifica_comps.csv` verfügbar für Leser, die eine der Berechnungen nachvollziehen möchten.

Wie diese Ausgabe zu lesen ist

Wer neu in der hedonischen Methodik ist, sollte die sechs Einträge der Reihe nach lesen. Die Abhängigkeiten sind real: Jeder Eintrag setzt das Vokabular der vorangehenden voraus, und Vorgreifen lädt zu Verwirrung ein, die die Einträge selbst nicht vorwegnehmen können.

Wer bereits mit hedonischer Regression in Berührung gekommen ist — das schließt die meisten Gutachter ein, die einen weiterführenden Statistikkurs besucht, mit einem AVM gearbeitet oder die Literatur zur Massenbewertung gelesen haben — kann die Ausgabe als Auffrischung und Nachschlagewerk nutzen. Die Einträge sind so geschrieben, dass sie atomar zitierbar sind; ein an anderer Stelle der Zeitschrift veröffentlichter Artikel, der auf den *impliziten Preis* verweist, kann per DOI auf Eintrag 004 verlinken, statt das Konzept neu herzuleiten.

Wer sich speziell für die Methodik des Residual Constraint Approach interessiert, sollte zuerst Eintrag 006 lesen, um die Latente-Variablen-Sichtweise zu verstehen, dann zu den Einträgen

003–005 zurückkehren für den hedonischen Apparat, den RCA erweitert, und schließlich zu den Einträgen 001–002 für das grundlegende Vokabular, auf dem alles ruht.

Wer sich für die Frage interessiert, *was eine Bewertungsanpassung vertretbar macht*, wird feststellen, dass die Abschnitte „Warum es für die Vertretbarkeit wichtig ist“ der einzelnen Einträge, in Folge gelesen, einen kurzen Essay über den methodischen Rang anpassungsbasierter Bewertung ergeben. Dieser Essay wird nirgendwo in der Ausgabe explizit zusammengesetzt, ist aber das implizite Thema jedes Eintrags.

Redaktionelle Haltung

Foundations ist keine Lehrbuchrubrik. Jeder Eintrag ist kurz genug, um in einer Sitzung gelesen zu werden, richtet sich an ein Publikum, das mit der Bewertungspraxis bereits vertraut ist, und soll eine präzise Formulierung eines Konzepts liefern statt einer umfassenden Darstellung. Ziel ist ein stabiler Referenzpunkt, den andere Artikel der Zeitschrift zitieren können.

Einträge werden redaktionell geprüft und können einen geladenen Kommentar eines Fachexperten tragen. Sie erscheinen unter einer Creative-Commons-Namensnennungslizenz und erhalten DOIs sowohl über Crossref (als kanonische Zitation) als auch über Zenodo (als versioniertes Archiv). Jeder Eintrag kann im Laufe der Zeit geschärft werden — kleinere Revisionen aktualisieren die Zenodo-Version, größere Revisionen aktualisieren beide DOIs mit einer redaktionellen Notiz zur Änderung. Der Crossref-DOI verweist stets auf die aktuellste Version; bestimmte historische Versionen sind über Zenodo adressierbar.

Künftige Ausgaben werden das *Foundations*-Vokabular in angrenzende Bereiche erweitern: räumliche Abhängigkeit, bedingte Erwartung, Endogenität, Verzerrung durch ausgelassene Variablen, Identifikation sowie den formalen Apparat, der Proxy-Variablen von latenten Variablen unterscheidet. Vorschläge für Konzepte, die eine Behandlung verdienen, sind willkommen und können an den Herausgeber gesandt werden.

Die sechs Einträge dieser Ausgabe:

1. **Heterogenes Gut** — warum Bewertungsmethodik überhaupt existieren muss.
2. **Merkmalsraum** — die Vektordarstellung der Bündelattribute.
3. **Hedonische Preisfunktion** — die Gleichgewichtsabbildung von Merkmalen auf Preise.
4. **Impliziter Preis** — der marginale Marktpreis jedes Merkmals.
5. **Hedonische Regression** — die Schätzung der Funktion aus beobachteten Transaktionen, mit einem vollständig durchgerechneten R-Beispiel.
6. **Latente Variable** — die Rückgewinnung strukturierten Signals aus Residuen.

Literatur.

- 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

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (000-overview.pl) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 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).

```

Heterogenes Gut

Foundations • The Valuation Engineer

Bert Craytor • 1. August 2026

Version 1.0.0

DOI: [10.67330/slscsv65](https://doi.org/10.67330/slscsv65)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Heterogene Güter sind Güter, die nicht austauschbar sind, weil jedes ein eigenes Bündel von Merkmalen darstellt, die Käufer separat bewerten, und deren Merkmale nicht separat gehandelt werden.

Intuitive Einordnung. Ein homogenes Gut ist eines, bei dem jedes beliebige Exemplar genügt: ein Barrel WTI-Rohöl, ein Scheffel Mais der Sorte No. 2 yellow, eine Kilowattstunde aus dem Netz, eine Aktie eines bestimmten Unternehmens. Die Handelsklassen und Spezifikationen hinter diesen Gütern existieren gerade deshalb, um einzelne Exemplare austauschbar zu machen, sodass Preise für die Kategorie gelten und nicht für das einzelne Stück. Ein heterogenes Gut ist eines, bei dem das konkrete Exemplar zählt — *dieses* Haus, nicht *jenes* Haus; *dieser* Gebrauchtwagen, nicht der daneben.

Zwei Eigenschaften machen ein Gut im fachlichen Sinne heterogen. Erstens: Käufer legen Wert auf mehrere unterscheidbare Attribute, die über die Exemplare im Markt hinweg variieren. Zweitens: Diese Attribute sind durch die Produktionstechnologie des Gutes gebündelt; sie sind nicht einzeln käuflich. Man kann nicht 200 Quadratfuß Wohnfläche (GLA) für sich allein kaufen. Man kann keinen Schulbezirk unabhängig von einem Grundstück erwerben. Die Merkmale kommen gebündelt daher, und der Markt setzt Preise für ganze Bündel fest, nicht für die Merkmale selbst.

Immobilien sind das kanonische heterogene Gut und zeigen die Eigenschaft ungewöhnlich stark. Lagemerkmale sind vollkommen unreplizierbar: Keine zwei Parzellen können denselben Standort einnehmen, also besitzt jede Immobilie mindestens ein Attribut (ihre Koordinaten), das sie von jeder anderen Immobilie im Markt unterscheidet. Physische, rechtliche und umweltbezogene Attribute schichten weitere Variationsdimensionen auf einen Vektor, der als Ganzes für jede Immobilie einzigartig ist.

Wo Gutachter dem Konzept begegnen. Jede Anpassung im Vergleichswertverfahren (Sales Comparison Approach) ist ein implizites Eingeständnis, dass das Bewertungsobjekt und die Vergleichsobjekte heterogene Güter sind. Wären Immobilien homogen, wären Verkaufspreise bereits direkt vergleichbar; kein Raster wäre nötig. Das Raster ist ein Versuch, die Preise unterschiedlicher Bündel in Einklang zu bringen, indem für Unterschiede in der Bündelzusammensetzung angepasst wird. Es lohnt festzuhalten, was das Raster *nicht* tut: Es zerlegt einen Verkaufspreis nie in Wertbeiträge einzelner Merkmale. Das traditionelle Raster hat keine Spalte für Wertbeiträge — nur Anpassungen, direkt geschätzt, Merkmal für Merkmal.

Das Sachwertverfahren (Cost Approach) begegnet der Heterogenität, indem es das Bündel aus seinen Komponenten rekonstruiert — Fundament, Rohbau, Ausbau, Abschreibung — und jede separat bepreist. Das Ertragswertverfahren (Income Approach) begegnet ihr, indem es das Bündel zu einem einzigen Cashflow-Profil abstrahiert, wobei Kapitalisierungs- und Renditensätze die Heterogenität auf Bündelebene in einen marktabgeleiteten Multiplikator aufnehmen. Alle drei Bewertungsansätze sind methodische Antworten auf die zugrunde liegende Tatsache, dass Immobilien

ein heterogenes Gut sind. Sie unterscheiden sich darin, *wie* sie mit dieser Tatsache umgehen, nicht darin, ob sie sich ihr stellen.

Gutachter begegnen der Heterogenität auch informell, wann immer sie über Comp-Auswahl, Paarvergleichsanalysen (Paired-Sales) oder „Äpfel-mit-Äpfeln“-Vergleiche nachdenken. Der Instinkt, Vergleichsobjekte zu finden, die in möglichst vielen Attributen ähnlich sind, ist die Praktikerantwort auf Heterogenität: Minimiere die Bündeldistanz zwischen Bewertungsobjekt und Comp, und die unbeobachtete Komponente des Bündels trägt weniger zur Differenz der Verkaufspreise bei.

Warum es für die Vertretbarkeit wichtig ist. Die Heterogenität von Immobilien ist der Grund, warum Bewertungsmethodik als Disziplin existieren muss. Wären Immobilien austauschbar, wäre der Marktwert direkt aus veröffentlichten Preisen ablesbar; Bewertung reduzierte sich auf Nachschlagen. Weil sie es nicht sind, muss der Wert aus den Preisen verwandter, aber nicht identischer Immobilien erschlossen werden, und jeder Inferenzschritt ist eine methodische Entscheidung, die begründet werden muss.

Heterogenität formal anzuerkennen klärt mehrere Praxisfragen, an denen die Vertretbarkeit nach USPAP und den Standards Rules hängt. Comp-Auswahl ist die Wahl von Objekten im Merkmalsraum, die dem Bewertungsobjekt nahe genug liegen, dass die aus ihnen gewonnenen impliziten Preise verallgemeinerbar sind. Die Angemessenheit einer Anpassung hängt davon ab, ob der zugrunde liegende implizite Preis durch Marktdaten stützbar ist. Die Abstimmung zwischen den Ansätzen spiegelt das Urteil des Gutachters darüber, welcher Ansatz die Heterogenität *dieser* Immobilie am glaubwürdigsten bewältigt. Jede dieser Vertretbarkeitsfragen wird schärfer, sobald „heterogenes Gut“ als die zugrunde liegende Bedingung benannt ist.

Durchgerechnetes Bewertungsbeispiel. Zwei jüngere Verkäufe im selben Wohnviertel von Pacifica, Kalifornien:

Attribut	Comp A	Comp B
GLA (sf)	1,650	1,650
Schlafzimmer / Bäder	3 / 2	3 / 2
Grundstück (sf)	9,000	7,200
Aussicht	Ja	Nein
Verkaufspreis	\$1,425,000	\$1,280,000

Unter einer Homogene-Güter-Sichtweise sind dies dieselbe Immobilie: identische GLA, identische Zimmerzahl, dasselbe Viertel. Unter der Heterogene-Güter-Sichtweise sind es unterschiedliche Bündel, die sich in zwei Merkmalen (Grundstück und Aussicht) unterscheiden, welche Käufer separat bewerten. Die Preisdifferenz von \$145,000 ist der *gemeinsame* Beitrag dieser beiden Attribute — zwei implizite Preise, verflochten in einem einzigen Paar.

Die impliziten Preise einzeln zurückzugewinnen erfordert mehr Paare, idealerweise mit Variation jeweils nur eines Attributs. Drei oder vier zusätzliche Verkäufe, die die Kombinationen Aussicht/keine Aussicht und großes/kleines Grundstück aufspannen, beginnen die impliziten Preise separat zu identifizieren. Mit genügend Variation über genügend Comps leistet die hedonische Regression die Entflechtung formal; bei begrenzter Variation ist der Gutachter auf eingegrenztes Paarvergleichsdenken beschränkt, dessen Verlässlichkeit explizit davon abhängt, wie nahe die Bündel einander im Merkmalsraum sind.

Das Beispiel zeigt, warum Heterogenität die Wurzelbedingung ist: Selbst zwei „identische“ Häuser sind nicht identisch, und der methodische Apparat der Bewertung existiert genau dazu, mit der Distanz zwischen ihnen umzugehen.

Literatur.

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.

Querverweise: Merkmalsraum; hedonische Preisfunktion; impliziter Preis.

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (001-heterogeneous-good.pl) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 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).
```

Merkmalsraum

Foundations • The Valuation Engineer

Bert Craytor • 3. August 2026

Version 1.0.0

DOI: [10.67330/3yymh010](https://doi.org/10.67330/3yymh010)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Der Merkmalsraum einer Klasse heterogener Güter ist der Vektorraum, dessen Koordinaten die von Käufern bewerteten Attribute indizieren und in dem jedes Gut durch den Vektor seiner Attributausprägungen dargestellt wird.

Intuitive Einordnung. Der vorangehende Eintrag hat festgehalten, dass ein heterogenes Gut ein eigenes Bündel von Merkmalen ist. Die natürliche Anschlussfrage lautet: Wie sollen wir solche Bündel formal beschreiben und vergleichen? Die Antwort, die auf Lancaster (1966) zurückgeht, besteht darin, jedes Gut als Punkt in einem Vektorraum zu behandeln, dessen Achsen die Merkmale sind, auf die Käufer Wert legen.

In dieser Darstellung ist eine Wohnimmobilie in Pacifica nicht mehr „ein Haus in Park Pacifica“, sondern ein Vektor mit Koordinaten für Wohnfläche (GLA), Grundstücksgröße, Aussicht, Zustand, Schulbezirk, Entfernung zum Ozean und so weiter. Zwei Immobilien, die alle diese Attributwerte teilen, würden denselben Punkt im Merkmalsraum einnehmen; Immobilien, die sich in irgendeinem Attribut unterscheiden, besetzen verschiedene Punkte. Die Geometrie des Raums kodiert, wie nah oder wie verschieden zwei Immobilien sind — und zwar so, dass alle Dimensionen erfasst werden, in denen Käufer Unterscheidungen treffen.

Die Dimensionalität des Raums hängt davon ab, welche Attribute für den untersuchten Markt relevant sind. Für ein typisches Wohnimmobiliengutachten mag der relevante Raum ein Dutzend Dimensionen haben: physische (GLA, Grundstück, Schlafzimmer, Bäder, Alter, Qualität, Zustand), lagebezogene (Schulbezirk, Aussicht, Nähe zu Annehmlichkeiten oder Störquellen) und rechtlich-ökonomische (Bauplanungsrecht, Dienstbarkeiten). Für ein Gewerbe-gutachten verschieben sich die relevanten Dimensionen zu Cashflow-Attributen (Mietvertragsstruktur, Mieterbonität, Belegung), aber die zugrunde liegende geometrische Idee ist dieselbe.

Was den Merkmalsraum nützlich macht, ist, dass er eine qualitative Praktikerintuition („diese Comps sind dem Bewertungsobjekt ähnlich“) in eine quantitative Behauptung verwandelt („diese Comps liegen dem Bewertungsobjekt im Merkmalsraum nahe“). Sobald Distanz definiert ist, lässt sich jede Entscheidung im Vergleichswertverfahren als Aussage über Lage im Raum untersuchen.

Wo Gutachter dem Konzept begegnen. Gutachter arbeiten ständig im Merkmalsraum, meist ohne ihn so zu nennen. Jede Zeile eines Vergleichsrasters ist eine Koordinatenachse; jeder Comp ist ein Zeilenvektor von Koordinatenwerten; jede Anpassung ist der Versuch, sich entlang einer Achse zu bewegen und die anderen festzuhalten.

Die acht Pacifica-Vergleichsobjekte, die wir in dieser Ausgabe durchgehend verwenden, illustrieren die Konstruktion direkt. Jeder Comp ist ein Punkt in einem vierdimensionalen Merkmalsraum, aufgespannt von GLA, Grundstücksgröße, Aussicht und Zustand:

Comp	GLA (sf)	Grundstück (sf)	Aussicht	Zustand	Verkaufspreis
A	1,650	9,000	ja	gut	\$1,425,000
B	1,650	7,200	nein	gut	\$1,280,000
C	1,850	7,500	nein	gut	\$1,407,500
D	1,450	7,000	nein	gut	\$1,155,000
E	1,700	8,000	ja	ausgezeichnet	\$1,470,000
F	1,550	7,100	nein	durchschnittl.	\$1,177,500
G	1,700	10,500	nein	gut	\$1,392,500
H	1,600	7,300	nein	gut	\$1,312,500

Comp A und Comp B aus dem vorangehenden Eintrag sind die ersten beiden Zeilen; die übrigen sechs erweitern das Beispiel auf Weisen, die wir im Rest der Ausgabe ausnutzen werden.

Jeder Comp ist ein Punkt in $\mathbb{R}^4$:

- Comp A liegt bei (1650, 9000, 1, 2).
- Comp B liegt bei (1650, 7200, 0, 2).
- Comp G liegt bei (1700, 10500, 0, 2).

Die Koordinaten für Aussicht und Zustand sind numerisch kodiert (Aussicht als $\{0, 1\}$, Zustand als $\{1, 2, 3\}$); der Preis ist *keine* Koordinate des Merkmalsraums, sondern eine Funktion der Position darin.

Warum es für die Vertretbarkeit wichtig ist. Das Merkmalsraum-Rahmenwerk schärft, was Comp-Auswahl eigentlich ist. Ein Vergleichsobjekt ist nicht bloß „eine ähnliche Immobilie“, sondern ein Punkt in der relevanten Region des Merkmalsraums nahe dem Bewertungsobjekt. Je näher ein Comp dem Bewertungsobjekt im Raum liegt, desto kleiner sind die erforderlichen Anpassungen und desto verlässlicher wird der angezeigte Wert.

Diese Umformung legt mehrere Praxisfragen offen, die das informelle Raster verdeckt:

- *Welche Dimensionen zählen?* Die Wahl der Achsen ist eine Modellierungsentscheidung. „Aussicht“ als Achse aufzunehmen ist nur sinnvoll, wenn der Markt Aussicht eigenständig bepreist; eine „Feng-Shui-Ausrichtung“ aufzunehmen ist nur sinnvoll, wenn ein Marktsegment sie bepreist. Vertretbare Bewertung erfordert, die Achsen zu verteidigen, nicht nur die Werte entlang der Achsen.
- *Welches Distanzmaß gilt?* Zwei Comps können sich um \$200,000 im Preis unterscheiden, weil sie im Merkmalsraum weit auseinanderliegen, aber die Distanzberechnung erfordert eine Skalierung für jede Achse. Untransformierte euklidische Distanz behandelt einen Quadratfuß GLA als äquivalent zu einem Quadratfuß Grundstück, was der Markt nicht tut.
- *Liegen die Comps in einer Region, in der sich die Preisfunktion gutartig verhält?* Sitzt das Bewertungsobjekt in einer dünn besetzten Ecke des Merkmalsraums, liegt kein Comp in seiner Nähe, und die Wertindikation ist notwendigerweise extrapolativ.

Durchgerechnetes Bewertungsbeispiel. Die acht Pacifica-Comps besetzen eine relativ kompakte Region des vierdimensionalen Merkmalsraums. Betrachten wir die Frage, welcher Comp einem hypothetischen Bewertungsobjekt bei (1,700, 7,500, 1, 2) — 1,700 sf, 7,500 sf Grundstück, Aussicht, Zustand „gut“ — am „nächsten“ liegt.

Verwenden wir naiv die euklidische Distanz mit unskalierten Achsen, dominiert die Grundstücksdimension, weil ihre Werte (7,000–10,500) numerisch viel größer sind als Aussicht (0–1) oder Zustand (1–3). Distanzen zum Bewertungsobjekt:

Comp	Distanz (unskaliert, euklidisch)	Anmerkungen
A	1,501	Aussicht passt; Grundstück dominiert
B	308	Aussicht differiert; Grundstück nah
C	230	Aussicht differiert; Grundstück nah
D	559	Aussicht differiert; GLA abweichend
E	502	Aussicht passt; Zustand differiert
F	422	Aussicht und Zustand differieren
G	3,000	Aussicht differiert; Grundstück weit entfernt
H	224	Aussicht differiert; am nächsten von allen

Nach Rohdistanz ist Comp H der nächste Nachbar. Aber Comp H hat keine Aussicht, das Bewertungsobjekt schon. Das naive Maß hat uns in die Irre geführt: Die numerische Skala der Grundstücksdimension hat die kategoriale Aussichtsdimension verdrängt, die für den Preis tatsächlich mehr zählt.

Ein vertretbares Distanzmaß würde jede Achse mit ihrem impliziten Preis skalieren — grob gesagt damit, wie viele Dollar eine Einheit Variation entlang dieser Achse wert ist. Sind die Achsen erst nach impliziten Preisen skaliert, wird Aussicht (im Wert von ~\$100,000) nicht mehr vom Grundstück (~\$25 pro Quadratfuß) dominiert. Comp A, der beim Merkmal Aussicht mit dem Bewertungsobjekt übereinstimmt, tritt dann als das geeignete primäre Vergleichsobjekt hervor.

Genau das liefert die Maschinerie der impliziten Preise, die in den folgenden Einträgen entwickelt wird: eine prinzipiengeleitete Skalierung des Merkmalsraums, die „Nähe“ von einer geometrischen Vermutung in eine marktabgeleitete Berechnung verwandelt.

Literatur.

- 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.

Querverweise: heterogenes Gut; hedonische Preisfunktion; impliziter Preis.

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (`002-characteristics-space.pl`) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 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).

```


Hedonische Preisfunktion

Foundations • The Valuation Engineer

Bert Craytor • 1. August 2026

Version 1.0.0

DOI: [10.67330/9x1a0489](https://doi.org/10.67330/9x1a0489)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Die hedonische Preisfunktion ist die Gleichgewichtsabbildung von Punkten im Merkmalsraum eines heterogenen Gutes auf Transaktionspreise, gebildet als Einhüllende der Gebotsfunktionen der Käufer und der Angebotsfunktionen der Verkäufer in einem Wettbewerbsmarkt.

Intuitive Einordnung. Wir haben nun zwei Zutaten: ein heterogenes Gut (ein Bündel von Merkmalen) und einen Merkmalsraum (einen Vektorraum, in dem jedes Bündel ein Punkt ist). Die hedonische Preisfunktion verbindet beide. Sie ist die Funktion $p(\mathbf{z})$, die jedem Punkt $\mathbf{z}$ im Merkmalsraum einen Preis zuordnet und uns sagt, was der Markt für ein Bündel mit genau dieser Attributkombination zahlt.

Die zentrale Einsicht von Rosen (1974) ist, dass $p(\mathbf{z})$ von keinem einzelnen Käufer oder Verkäufer gewählt wird. Sie entsteht aus dem Marktgleichgewicht. Jeder Käufer hat eine *Gebotsfunktion*, die das Maximum ausdrückt, das er für ein beliebiges Bündel zahlen würde, gegeben seine Präferenzen und sein Einkommen. Jeder Verkäufer hat eine *Angebotsfunktion*, die das Minimum ausdrückt, das er akzeptieren würde, gegeben seine Kosten und Alternativen. Die hedonische Preisfunktion ist die Einhüllende dieser Gebote und Angebote — die Grenzlinie, entlang derer Käufer und Verkäufer tatsächlich abschließen.

Für Immobilien ist $p(\mathbf{z})$ die Antwort auf die Frage: Wenn man ein vollständiges Merkmalsbündel spezifizierte, welchen Preis würde der Markt dafür hervorbringen? Die Antwort ist nicht willkürlich. Sie spiegelt die aggregierten Präferenzen aller Käufer im Markt und die aggregierten Kosten und Alternativen aller Verkäufer wider, vermittelt durch die Gleichgewichtszuordnung von Käufern zu Bündeln.

Die hedonische Preisfunktion ist das zentrale theoretische Objekt der hedonischen Analyse. Implizite Preise sind ihre partiellen Ableitungen. Die hedonische Regression ist ihre empirische Schätzung. Die Vertretbarkeit jeder Anpassung im Vergleichswertverfahren reduziert sich letztlich auf eine Behauptung über diese Funktion.

Wo Gutachter dem Konzept begegnen. Gutachter denken fortwährend über $p(\mathbf{z})$ nach, wiederum meist ohne es so zu nennen. Mehrere Praxistätigkeiten sind implizite Manipulationen der hedonischen Preisfunktion:

Anpassungen im Vergleichswertverfahren. Wenn ein Gutachter Comp B um \$100,000 nach oben anpasst, um zu berücksichtigen, dass das Bewertungsobjekt eine Aussicht hat und Comp B nicht, behauptet der Gutachter etwas über $p(\mathbf{z})$: nämlich dass die Funktion, bei festgehaltenen übrigen Merkmalen, um \$100,000 steigt, wenn das Aussichtsmerkmal von 0 auf 1 wechselt. Das ist eine Aussage über die lokale Steigung von p in der Aussichtsdimension.

Abstimmung zwischen den Ansätzen. Das Sachwertverfahren konstruiert eine Wertindikation von unten, indem es die Komponentenkosten eines replizierten Bündels summiert. Das Ertragswertverfahren leitet den Wert aus Cashflows ab, die im Gleichgewicht den Preis von Bündeln widerspiegeln, die ähnliche Cashflows erzeugen. Das Vergleichswertverfahren liest den Preis direkt

aus beobachteten Transaktionen ähnlicher Bündel ab. In einem gut funktionierenden Markt sollten alle drei Ansätze auf denselben Punkt von $p(\mathbf{z})$ konvergieren. Wenn sie auseinanderlaufen, muss der Gutachter beurteilen, welcher Ansatz die Funktion für die Lage des Bewertungsobjekts im Merkmalsraum am verlässlichsten gemessen hat.

Marktanalyse und Zeitanpassungen. Die hedonische Preisfunktion verschiebt sich im Zeitverlauf, wenn Angebot, Nachfrage, Finanzierung und Erwartungen sich ändern. Eine jährliche Wertsteigerungsrate von 6 Prozent ist eine Behauptung darüber, wie sich $p(\mathbf{z})$ als Ganzes verschoben hat; der Befund, dass Immobilien mit Aussichtsprämie schneller an Wert gewonnen haben als Immobilien ohne Aussicht, ist eine Behauptung, dass die Funktion sich über den Merkmalsraum hinweg ungleichmäßig verschoben hat.

Warum es für die Vertretbarkeit wichtig ist. Die hedonische Preisfunktion explizit zu benennen klärt mehrere grundlegende Vertretbarkeitsfragen:

Existiert die Funktion? Das Rosen-Rahmenwerk unterstellt einen Wettbewerbsmarkt mit genügend Transaktionen, damit eine Gleichgewichtsfunktion sinnvoll ist. In dünnen Märkten, in sich schnell verändernden Märkten oder in Teilmärkten mit idiosynkratischen Käuferpools kann die Funktion schlecht definiert oder instabil sein. Ein Gutachter, der in einem solchen Markt arbeitet, sollte anerkennen, dass die Inferenzmaschinerie auf einem wackligen theoretischen Fundament operiert.

Ist die Funktion überall dieselbe? Rosens Rahmenwerk liefert eine Funktion pro Marktsegment, nicht eine Funktion global. Pacifica und San Francisco sind unterschiedliche Märkte mit unterschiedlichen hedonischen Funktionen; Einfamilienhäuser und Eigentumswohnungen sind selbst innerhalb einer Stadt unterschiedliche Teilmärkte. Ein Gutachter, der Segmente vermengt, unterstellt implizit, dass eine einzige Funktion beide überspannt, was meist falsch ist.

Welche funktionale Form hat p ? Die Funktion muss weder linear noch log-linear noch von irgendeiner anderen bestimmten Form sein. Rosen zeigte nur, dass p die Einhüllende von Geboten und Angeboten ist; ihre Gestalt spezifizierte er nicht. Empirisch zeigt p Nichtlinearitäten (der Grenzwert eines zusätzlichen Schlafzimmers ist oft nicht konstant in der GLA), Interaktionen (die Aussichtsprämie hängt von der Grundstücksausrichtung ab) und Schwellenwerte (eine Küche unterhalb einer gewissen Qualitätsschwelle kann einen Abschlag anziehen, der zu ihrem Zustandswert außer Verhältnis steht). Vertretbare Schätzung erfordert entweder eine aus dem Markt stützbar funktionale Form oder einen flexiblen Schätzer, der die Gestalt aus den Daten zurückgewinnen kann.

Durchgerechnetes Bewertungsbeispiel. Für das untersuchte Viertel in Pacifica arbeiten wir mit einer einfachen additiven funktionalen Form, die der nächste Eintrag ausführlicher durchexerzieren wird:

$$p(\mathbf{z}) = \beta_0 + \beta_{\text{GLA}} z_{\text{GLA}} + \beta_{\text{lot}} z_{\text{lot}} + \beta_{\text{view}} z_{\text{view}} + \beta_{\text{cond}} z_{\text{cond}}.$$

Diese Form enthält zwei starke Annahmen: dass der Beitrag jedes Merkmals linear in seiner Ausprägung ist und dass die Merkmale nicht interagieren. Beide Annahmen sind testbar und werden häufig verletzt; wir übernehmen sie hier als Ausgangspunkt.

Für die acht Comps des durchgehenden Beispiels dieser Ausgabe wird unsere Analyse in den nächsten Einträgen Werte von β schätzen, die eine Funktion etwa der folgenden Gestalt ergeben:

Parameter	Geschätzter Wert
β_0 (Achsenabschnitt)	\$80,777
β_{GLA} (pro sf)	\$575
β_{lot} (pro sf)	\$22.70
β_{view} (ja/nein)	\$86,179
β_{cond} (pro Stufe)	\$49,824

In die additive Form eingesetzt, sagt diese Funktion einen Preis für jeden Punkt des vierdimensionalen Merkmalsraums voraus. Setzt man die Koordinaten von Comp A ein, ergibt sich ein prognostizierter Preis nahe seinem beobachteten Wert von \$1,425,000; setzt man das hypothetische Bewertungsobjekt bei (1700, 7500, 1, 2) ein, ergibt sich eine Wertindikation für diese Immobilie.

Der Kern dieses Eintrags sind nicht diese konkreten Zahlen, sondern der konzeptionelle Schritt: *Die Funktion existiert im Prinzip, bevor wir sie schätzen.* Die Funktion ist eine Eigenschaft des Marktes. Unsere Schätzung ist ein Versuch, sie zurückzugewinnen. Diese Rückgewinnung kann unvollständig, verzerrt oder verrauscht sein, und die nächsten drei Einträge handeln davon, wie man sie verantwortungsvoll durchführt.

Literatur.

- 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.

Querverweise: heterogenes Gut; Merkmalsraum; impliziter Preis; hedonische Regression.

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (003-hedonic-price-function.pl) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 003-hedonic-price-function.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 003
%% Title: Hedonic Price Function
%% Author: Bert Craytor
%% Concept: hedonic_price_function
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(hedonic_price_function, "Hedonic Price Function", economic_theory,
```

```

%      "The hedonic price function is the equilibrium mapping from
%      points in the characteristics space of a heterogeneous good to
%      transaction prices, formed as the envelope of buyer bid
%      functions and seller offer functions in a competitive market.",
%      published, 'dictionary-2026.1').
%
% term(bid_function, "Bid Function", economic_theory, ...).
% term(offer_function, "Offer Function", economic_theory, ...).
% term(sales_comparison_approach, ...).
% term(cost_approach, ...).
% term(income_approach, ...).
%
% formalized_by(hedonic_price_function, rosen_1974).
% depends_on(hedonic_price_function, heterogeneous_good).
% depends_on(hedonic_price_function, characteristics_space).
% -----

:- use_terms([ hedonic_price_function, heterogeneous_good,
               characteristics_space, implicit_price, hedonic_regression,
               bid_function, offer_function,
               sales_comparison_approach, cost_approach, income_approach,
               adjustment, reconciliation ]).
:- use_relations([related/2, depends_on/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_price_function).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.003').

% =====
% LOCAL VOCABULARY
% =====

local_term(equilibrium).
local_term(envelope).
local_term(functional_form).
local_term(market_segment).
local_term(linearity).
local_term(separability).
local_term(time_adjustment).
local_term(coefficient).
local_term(intercept).

% =====
% FORMAL ATTRIBUTION
% =====

attributed_formalization(hedonic_price_function, rosen_1974).

% =====
% CONSTRUCTION: ENVELOPE OF BID AND OFFER FUNCTIONS
% =====
%
% The Rosen (1974) construction:  $p(z)$  is the envelope of buyer bid
% functions and seller offer functions.

```

```

primitive_of_construction(bid_function, buyer).
primitive_of_construction(offer_function, seller).

constructed_as_envelope_of(hedonic_price_function,
                           [bid_function, offer_function]).

emerges_from(hedonic_price_function, market_equilibrium).

% Property that flows from the construction: the function is not
% chosen by any single market participant.
property(hedonic_price_function, not_set_by_individual_participant).
property(hedonic_price_function, equilibrium_object).
property(hedonic_price_function, defined_per_market_segment).

% =====
% APPRAISER ENCOUNTERS - IMPLICIT MANIPULATIONS OF p(z)
% =====

implicit_manipulation_of_hedonic(
    adjustment,
    sales_comparison_approach,
    "Asserts a local slope of p in the relevant characteristic dimension.").
implicit_manipulation_of_hedonic(
    reconciliation,
    multi_approach_valuation,
    "Three approaches should converge on the same point of p; divergence requires editorial
    ↪ judgment.").
implicit_manipulation_of_hedonic(
    time_adjustment,
    market_analysis,
    "p shifts over time; appreciation rates are statements about that shift.").

% =====
% DEFENSIBILITY QUESTIONS
% =====
%
% The article names three questions the hedonic-price-function framing
% sharpens.

defensibility_question(function_existence,
    "In thin or rapidly-changing markets, the equilibrium function may be poorly defined; the
    ↪ inferential machinery operates on a shaky foundation.").
defensibility_question(segment_specificity,
    "One function per market segment, not one function globally. Conflating segments is usually
    ↪ wrong.").
defensibility_question(functional_form,
    "Rosen showed only that p is an envelope, not its shape. Empirically nonlinear with
    ↪ interactions and thresholds; estimation requires defending the form.").

% =====
% WORKED EXAMPLE: ADDITIVE LINEAR SPECIFICATION
% =====
%
% For the Pacifica neighborhood, the article adopts the additive linear
% specification:
%  $p(z) = \beta_0 + \beta_{\text{gla}} * z_{\text{gla}} + \beta_{\text{lot}} * z_{\text{lot}} + \beta_{\text{view}} * z_{\text{view}} + \beta_{\text{cond}} * z_{\text{cond}}$ 
%
% The next entry (004) takes the coefficients as implicit prices; entry
% 005 estimates them by hedonic regression.

```

```

:- consult('../..//kb/pacifica_comps').

specification_form(additive_linear).
specification_assumption(linearity_in_levels,
    "Contribution of each characteristic is linear in its level.").
specification_assumption(no_interactions,
    "Characteristics do not interact in the price function.").
specification_caveat(
    "Both assumptions are testable and frequently violated; adopted here as a starting point.").

% Coefficient values reported in the article (estimated in entry 005,
% asserted here so this entry's worked example can be queried).
coefficient(intercept,      80777.07).
coefficient(beta_gla,       574.93).
coefficient(beta_lot,       22.70).
coefficient(beta_view,      86179.03).
coefficient(beta_cond,      49824.33).

%! predicted_price(?Property, -Predicted) is det.
%
% The fitted additive-linear hedonic price function applied to a
% property's characteristics. Symbolic view/cond are mapped to {0,1}
% and {1,2,3} respectively.

predicted_price(P, Predicted) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond),
    coefficient(intercept, B0),
    coefficient(beta_gla, Bg),
    coefficient(beta_lot, Bl),
    coefficient(beta_view, Bv),
    coefficient(beta_cond, Bc),
    Predicted is B0 + Bg * GLA + Bl * Lot + Bv * View + Bc * Cond.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V
    ; A == view, V == 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).
```

Impliziter Preis

Foundations • The Valuation Engineer

Bert Craytor • 2. August 2026

Version 1.0.0

DOI: [10.67330/md6wke17](https://doi.org/10.67330/md6wke17)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Der implizite Preis eines Merkmals ist die partielle Ableitung der hedonischen Preisfunktion nach diesem Merkmal, ausgewertet an einem Punkt des Merkmalsraums; er stellt den marginalen Marktpreis einer zusätzlichen Einheit des Merkmals dar, bei festgehaltenen übrigen Merkmalen.

Intuitive Einordnung. Die hedonische Preisfunktion $p(\mathbf{z})$ ordnet jedem Bündel einen Preis zu. Die meisten praktischen Fragen der Bewertung betreffen jedoch nicht den Preis eines vollständigen Bündels, sondern den Preis einer bestimmten *Komponente* eines Bündels. Wie viel ist die Aussicht wert? Wie viel bringen hundert zusätzliche Quadratfuß? Wie viel trägt der Unterschied zwischen dem Zustand „gut“ und „ausgezeichnet“ bei?

Der implizite Preis beantwortet diese Fragen. Der Begriff geht auf Rosen (1974) zurück, der zeigte, dass die beobachteten Preise differenzierter Produkte zusammen mit ihren Merkmalen „eine Menge impliziter oder ‘hedonischer’ Preise definieren“; Lancaster (1966) lieferte die zugrunde liegende Sicht von Gütern als Merkmalsbündeln. Für jedes Merkmal z_i ist der implizite Preis die Rate, mit der sich $p(\mathbf{z})$ ändert, wenn z_i variiert und alle übrigen Merkmale festgehalten werden:

$$\pi_i(\mathbf{z}) = \frac{\partial p(\mathbf{z})}{\partial z_i}.$$

Der implizite Preis hat die Einheit Dollar pro Einheit des Merkmals — Dollar pro Quadratfuß GLA, Dollar pro Quadratfuß Grundstück, Dollar pro Stufe des Zustands. Für ein binäres Merkmal wie Aussicht ist der implizite Preis die volle Prämie, die für den Wechsel von keine Aussicht (0) zu Aussicht (1) gezahlt wird.

Das Wort „implizit“ erfasst den zentralen Sachverhalt: Diese Preise werden *nicht* ausgeschrieben. Niemand bietet an, „einen Quadratfuß GLA“ für \$575 zu verkaufen. Der Preis steckt implizit in den ausgeschriebenem Gleichgewichts-Bündelpreisen, und der hedonische Apparat existiert, um ihn zu extrahieren.

Implizite Preise sind der bewertungsrelevante Output des hedonischen Rahmenwerks. Die Anpassung in einem Vergleichsraster ist im Prinzip eine Schätzung eines impliziten Preises multipliziert mit einer Merkmalsdifferenz. Jede Anpassung ist eine Wette auf den Wert von π_i am relevanten Punkt des Merkmalsraums.

Wo Gutachter dem Konzept begegnen. Das Anpassungsraster des Vergleichswertverfahrens ist eine Tabelle von Anwendungen impliziter Preise, auch wenn Gutachter sie nicht so nennen. Ein Rastereintrag, der einen Comp um „+\$10,000 für 200 sf weniger GLA als das Bewertungsobjekt“ anpasst, behauptet, dass der implizite Preis der GLA an der relevanten Stelle des Merkmalsraums \$50 pro Quadratfuß beträgt. Ein Rastereintrag, der „\$100,000 für unterlegene Aussicht“ anpasst, behauptet einen impliziten Preis von \$100,000 für das binäre Aussichtsmerkmal. Die gesamte Vertretbarkeit des Vergleichswertverfahrens ruht darauf, ob diese impliziten Preise durch den Markt

gestützt werden können. In der Praxis schätzen Gutachter Anpassungen jedoch direkt — aus Matched Pairs, aus Erfahrung oder aus der üblichen Praxis der Kollegen — ohne je die zugrunde liegenden Wertbeiträge zu berechnen; das traditionelle Raster verzeichnet nur die Anpassungen selbst. Die Lesart über implizite Preise ist eine Interpretation dessen, was diese Anpassungen behaupten, keine Beschreibung, wie sie hergeleitet werden.

Diese Benennung klärt auch, warum die Paarvergleichsanalyse die grundlegende Technik der Anpassungsherleitung ist. Zwei Comps, die sich in genau einem Merkmal unterscheiden, liefern eine direkte Schätzung des impliziten Preises dieses Merkmals: die Preisdifferenz geteilt durch die Merkmalsdifferenz. Die Verlässlichkeit der Paarvergleichsanalyse hängt davon ab, ob die beiden Comps einander im Merkmalsraum wirklich nahe genug sind, dass der aus ihnen gewonnene implizite Preis auf das Bewertungsobjekt verallgemeinerbar ist.

Drei Eigenschaften impliziter Preise, die für die Praxis zählen:

Sie sind lokal. Der implizite Preis der GLA am 1,650-sf-Punkt ist nicht notwendigerweise derselbe wie am 2,500-sf-Punkt. Die hedonische Preisfunktion ist im Allgemeinen nichtlinear, also variieren ihre partiellen Ableitungen mit der Lage im Merkmalsraum. Eine Anpassung, die aus einer Paarvergleichsanalyse bei 1,500 sf hergeleitet wurde, gilt nicht notwendigerweise für ein Bewertungsobjekt mit 3,000 sf.

Sie sind konditional. Der implizite Preis der Aussicht in einem Strandlagen-Teilmarkt ist nicht derselbe wie in einem Binnenland-Teilmarkt. Die partielle Ableitung wird bei festgehaltenen übrigen Merkmalen gebildet, aber der Wert dieser Ableitung hängt davon ab, *wo* die übrigen Merkmale festgehalten werden.

Sie sind marktabhängig. Implizite Preise sind Gleichgewichtspreise, sie spiegeln also die Präferenzen des Käuferpools zum Zeitpunkt der Transaktionen wider. Eine Marktverschiebung kann implizite Preise ändern, auch wenn die Merkmale selbst unverändert sind. Zeitanpassungen sind implizite Eingeständnisse, dass die impliziten Preise von gestern nicht die von heute sein müssen.

Warum es für die Vertretbarkeit wichtig ist. Die Sichtweise über implizite Preise legt Vertretbarkeitsfragen der Anpassung offen, die das informelle Raster verdecken kann:

Wo im Merkmalsraum befinden Sie sich? Ein impliziter Preis ist nur vertretbar, wenn er nahe der Lage des Bewertungsobjekts im Merkmalsraum geschätzt wurde. Eine aus Binnenland-Comps abgeleitete Aussichtsprämie, angewandt auf ein Bewertungsobjekt in Strandlage, extrapoliert über eine Region des Merkmalsraums, in der das Verhalten der Funktion unbekannt ist.

Wie ist die lokale Geometrie? Wenn zwei Merkmale interagieren — wenn etwa Aussicht auf einem größeren Grundstück mehr wert ist —, dann ist ihre getrennte additive Behandlung eine Fehlspezifikation. Der implizite Preis der Aussicht hängt von der Grundstücksgröße ab, und eine Anpassung, die die Interaktion ignoriert, unterschätzt die tatsächliche Struktur der Funktion.

Wie wurde der implizite Preis gewonnen? Eine Paarvergleichsanalyse, ein Regressionskoeffizient, eine Branchenauswertung und ein „fachliches Ermessen“ zeigen alle auf dieselbe Größe — den impliziten Preis —, aber mit sehr unterschiedlichem epistemischem Rang. Die Vertretbarkeit des Gutachters hängt davon ab, dass er angibt, welche Quelle den in jeder Anpassung verwendeten impliziten Preis hervorgebracht hat.

Durchgerechnetes Bewertungsbeispiel. Für das durchgehende Pacifica-Beispiel mit der im vorangehenden Eintrag übernommenen additiven Spezifikation sind die impliziten Preise einfach die Koeffizienten der linearen Funktion:

$$\pi_{\text{GLA}} = \beta_{\text{GLA}}, \quad \pi_{\text{lot}} = \beta_{\text{lot}}, \quad \pi_{\text{view}} = \beta_{\text{view}}, \quad \pi_{\text{cond}} = \beta_{\text{cond}}.$$

Mit den geschätzten Koeffizienten aus Eintrag 003:

Merkmal	Impliziter Preis	Einheit
GLA	\$575	pro sf
Grundstück	\$22.70	pro sf
Aussicht	\$86,179	pro ja/nein
Zustand	\$49,824	pro Stufe

Diese impliziten Preise implizieren eine Paarvergleichsanalyse. Zurück zur Differenz zwischen Comp A und Comp B aus Eintrag 001:

- Comp A hat Aussicht ($\Delta z_{\text{view}} = +1$); Beitrag +\$86,179.
- Comp A hat ein um 1,800 sf größeres Grundstück ($\Delta z_{\text{lot}} = +1,800$); Beitrag $1,800 \times \$22.70 = +\$40,860$.
- Alle übrigen Merkmale identisch.
- Prognostizierte Preisdifferenz: $\$86,179 + \$40,860 = \$127,039$.

Die beobachtete Preisdifferenz betrug \$145,000. Die Zerlegung nach impliziten Preisen erklärt davon etwa \$127,000; die übrigen \$18,000 spiegeln die Stichprobenvariabilität des kleinen Datensatzes wider (und, allgemeiner, jene Arten unbeobachteter Faktoren, die der Eintrag zur latenten Variable aufgreifen wird).

Zwei operative Punkte sind festzuhalten. Erstens leistet die Linearitätsannahme echte Arbeit: Unter einem nichtlinearen $p(\mathbf{z})$ könnte dieselbe physische Grundstücksdifferenz auf einem kleinen Grundstück mehr wert sein als auf einem großen, und die obige additive Zerlegung wäre verzerrt. Zweitens ist der implizite Preis von \$22.70 pro Quadratfuß Grundstück der Grenzwert eines zusätzlichen Quadratfußes, nicht der Gesamtwert des Grundstücks. Der Gesamtbeitrag eines 9,000-sf-Grundstücks zum Bündelpreis beträgt $9,000 \times \$22.70 = \$204,300$, aber ein weiterer Quadratfuß erhöht den Bündelpreis nur um \$22.70.

Diese Unterscheidung — marginal vs. total — ist die Quelle erheblicher Verwirrung in der alltäglichen Bewertungssprache und einer der wichtigsten konzeptionellen Gewinne daraus, den impliziten Preis explizit zu benennen.

Literatur.

- 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.

Querverweise: heterogenes Gut; Merkmalsraum; hedonische Preisfunktion; hedonische Regression.

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (004-implicit-price.pl) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 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).

```

Hedonische Regression

Foundations • The Valuation Engineer

Bert Craytor • 2. August 2026

Version 1.0.0

DOI: [10.67330/ams1t139](https://doi.org/10.67330/ams1t139)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Hedonische Regression ist die Schätzung der hedonischen Preisfunktion aus beobachteten Transaktionen in einem Markt heterogener Güter; sie liefert empirische Schätzungen der impliziten Preise der Merkmale, die in das Modell eingehen.

Intuitive Einordnung. Die vorangehenden Einträge beschrieben die hedonische Preisfunktion und ihre impliziten Preise als theoretische Objekte. Die Funktion existiert im Prinzip, aber in der Praxis beobachten wir nur eine endliche Menge von Transaktionen: Bündel, die gehandelt wurden, und die Preise, zu denen sie gehandelt wurden. Die hedonische Regression ist die empirische Brücke zwischen beidem.

Der Aufbau ist geradlinig. Jede Transaktion liefert ein Paar $(\mathbf{z}_i, p_i)$: den Merkmalsvektor des gehandelten Bündels und seinen Verkaufspreis. Die Sammlung dieser Paare ist eine Stichprobe aus der (unbeobachteten) hedonischen Preisfunktion. Die Regression passt dieser Stichprobe eine funktionale Form an und schätzt Parameter, die die Funktion indizieren. Sind die Parameter geschätzt, liefern partielle Ableitungen der angepassten Funktion Schätzungen der impliziten Preise.

Drei Modellierungsentscheidungen müssen getroffen werden, bevor die Regression laufen kann:

1. *Welche Merkmale gehen in das Modell ein?* Die Wahl bestimmt die Dimensionen der geschätzten Funktion. Wird ein relevantes Merkmal ausgelassen, sickert sein Effekt in die Koeffizienten korrelierter aufgenommener Merkmale (Verzerrung durch ausgelassene Variablen, Omitted-Variable Bias).
2. *Welche funktionale Form?* Linear, log-linear, log-log, polynomial, basiserweitert, semiparametrisch, maschinelles Lernen. Die Wahl beschränkt, welche Gestalt die geschätzte Funktion annehmen kann und welche Muster impliziter Preise sie zurückgewinnen kann.
3. *Welcher Schätzer?* OLS, verallgemeinerte kleinste Quadrate, robuste M-Schätzer, Quantilsregression, räumlich ökonometrische Schätzer, penalisierte Regression. Die Wahl bestimmt, was der Schätzer optimiert und welche Annahmen seine Inferenz stützen.

Jede Entscheidung ist methodisch substanziell, und jede hat Konsequenzen für die Vertretbarkeit.

Wo Gutachter dem Konzept begegnen. Die hedonische Regression ist das Fundament der Massenbewertung und der Motor der meisten automatisierten Bewertungsmodelle (AVMs). Wenn ein County Assessor jährlich 100,000 Parzellen neu bewertet, leistet eine hedonische Regression auf jüngeren Verkäufen die Arbeit. Wenn Zillow oder Redfin eine automatisierte Schätzung produziert, erzeugt sie ein Modell hedonischen Typs (oft ein Baum-Ensemble statt OLS, konzeptionell aber weiterhin ein hedonischer Schätzer). Wenn das Zweitmarkt-AVM eines Kreditgebers eine Immobilie für eine Refinanzierung durchwinkt oder ablehnt, ist dieselbe Maschinerie beteiligt.

Auch das Einzelgutachten stützt sich zunehmend auf hedonische Regression, selbst wenn die endgültige Wertindikation aus dem Vergleichswertverfahren stammt. Aus einer Regression auf einer

Nachbarschaftsstichprobe geschätzte implizite Preise können Anpassungshöhen weit vertretbarer stützen als Gutachterintuition oder Faustregel-Prozentsätze. Die Regressionskoeffizienten werden zu Beweismitteln, nicht Meinungen, für die im Raster angewandten Anpassungen.

Warum es für die Vertretbarkeit wichtig ist. Regressionsbasierte Schätzungen tragen ein Vertretbarkeitsprofil, das sich von der Paarvergleichsanalyse unterscheidet. Die Vorteile: Eine Regression nutzt alle verfügbaren Verkäufe statt jeweils zwei; sie kann den Effekt eines Merkmals isolieren und dabei viele andere gleichzeitig kontrollieren; und sie liefert Standardfehler, die die Unsicherheit jedes impliziten Preises quantifizieren. Die Kompromisse: Eine Regression erlegt eine funktionale Form auf, die zur wahren Preisfunktion möglicherweise nicht passt; sie benötigt mehr Daten als eine Paarvergleichsanalyse; und ihre Ergebnisse können empfindlich auf die Wahl von Stichprobe und Spezifikation reagieren.

Mehrere Vertretbarkeitsfragen kehren in der regressionsbasierten Bewertung wieder:

Ist die Stichprobe angemessen? Die Regression schätzt die hedonische Funktion, wie sie im Marktsegment und Zeitraum der Stichprobe wirkt. Eine Regression auf Binnenland-Verkäufen, angewandt auf ein Bewertungsobjekt in Strandlage, ist ein Vertretbarkeitsversagen, egal wie sauber die Statistik aussieht.

Ist die Spezifikation angemessen? Eine lineare Spezifikation ist eine starke Annahme. Zeigt die wahre Funktion oberhalb einer Schwelle abnehmende Erträge der GLA, wird die lineare Anpassung die impliziten Preise in diesem Bereich falsch schätzen. Diagnostik, Residuenplots und alternative Spezifikationen gehören zum vertretbaren Arbeitsablauf.

Was bedeutet der Standardfehler? Ein Standardfehler erfasst die Stichprobenvariabilität unter der angenommenen Spezifikation. Er erfasst nicht Modellfehlspezifikation, Verzerrung durch ausgelassene Variablen oder Stichprobenselektion. Ein kleiner Standardfehler bei falscher Spezifikation ist gefährlicher als ein großer Standardfehler bei richtiger.

Triangulieren die Ergebnisse mit anderen Ansätzen? Ein vertretbarer regressionsbasierter impliziter Preis sollte mit dem konsistent sein, was die Paarvergleichsanalyse nahelegt, was die Komponentenkosten des Sachwertverfahrens implizieren und was marktkundige Gutachter glauben. Ein Regressionskoeffizient im Widerspruch zu allen dreien ist eine Einladung zur Untersuchung, nicht zur Veröffentlichung.

Durchgerechnetes Bewertungsbeispiel. Wir können nun die volle Maschinerie der hedonischen Regression auf die acht Pacifica-Comps anwenden. Der Datensatz ist verfügbar unter `shared/data/pacifica_comps`

R-Code:

```
comps <- read.csv("shared/data/pacifica_comps.csv")
fit <- lm(price ~ gla + lot + view + cond, data = comps)
summary(fit)
```

Ausgabe:

```
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
```

Interpretation. Das angepasste Modell schätzt die hedonische Preisfunktion als:

$$\hat{p}(\mathbf{z}) = 80,777 + 575 z_{\text{GLA}} + 22.70 z_{\text{lot}} + 86,179 z_{\text{view}} + 49,824 z_{\text{cond}}.$$

Jeder Koeffizient ist der geschätzte implizite Preis des entsprechenden Merkmals. GLA trägt \$575 pro Quadratfuß bei (mit dem engsten Standardfehler und der stärksten Signifikanz); Aussicht trägt etwa \$86,000 zu einer Pacifica-Immobilie bei; eine zusätzliche Zustandsstufe (z.B. von „gut“ zu „ausgezeichnet“) ist rund \$50,000 wert.

Mehrere Vorbehalte sind unmittelbar sichtbar:

Die Stichprobe ist winzig. Acht Beobachtungen und vier Prädiktoren lassen drei Freiheitsgrade für die Residuen. Die Standardfehler sind entsprechend breit, und die p -Werte für Grundstück, Aussicht und Zustand verfehlen die konventionelle 0.05-Schwelle trotz ihrer ökonomischen Plausibilität. Das ist ein Kleinstichproben-Artefakt, kein Beleg dafür, dass diese Merkmale nicht zählen. Mit 80 statt 8 Comps würden die p -Werte drastisch schrumpfen.

Die Gesamtanpassung ist stark. Ein adjustiertes $R^2 = 0.93$ und ein Modell- p -Wert von 0.012 zeigen an, dass die aufgenommenen Merkmale gemeinsam den Großteil der Preisvariation in der Stichprobe erklären. Die Kombination aus starker gemeinsamer Anpassung und schwacher individueller Signifikanz ist ein Lehrbuchsignal für Multikollinearität, die hier strukturell ist: GLA, Grundstück und Zustand kovariieren in der Stichprobe.

Ein Residuum sticht heraus. Das maximale Residuum beträgt +\$46,455 und entspricht Comp H. Die übrigen sieben Residuen liegen alle innerhalb eines Bandes von $\pm \$18,000$. Irgendetwas an Comp H wird von der angepassten Funktion nicht erfasst — das Modell unterschätzt seinen Preis systematisch, um einen Betrag, der fast dreimal so groß ist wie das nächstgrößte Residuum.

Dieses Ausreißer-Residuum ist kein Rauschen. Es ist das Signal, dass ein unbeobachtetes Merkmal von Comp H auf eine Weise zu dessen Preis beiträgt, die die aufgenommenen Variablen nicht zurückgewinnen können. Der Eintrag zur latenten Variable greift genau diese Frage auf.

Validierung gegen das Paarvergleichsdenken. Die Paarvergleichsanalyse aus Eintrag 004 prognostizierte eine Differenz von \$127,039 zwischen Comp A und Comp B allein aus Aussicht und Grundstück. Die impliziten Preise der Regression sind identisch (da sie die Quelle jener Berechnung sind). Die Regression verallgemeinert die Paarvergleichslogik auf alle acht Comps gleichzeitig, bringt aber bei dieser Stichprobengröße nichts grundsätzlich Neues hervor; sie bestätigt die Paaranalyse und berichtet dabei Standardfehler, die das Paarvergleichsdenken nicht liefern kann.

Literatur.

- 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.

Querverweise: *heterogenes Gut; Merkmalsraum; hedonische Preisfunktion; impliziter Preis; latente Variable.*

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (005-hedonic-regression.pl) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```
%% 005-hedonic-regression.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 005
%% Title: Hedonic Regression
%% Author: Bert Craytor
%% Concept: hedonic_regression
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('.././.././../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(hedonic_regression, "Hedonic Regression", econometric_method,
%      "Hedonic regression is the estimation of the hedonic price
%      function from observed transactions in a heterogeneous-goods
%      market, yielding empirical estimates of the implicit prices of
%      the characteristics that enter the model.",
%      published, 'dictionary-2026.1').
%
% term(mass_appraisal, ...).
% term(automated_valuation_model, ...). % synonym: avm
% term(omitted_variable_bias, ...).
% term(multicollinearity, ...).
% term(residual, ...).
% term(ordinary_least_squares, ...).
%
% introduced_by(hedonic_regression, court_1939).
% introduced_by(hedonic_regression, griliches_1961).
% formalized_by(hedonic_regression, rosen_1974).
% depends_on(hedonic_regression, hedonic_price_function).
% depends_on(hedonic_regression, implicit_price).
% -----

:- use_terms([ hedonic_regression, hedonic_price_function, implicit_price,
               characteristics_space, heterogeneous_good, latent_variable,
               mass_appraisal, automated_valuation_model,
               omitted_variable_bias, multicollinearity, residual,
               ordinary_least_squares,
               sales_comparison_approach, paired_sales_analysis, adjustment ]).
:- use_relations([related/2, depends_on/2,
                 introduced_by/2, formalized_by/2]).
```

```

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_regression).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.005').

% =====
% LOCAL VOCABULARY
% =====

local_term(specification).
local_term(coefficient).
local_term(standard_error).
local_term(p_value).
local_term(r_squared).
local_term(adjusted_r_squared).
local_term(residual_standard_error).
local_term(degrees_of_freedom).
local_term(f_statistic).
local_term(significance_level).
local_term(diagnostic).

% =====
% THE EMPIRICAL BRIDGE
% =====
%
% Hedonic regression is the bridge from theoretical  $p(z)$  to a sample of
% observed  $(z_i, p_i)$  pairs.

setup_of_regression(
    each_transaction_supplies_pair("(z_i, p_i): characteristics vector and sale price"),
    sample_from("the unobserved hedonic price function"),
    fits("a functional form to that sample, estimating parameters")).

% =====
% THREE MODELING CHOICES (per the article)
% =====

modeling_choice(which_characteristics,
    "Determines the dimensions of the estimated function. Omission causes omitted_variable_bias.").
modeling_choice(functional_form,
    "Constrains what shape the estimated function can take. Linear, log-linear, polynomial, basis-
    ↪ expanded, semiparametric, ML.").
modeling_choice(estimator,
    "OLS, GLS, robust, quantile, spatial, penalized. Determines optimization criterion and
    ↪ inference assumptions.").

% =====
% PRACTICE CONTEXTS WHERE REGRESSION IS THE ENGINE
% =====

regression_engine_context(mass_appraisal,
    "Assessor revaluation of 100,000 parcels annually.").
regression_engine_context(automated_valuation_model,
    "Zillow/Redfin/lender AVMs; tree ensembles are conceptually hedonic estimators.").
regression_engine_context(single_property_appraisal,
    "Regression coefficients support adjustments more defensibly than appraiser intuition.").

```

```
% =====
% DEFENSIBILITY PROFILE: REGRESSION vs PAIRED-SALES
% =====

advantage_over_paired_sales(uses_all_data,
  "Regression uses all available sales rather than two at a time.").
advantage_over_paired_sales(simultaneous_control,
  "Isolates one characteristic while controlling for many others.").
advantage_over_paired_sales(quantified_uncertainty,
  "Produces standard errors quantifying uncertainty in each implicit price.").

tradeoff(functional_form_constraint,
  "Regression imposes a form that may not match the true price function.").
tradeoff(data_requirement,
  "Requires more observations than paired-sales analysis.").
tradeoff(sample_specification_sensitivity,
  "Results can be sensitive to the choice of sample and specification.").

% =====
% RECURRING DEFENSIBILITY QUESTIONS
% =====

defensibility_question(sample_appropriateness,
  "The regression estimates the function for the sample's market segment and time period;
  ↪ mismatch is a defensibility failure regardless of statistical fit.").
defensibility_question(specification_appropriateness,
  "Linear is a strong assumption; diagnostics, residual plots, and alternative specifications
  ↪ belong in the defensible workflow.").
defensibility_question(standard_error_interpretation,
  "Standard errors capture sampling variability under the assumed specification, NOT
  ↪ misspecification, omitted-variable bias, or sample selection.").
defensibility_question(triangulation,
  "Regression coefficients should be consistent with paired-sales, cost components, and local
  ↪ appraiser opinion; isolated coefficients are an invitation to investigate.").

% =====
% WORKED EXAMPLE: lm(price ~ gla + lot + view + cond)
% =====

:- consult('../..\\kb\\pacificacomps').

% The fitted model.
fit_call("lm(formula = price ~ gla + lot + view + cond, data = comps)").
fit_estimator(ordinary_least_squares).
fit_software(r_lm).

% Fitted coefficients, with standard errors and significance from
% the R output reproduced in the article.
fitted_coefficient(intercept, 80777.07, 164509.35, 0.491, 0.6571, ' ').
fitted_coefficient(gla, 574.93, 109.76, 5.238, 0.0135, '*').
fitted_coefficient(lot, 22.70, 10.38, 2.187, 0.1165, ' ').
fitted_coefficient(view, 86179.03, 31342.97, 2.750, 0.0708, '.').
fitted_coefficient(cond, 49824.33, 27761.27, 1.795, 0.1706, ' ').

% Diagnostics from the R output.
residual_summary(min, -17995).
residual_summary(q1, -8243).
residual_summary(median, -5433).
```

```

residual_summary(q3, -1409).
residual_summary(max, 46455).

model_summary(residual_standard_error, 30481).
model_summary(degrees_of_freedom, 3).
model_summary(r_squared, 0.9707).
model_summary(adjusted_r_squared, 0.9317).
model_summary(f_statistic, 24.86).
model_summary(model_p_value, 0.01231).

% =====
% INTERPRETIVE CAVEATS FROM THE ARTICLE
% =====

caveat(tiny_sample,
    "n=8, four predictors, three residual df. Standard errors are correspondingly wide; lot/view/
    ↪ cond fail to reach the 0.05 threshold despite economic plausibility. Small-sample artifact.").
    ↪

caveat(structural_multicollinearity,
    "Adj R^2 = 0.93 with model p = 0.012 indicates strong joint fit; weak individual significance
    ↪ is a textbook multicollinearity signature. Here GLA, lot, and condition covary in the sample
    ↪ .").

caveat(comp_h_residual,
    "Comp H's residual is +$46,455; the other seven fall within +/- $18,000. The model
    ↪ systematically under-predicts Comp H by ~3x the next-largest residual. NOT noise. Signal of
    ↪ an unobserved characteristic.").

% =====
% RELATING TO PAIRED-SALES (entry 004)
% =====
%
% The regression's implicit prices ARE the values used in entry 004's
% paired-sales decomposition of A vs B. The regression generalizes that
% logic to all 8 comps simultaneously while reporting standard errors
% that paired-sales reasoning cannot produce.

confirms_paired_sales_from(entry_004,
    "A vs B differential of $127,039 from view and lot alone - coefficients are the same;
    ↪ regression confirms the paired analysis while adding standard errors.").

% =====
% RESIDUAL OF COMP H POINTS FORWARD TO ENTRY 006
% =====

points_forward_to(latent_variable,
    "Comp H's anomalous +$46,455 residual is the signal that an unobserved characteristic is
    ↪ contributing to price. Entry 006 takes up the question.").

% =====
% PREDICTED PRICES FROM THE FITTED MODEL
% =====
%
% Computes predicted price from the fitted coefficients. The residual
% for each comp is sale_price - predicted_price.

predicted_price(P, Predicted) :-
    coord(P, gla_sqft, GLA),

```

```

    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond),
    fitted_coefficient(intercept, B0, _, _, _, _),
    fitted_coefficient(gla,      Bg, _, _, _, _),
    fitted_coefficient(lot,      Bl, _, _, _, _),
    fitted_coefficient(view,     Bv, _, _, _, _),
    fitted_coefficient(cond,     Bc, _, _, _, _),
    Predicted is B0 + Bg * GLA + Bl * Lot + Bv * View + Bc * Cond.

residual_of(P, R) :-
    sale_price(P, Observed),
    predicted_price(P, Predicted),
    R is Observed - Predicted.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V
    ; A == view, V == yes -> N = 1
    ; A == view, V == no  -> N = 0
    ; V = N
    ).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).
cross_reference(latent_variable).

```

Latente Variable

Foundations • The Valuation Engineer

Bert Craytor • 2. August 2026

Version 1.0.0

DOI: [10.67330/11n3g168](https://doi.org/10.67330/11n3g168)

Deutsche Fassung — Übersetzung mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Formale Definition. Eine latente Variable ist ein unbeobachtetes Merkmal, das beobachtete Ergebnisse beeinflusst und dessen Wert aus seinen Wirkungen auf beobachtbare Variablen erschlossen werden muss, statt direkt gemessen zu werden.

Intuitive Einordnung. Die ersten fünf Einträge behandelten den Merkmalsvektor $\mathbf{z}$, als ließe er sich vollständig messen: GLA, Grundstück, Aussicht, Zustand und so weiter, allesamt als Datenspalten verfügbar. In der Praxis ist kein Merkmalsvektor jemals vollständig. Manche Merkmale fehlen in den Daten, weil sie schwer zu messen sind (Qualität des Innenausbaus, Fließen des Grundrisses, mikrolokale Feinausrichtung), weil sie in den üblichen Datenquellen nicht vorkommen (MLS-Datensätze erfassen weder Lichtqualität noch Schallschutz noch jüngste ungenehmigte Verbesserungen) oder weil sie noch nicht als relevant erkannt wurden (Merkmale auf Nachbarschaftsebene, die sich erst nach der Analyse als preisrelevant erweisen).

Diese unbeobachteten-aber-relevanten Merkmale sind *latente Variablen*. Sie existieren, sie beeinflussen den Preis, aber sie stehen nicht in den Spalten unserer Regression. Ihre Wirkungen verschwinden nicht dadurch, dass sie ungemessen bleiben; sie sammeln sich in den Residuen des Modells, wo sie aus Sicht der aufgenommenen Variablen als „Rauschen“ erscheinen, aus Sicht des zugrunde liegenden datenerzeugenden Prozesses aber tatsächlich strukturiertes Signal sind.

Latente Variablen sind die Brücke zwischen dem idealisierten hedonischen Rahmenwerk und der unordentlichen Realität der Bewertungsdaten. Sie ausdrücklich anzuerkennen ist Voraussetzung für ehrliches Nachdenken über Modellunsicherheit, Residuenstruktur und die Grenzen dessen, was beobachtbare Merkmale erklären können.

Wo Gutachter dem Konzept begegnen. Gutachter begegnen latenten Variablen in drei wiederkehrenden Zusammenhängen.

Die Anpassungen für Qualität und Zustand. „Qualität“ und „Zustand“ sind als Bewertungskategorien selbst schon Versuche, ein mehrdimensionales latentes Konstrukt auf eine einzige Ordinalskala zusammenzuziehen. Eine Zustandsbewertung C3 nach UAD 3.6 soll den aufgestauten Instandhaltungsrückstand einer Immobilie, den Verschleiß der Oberflächen, das Alter der technischen Anlagen und die Präsentation zusammenfassen — vier bis zehn zugrunde liegende Merkmale, verdichtet zu einer einzigen Zahl. Die Ein-Zahl-Zusammenfassung ist ein Proxy für ein latentes Konstrukt; zur Kunst des Gutachters gehört das Urteil darüber, wie gut der Proxy die zugrunde liegende Wahrheit im Einzelfall erfasst.

Das unerklärte Residuum im Raster. Wenn drei angepasste Comps einen Wert anzeigen, dabei aber stärker voneinander abweichen als erwartet, ist die Abweichung informativ. Sie ist selten reines Stichprobenrauschen; häufiger spiegelt sie ein Merkmal wider, das der Gutachter noch nicht identifiziert und angepasst hat. Der Abweichung im Residuum nachzugehen — „Warum zeigt dieser Comp \$50,000 über den anderen an?“ — ist die informelle Version der Latente-Variablen-Inferenz beim Gutachter.

Die jüngste Verbesserung, die niemand sehen kann. Eine Immobilie kann eine Küchenrenovierung, eine Dacherneuerung oder den genehmigten Ausbau einer ADU hinter sich haben, die sich noch nicht in MLS- oder Katasterdaten niederschlägt. Aus Sicht des Gutachters ist die Verbesserung vor Ort beobachtbar, in den Daten aber nicht. Aus Sicht der Regression ist die Verbesserung latent. Der Gutachter muss die Beobachtung vor Ort mit der datengestützten Indikation in Einklang bringen, und beide sollten auf den wahren Beitrag des latenten Merkmals zulaufen.

Warum es für die Vertretbarkeit wichtig ist. Latente Variablen führen mehrere Vertretbarkeitsüberlegungen ein, die das additive lineare Modell aus Eintrag 005 nicht unmittelbar behandelt.

Verzerrung durch ausgelassene Variablen. Ist eine latente Variable mit einer beobachteten Variable korreliert, sickert ihr Effekt in den Koeffizienten der beobachteten Variable. Wenn kürzlich verkaufte Pacifica-Häuser tendenziell renovierte Küchen haben (ein latentes Merkmal) und das Vertragsdatum des Hausverkaufs in den Daten steht, dann wird der Koeffizient des Verkaufsalters (auch `sale_age` oder „Tage seit Verkauf“) einen Teil des Küchenrenovierungseffekts aufnehmen. Der geschätzte implizite-Preis-Beitrag von `sale_age` wird verzerrt — er spiegelt `sale_age plus` einen Anteil des ungemessenen Küchenrenovierungsstatus wider. Vertretbare Bewertung erkennt an, dass die Koeffizienten beobachteter Variablen bedingt sind auf das, was im Modell fehlt.

Interpretation der Residuen. Die Standardregressionstheorie behandelt Residuen als gaußsches Rauschen mit Erwartungswert null. Die Latente-Variablen-Sicht behandelt Residuen als Mischung aus Rauschen und Signal: ein Teil ist Stichprobenvariation, ein Teil ist die kumulierte Wirkung unbeobachteter Merkmale. Zu bestimmen, was wovon ist, ist methodisch substanziell. Ein Residuum, das mit der Lage, mit dem Verkaufsdatum, mit der Identität des Gutachters oder mit irgendeinem anderen Beobachtbaren korreliert, legt nahe, dass die zugehörige latente Variable teilweise zurückgewonnen werden kann.

Modellwahl. Mehr beobachtete Variablen aufzunehmen ist das naheliegende Heilmittel gegen die Verzerrung durch ausgelassene Variablen, aber die latenten Variablen, auf die es am meisten ankommt, sind oft gerade diejenigen, die sich am schwersten in das Modell aufnehmen lassen (Feinkörnigkeit des Zustands, Qualität jüngster Verbesserungen, Nachbarschafts-Mikroeffekte). Die vertretbare Antwort besteht mitunter darin anzuerkennen, dass das Modell sie nie erfassen wird, und die unerklärte Varianz dem Residuum zuzuweisen, statt zusätzliche Variablen mit Gewalt anzupassen.

Durchgerechnetes Bewertungsbeispiel. Die Regression in Eintrag 005 erzeugte ein ungewöhnliches Residuum für Comp H: Das Modell unterschätzte den Preis von Comp H um \$46,455, während die übrigen sieben Residuen in ein Band von $\pm \$18,000$ fielen. Die beobachteten Merkmale von Comp H — GLA 1,600, Grundstück 7,300, keine Aussicht, Zustand „gut“ — platzieren ihn klar in der Mitte der Stichprobe; nichts an seinem beobachtbaren Profil sagt einen ungewöhnlich hohen Preis voraus. Und doch wurde er für \$46,455 mehr verkauft, als das Modell sagt.

Die naheliegende Hypothese lautet, dass Comp H ein latentes Merkmal besitzt, das zu seinem Preis beiträgt, aber nicht in den Daten auftaucht. Bei der Konstruktion des Datensatzes für diese Ausgabe ist genau das der Fall: Comp H hat eine jüngste Küchenrenovierung erfahren, die dem Bündelwert etwa \$60,000 hinzufügt. Die Renovierung stand nicht in der Variablenliste der Regression, sodass sich der Beitrag im Residuum statt in einem Koeffizienten ansammelte. Das geschätzte Residuum (\$46,455) ist etwas kleiner als der wahre latente Beitrag (\$60,000), weil die Regression kleine Anteile des latenten Effekts auf die übrigen Koeffizienten verteilt und diese in korrelierten Dimensionen nach oben verzerrt.

Daraus folgen drei operative Konsequenzen:

Behandeln Sie das Ausreißer-Residuum als Evidenz, nicht als Fehler. Die erste angemessene Reaktion auf das Residuum von Comp H besteht nicht darin, ihn als Ausreißer zu

verwerfen, sondern zu untersuchen, was an der Immobilie die Unterprognose erklärt. In diesem Fall würde eine Besichtigung vor Ort oder eine Recherche in den Genehmigungsakten die Renovierung zutage fördern. Das Versagen des Modells hat diagnostischen Wert.

Erkennen Sie an, dass die übrigen Koeffizienten leicht verzerrt sind. Die Koeffizienten für Aussicht, Zustand und Grundstück haben über Korrelationen mit den beobachteten Merkmalen von Comp H kleine Bruchteile des Renovierungseffekts aufgenommen. Die Verzerrung ist bei $n = 8$ klein und würde mit mehr Comps und mehr Kovariaten weiter schrumpfen, aber sie ist nicht null.

Die Residuen als Ganzes tragen zurückgewinnbare Struktur. Dies ist die substanzielle These der vom Autor entwickelten Methodik des Residual Constraint Approach (RCA) (Craytor, 2025): In Daten der Wohnimmobilienbewertung sind Residuen systematisch kein iid-gaußsches Rauschen. Sie spiegeln den strukturierten Beitrag unbeobachteter Merkmale wider — Feinkörnigkeit des Zustands, Feinkörnigkeit der Qualität, funktionale Nützlichkeit und andere latente Konstrukte —, der sich teilweise zurückgewinnen lässt, indem man das Residuum selbst als Latente-Variablen-Signal behandelt, das modelliert statt verworfen werden sollte.

Die Behandlung von Zustand und Qualität im RCA macht diese zweistufige Struktur operational. Die ordinalen Bewertungen des Gutachters — die oben besprochenen grobkörnigen Proxys, die man mit Zustand-1 (Zustand grobkörnig) und Qualität-1 (Qualität grobkörnig) bezeichnen kann — sind grob genug, dass man sie ebenso gut selbst als latente Variablen behandeln könnte; sie lassen sich aber dennoch als Regressoren aufnehmen, um näherungsweise implizite-Preis-Beiträge zu erhalten. Das Protokoll verschiebt diese geschätzten Beiträge dann aus den gemessenen Merkmalen im Regressionsteil des Verkaufsrasters zurück in das Residuum, von wo aus sie den feinkörnigen latenten Variablen Zustand-2 (Zustand feinkörnig) und Qualität-2 (Qualität feinkörnig) neu zugewiesen und anhand zusätzlicher Evidenz wie den öffentlichen MLS-Bemerkungen verfeinert werden. Die grobkörnige Bewertung liefert eine erste Wertzuweisung; die feinkörnige latente Variable trägt die zurückgewinnbare Struktur. Am Ende ersetzt der feinkörnige Wertbeitrag den grobkörnigen Wertbeitrag.

Die Ein-Comp-Illustration hier ist der einfachste Fall der RCA-Intuition. Mit mehr Daten und reicherer Modellierung lässt sich die Struktur im Residuum in erkennbare latente Faktoren zerlegen, von denen jeder auf Weisen zum Marktpreis des Bündels beiträgt, die die hedonische Standardspezifikation nicht erreichen kann. Latente Variablen sind kein Störfaktor, den man wegkontrollieren muss; sie sind die nächste Grenze der vertretbaren Bewertungsmodellierung.

Literatur.

- 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)

Querverweise: Merkmalsraum; hedonische Preisfunktion; impliziter Preis; hedonische Regression.

Addendum: Prolog-Formalisierung

Die folgende maschinenlesbare Begleitdatei (006-latent-variable.pl) formalisiert diesen Eintrag in der Prolog-Wissensbasis der Zeitschrift. Sie wird zudem als browsbare Fahne (Galley) neben diesem Artikel veröffentlicht.

```

%% 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).
```

Aktuelle Gerichtsentscheidungen

Aktuelle Gerichtsentscheidungen — Vereinigte Staaten

Aktuelle Gerichtsentscheidungen • The Valuation Engineer

William Bert Craytor • Jahrgang 1, Ausgabe 1 • Sommer 2026

DOI: 10.67330/fz34f711

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Fallzusammenfassungen nach Jurisdiktion

Kalifornien

The Retail Property Trust v. Orange County Assessment Appeals Board No. 1. California Court of Appeal (kalifornisches Berufungsgericht), entschieden im April 2026.

Das Court of Appeal befasste sich mit der Frage, ob nach Section 170 des California Revenue and Taxation Code (kalifornisches Steuer- und Abgabengesetzbuch) eine Ermäßigung der Grundsteuer-Veranlagung für Zugangsbeschränkungen zu gewähren ist, die durch behördliche Anordnungen während der COVID-19-Pandemie verhängt wurden. Das klagende Einkaufszentrum machte geltend, der beschränkte Zugang während der Pandemie stelle einen Schaden („damage“) am Grundstück im Sinne von Section 170 dar und begründe damit einen Anspruch auf Ermäßigung der Veranlagung.

Bedeutung für Gutachter. Die gerichtliche Auslegung des Begriffs „damage“ nach Section 170 wird künftige Ermäßigungsanträge prägen, die auf extern verursachte Minderungen der Nutzbarkeit eines Grundstücks gestützt werden — gleich ob diese aus Anordnungen des Gesundheitsschutzes, aus Infrastrukturvorhaben oder aus sonstigen behördlichen Maßnahmen mit Auswirkungen auf die Zugänglichkeit herrühren. Gutachter, die sich zu den Wertauswirkungen von Zugangsbeschränkungen bei Gewerbeimmobilien äußern sollen, sollten die Erwägungen des Gerichts dazu prüfen, ob eine eingeschränkte Nutzbarkeit — im Unterschied zu einer körperlichen Beschädigung — einen für Zwecke der Veranlagungsermäßigung ausgleichsfähigen Schaden darstellt.

Colorado

Cheroutes v. Rocket Mortgage & Solidifi. U.S. District Court, D. Colorado (US-Bundesbezirksgericht für den District of Colorado); anhängig, mit Sachstandsmeldungen im Jahr 2026.

Die zugrunde liegende Klage macht eine gutachterliche Voreingenommenheit (*appraisal bias*) nach dem Fair Housing Act (US-Bundesgesetz gegen Diskriminierung auf dem Wohnungsmarkt) im Zusammenhang mit einer Wertermittlung aus dem Jahr 2021 für ein Zweifamilienhaus in Denver geltend, das mit rund 640.000 US-Dollar bewertet wurde — mehr als 220.000 US-Dollar unter dem im Vorjahr ermittelten Wert. Ein Bundesrichter hat den Abweisungsantrag der Beklagten (*motion to dismiss*) bereits zurückgewiesen und entschieden, dass Rocket Mortgage und die Appraisal Management Company Solidifi sich gegen die Ansprüche verteidigen müssen, statt eine Verantwortlichkeit mit der Erwägung von sich zu weisen, Kreditgeber und AMCs könnten für das Verhalten unabhängiger, auf Honorarbasis tätiger Gutachter nicht haften. Das Verfahren befindet sich 2026 im Beweisermittlungsverfahren (*discovery*).

Bedeutung für Gutachter. Das Verfahren ist derzeit der wichtigste anhängige Fall, an dem die Haftung von AMC's und Kreditgebern nach dem Fair Housing Act für das Verhalten freiberuflicher Gutachter erprobt wird. Sollte das Gericht die AMC und den Kreditgeber am Ende in der Sache für haftbar halten, würde dies den Kreis der Parteien, die aus Wertermittlungen resultierenden Fair-Housing-Ansprüchen ausgesetzt sind, erheblich erweitern und sich voraussichtlich auf die Auftragsverträge zwischen AMC und Gutachter durchschlagen.

Florida

Sammelklage in Florida gegen AMC-Gebühren, die Darlehensnehmern von Wohnimmobilien berechnet werden. Eingereicht am 2. Januar 2026.

Die Klage wendet sich gegen die Gebühren von Appraisal Management Companies, die Darlehensnehmern von Wohnimmobilienhypotheken berechnet werden. Sie macht geltend, die Marge zwischen dem an den freiberuflichen Gutachter gezahlten Honorar und der dem Darlehensnehmer beim Abschluss berechneten Gebühr überschreite die bundesrechtlich festgelegten Grenzen und werde unzureichend offengelegt. Der Fall liegt an der Schnittstelle des Dodd-Frank-Erfordernisses einer „customary and reasonable“ (üblichen und angemessenen) Vergütung (12 U.S.C. § 1639e) und der Offenlegungsregime des Truth in Lending Act sowie des Real Estate Settlement Procedures Act.

Bedeutung für Gutachter. Der Rechtsstreit stellt die Struktur der AMC-Vergütung auf den Prüfstand und könnte, sofern er als Sammelklage zugelassen und in der Sache entschieden wird, die Offenlegungspraxis branchenweit umgestalten. Das Ergebnis kann sich mittelbar auf das Honorarumfeld freiberuflicher Gutachter auswirken, die über AMC's arbeiten, indem es die Spanne zwischen dem von der AMC gezahlten und dem dem Darlehensnehmer berechneten Betrag gerichtlicher Kontrolle unterwirft.

Georgia

Die Green-Rock-Verfahren zu Naturschutz-Dienstbarkeiten (*conservation easements*). United States Tax Court (US-Bundesfinanzgericht); die Entscheidungen wurden am 2. März 2026 in *Tax Notes* besprochen.

Der Tax Court arbeitete bei der Bewertung von Naturschutz-Dienstbarkeiten, die an benachbarten Grundstücken (Green Rock Property und Green Rock South) in der Piedmont-Provinz Georgias bestellt worden waren, jede Stufe des Tests der höchsten und besten Nutzung (*highest and best use*) durch. Die Sachverständigen der Antragsteller machten geltend, die höchste und beste Nutzung der Grundstücke vor Bestellung der Dienstbarkeit sei der Betrieb eines Granitsteinbruchs gewesen, und schätzten mittels Discounted-Cashflow-Analyse einen Marktwert vor Bestellung der Dienstbarkeit von rund 13,75 Millionen US-Dollar gegenüber einem Wert nach Bestellung von 180.000 US-Dollar.

Das Gericht äußerte zu jeder Stufe des Tests Vorbehalte. Zur rechtlichen Zulässigkeit war es nicht davon überzeugt, dass eine Umwidmung zur Zulassung des Steinbruchbetriebs hinreichend wahrscheinlich gewesen wäre; es hob insbesondere hervor, dass eine Änderung der bauplanungsrechtlichen Einstufung für die Green Rock Property kurz nach der Beauftragung eines neuen Wertgutachtens zur Naturschutz-Dienstbarkeit erwirkt worden war — eine zeitliche Abfolge, der das Gericht nur schwer Glauben schenken konnte. Zur physischen Möglichkeit hielt das Gericht die geologischen Nachweise der Antragsteller für unzureichend, um zu belegen, dass auf dem Grundstück an der Harman Road ein tragfähiger Steinbruch hätte entwickelt werden können — ungeachtet des allgemein bekannten Vorkommens abbaufähigen Granits in der Piedmont-Provinz. Das Gericht entschied ferner, dass

selbst bei unterstellter rechtlicher Zulässigkeit und physischer Möglichkeit dies ohne Erfüllung der übrigen Stufen der Prüfung nicht ausschlaggebend wäre.

Bedeutung für Gutachter. Die Entscheidung ist ein durchgearbeitetes Beispiel für die disziplinierte, schrittweise Anwendung des vierstufigen Tests der höchsten und besten Nutzung. Sie bekräftigt die mit *Ranch Springs, LLC v. Commissioner*, 164 T.C. No. 6 (2025), und *Beaverdam Creek Holdings, LLC v. Commissioner*, T.C. Memo. 2025-53, gezogene Linie, dass der Tax Court spekulative Behauptungen zur höchsten und besten Nutzung auch dann nicht akzeptiert, wenn die allgemeine Machbarkeit der vorgeschlagenen Nutzung für die umliegende Region dokumentiert ist. Gutachter, die Bewertungen von Dienstbarkeiten erstellen, sollten für jede Stufe grundstücksbezogene Nachweise erarbeiten, statt sich auf regionale oder analogiegestützte Machbarkeit zu verlassen, und sollten bei zeitlichen Mustern besonders vorsichtig sein — etwa bei Umwidmungen, die zeitgleich mit der Beauftragung des Gutachtens erwirkt werden —, die richterliche Skepsis hervorrufen können.

Maryland

Erwerb des Pimlico Race Course durch die Stadt Baltimore. Enteignungsverfahren (*eminent domain*), laufend 2025–2026.

Der Erwerb des Pimlico Race Course durch die City of Baltimore im Wege der Enteignung wirft zur angemessenen Entschädigung (*just compensation*) Fragen nach dem Verhältnis zwischen dem Wert des laufenden Unternehmens (*going-concern value*) und dem Wert des Grundeigentums bei Spezialimmobilien auf. Da der Betrieb von Pimlico funktional untrennbar mit der baulichen Anlage verbunden ist, ist eine herkömmliche Vergleichswertanalyse für das Grundeigentum nur begrenzt brauchbar, und Ertragswertermittlungen hängen stark von der unterstellten Fortführung des Rennbetriebs ab.

Bedeutung für Gutachter. Das Verfahren verdient Beachtung durch jeden Gutachter, der eine Einzweckanlage bewerten soll, bei der Betrieb und Immobilie funktional untrennbar sind — etwa Rennbahnen, Filmstudios, Freizeitparks, Marinas sowie bestimmte Gesundheits- oder Hotelimmobilien. Die aus streitigen Verfahren hervorgehende Methodik wird voraussichtlich beeinflussen, wie künftige Enteignungsverfahren Going-Concern-Wertanteile bei der Berechnung der angemessenen Entschädigung behandeln.

Bundesrechtliche Klagen gegen Gutachter nach dem Fair Housing Act

Braxton v. Bank of America. Eingereicht im April 2026.

Eine Wohnungseigentümerin hat die Bank of America und einen in Virginia ansässigen Gutachter verklagt und macht rassistische Diskriminierung unter Verstoß gegen den Fair Housing Act geltend. Die Klägerin trägt vor, sie und ihre Mieter seien bei der Besichtigung anwesend gewesen und das Verhalten des Gutachters habe rassistische Ressentiments erkennen lassen. Der Fall setzt das Muster fort, das frühere Verfahren zur gutachterlichen Voreingenommenheit begründet haben — darunter die Sache *Connolly/Mott v. loanDepot* im District of Maryland und die Sache *Austin v. Miller* in Marin County, Kalifornien —, in denen einzelne Kläger Ansprüche nach dem Fair Housing Act gegen Gutachter, AMCs und Kreditgeber geltend machen.

Bedeutung für Gutachter. Das anhaltende Tempo neuer Klagen wegen gutachterlicher Voreingenommenheit gegen einzelne Gutachter deutet zusammen mit dem Rückzug der Bundesbehörden bei der Durchsetzung des *disparate impact* darauf hin, dass sich das Haftungsfeld verschoben hat: Private Kläger und Durchsetzungsstellen der Einzelstaaten sind nunmehr die wichtigsten Quellen der

Haftung für gutachterliche Voreingenommenheit, nicht mehr die Bundesbehörden. Gutachter sollten ihre Dokumentationspraxis daraufhin überprüfen, dass die Auswahl der Vergleichsobjekte, die Herleitung der Anpassungen und die abschließende Wertabgleichung einer nachträglichen Prüfung eigenständig standhalten, und sollten sich bewusst sein, dass Äußerungen oder Verhaltensweisen während der Besichtigung selbst häufig die tatsächliche Grundlage der Klägervorbringen bilden.

Aktuelle Gerichtsentscheidungen — Europäische Union

Aktuelle Gerichtsentscheidungen • The Valuation Engineer

William Bert Craytor • Jahrgang 1, Ausgabe 1 • Sommer 2026

DOI: 10.67330/r1yn1y02

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Umfang und Quellen

Dieser Überblick stützt sich, soweit zugänglich, auf öffentlich verfügbare Primärquellen — insbesondere auf die amtliche Rechtsprechungsdatenbank des Gerichtshofs der Europäischen Union (curia.europa.eu), HUDOC für den Europäischen Gerichtshof für Menschenrechte, die veröffentlichten Entscheidungen des Bundesfinanzhofs (bundesfinanzhof.de), das Bundesverfassungsgericht (bverfg.de), Légifrance für die französische Rechtsprechung sowie das Upper Tribunal (Lands Chamber) für das Vereinigte Königreich — und auf Sekundärkommentierung in berufsständischen und wissenschaftlichen Publikationen. Entscheidungen in den jeweiligen Landessprachen wurden im Original gelesen und werden hier zusammenfassend beschrieben; die Verfahrensbezeichnungen behalten ihre ursprüngliche Schreibweise und Diakritika bei.

Kostenpflichtige juristische Recherchedatenbanken (juris.de, beck-online, Westlaw UK, Lamy, Aranzadi) werden nicht genutzt. Entscheidungen nationaler Gerichte unterhalb der höchstgerichtlichen Ebene werden selektiv erfasst, wobei Entscheidungen von klarer präjudizieller oder methodischer Bedeutung Vorrang haben. Die hier angeführten Fundstellen sollten vor jeder Verwendung in der Praxis oder für weitere Veröffentlichungen anhand der amtlichen Entscheidungssammlung des jeweiligen Gerichts überprüft werden.

Diese Bestandsaufnahme behandelt Deutschland, Frankreich und das Vereinigte Königreich als die wichtigsten nationalen Jurisdiktionen. Italien, Spanien, die Niederlande und die nordischen Länder werden in dieser Folge nicht gesondert erfasst, sollen aber in nachfolgenden vierteljährlichen Folgen ergänzt werden, sofern der Stoff es rechtfertigt. Die Schweiz wird, obwohl außerhalb der Europäischen Union gelegen, in künftigen Folgen erfasst werden, weil die schweizerischen Bewertungsstandards die europäische Praxis insgesamt methodisch stark beeinflussen.

Europäischer Regulierungsrahmen

Zwei Regulierungsvorhaben auf Ebene der Europäischen Union wirken sich wesentlich auf das Arbeitsumfeld von Gutachtern aus, obwohl keines von ihnen selbst eine Gerichtsentscheidung ist:

KI-Verordnung der EU und automatisierte Bewertungsmodelle. Die Verordnung (EU) 2024/1689 (EU AI Act, KI-Verordnung) ist 2024 in Kraft getreten; die Vorschriften für Hochrisiko-KI-Systeme werden gestaffelt bis August 2026 wirksam. Am 19. Mai 2026 veröffentlichte die Europäische Kommission einen Entwurf von Leitlinien zur Einstufung von Hochrisiko-KI-Systemen zur Konsultation der Interessenträger. Anhang III der KI-Verordnung führt die Kreditwürdigkeitsprüfung natürlicher Personen unter den acht Kategorien vermutungsweise hochriskanter Anwendungsfälle auf, und der Leitlinienentwurf der Kommission entfaltet diese Kategorie im Einzelnen. Da automatisierte Bewertungsmodelle (*automated valuation models*, AVMs), die bei der Hypothekenvergabe,

der Überwachung von Hypothekenportfolios und in der Versicherungszeichnung eingesetzt werden, funktional untrennbar mit der Kreditwürdigkeitsprüfung für die zugrunde liegenden Darlehen verbunden sind, werden AVMs, die von Kreditinstituten oder Versicherern innerhalb der Europäischen Union betrieben werden, in den meisten Konstellationen als Hochrisiko-KI-Systeme einzustufen sein. Zu den Folgen für die Rechtsbefolgung zählen verpflichtende Risikomanagementsysteme, Anforderungen an die Datenverwaltung, automatische Protokollierung mit einer Aufbewahrungsfrist von mindestens sechs Monaten, Mechanismen menschlicher Aufsicht, die es den Betreibern erlauben, Systementscheidungen nachzuvollziehen und zu übersteuern, sowie Pflichten zur Beobachtung nach dem Inverkehrbringen.

Bedeutung für Gutachter. Europäische Gutachter sollten damit rechnen, dass AVM-gestützte Arbeitsabläufe ihrer Auftraggeber auf Kreditgeberseite erheblich strengeren Anforderungen an Dokumentation und Erklärbarkeit unterliegen werden. Soweit Gutachter Modelleingangsdaten, Kalibrierungsdaten oder die menschliche Aufsichtsprüfung von AVM-Ergebnissen beisteuern, werden diese Tätigkeiten selbst zu Bestandteilen der Hochrisiko-KI-Compliance des Kreditgebers.

Gebäuderichtlinie (Neufassung). Die Richtlinie (EU) 2024/1275, die neu gefasste Richtlinie über die Gesamtenergieeffizienz von Gebäuden (*Energy Performance of Buildings Directive*, EPBD), wurde am 8. Mai 2024 im Amtsblatt veröffentlicht und trat am 28. Mai 2024 in Kraft. Die Umsetzungsfrist für die Mitgliedstaaten endet am 29. Mai 2026. Die neu gefasste EPBD verpflichtet die Mitgliedstaaten unter anderem, die Sanierung der energetisch schlechtesten sechzehn Prozent des Nichtwohngebäudebestands und der energetisch schlechtesten dreiundvierzig Prozent des Wohngebäudebestands vorrangig anzugehen; sie legt Mindestanforderungen an die Gesamtenergieeffizienz (MEPS) fest, verlangt Nullemissionsgebäude für öffentliche Neubauten ab 2028 und für alle Neubauten ab 2030 und schreibt ab 2030 eine Bewertung des Lebenszyklus-Treibhauspotenzials für Neubauten vor.

Bedeutung für Gutachter. Die neu gefasste EPBD wird Immobilienbewertungen über mehrere Kanäle beeinflussen: Mindestanforderungen an die Energieeffizienz schaffen eine neue Kategorie regulatorisch bedingter Veralterung; die Einstufung im Energieausweis (*energy performance certificate*, EPC) wird zu einer wesentlichen Preisdeterminante sowohl für Wohn- als auch für Gewerbebestände; und die zur Erfüllung erforderlichen Sanierungskosten müssen im Sachwertverfahren und beim Abzug für die verbleibende wirtschaftliche Restnutzungsdauer berücksichtigt werden. Gutachter sollten zudem damit rechnen, dass Kreditgeber, die der Wohnimmobilienkreditrichtlinie (*Mortgage Credit Directive*) unterliegen, in Beleihungsbewertungen zunehmend eine ausdrückliche Behandlung der EPC-Einstufung und der MEPS-bedingten Investitionsbedarfe verlangen werden.

Hinweise für nachfolgende Folgen

Die auf diese Bestandsaufnahme folgenden vierteljährlichen Folgen konzentrieren sich auf Veränderungen: neue Entscheidungen, die Erledigung der hier bezeichneten anhängigen Verfahren (insbesondere die Behandlung von 1 BvR 472/26 durch das Bundesverfassungsgericht) sowie Umsetzungsgesetzgebung der Mitgliedstaaten zur Überführung der neu gefassten EPBD in nationales Recht. Der schweizerische Rechtsrahmen, die Behandlung italienischer und spanischer höchstrichterlicher Entscheidungen sowie Entwicklungen beim französischen *Conseil constitutionnel* werden ergänzt, sofern der Stoff es rechtfertigt.

Fallzusammenfassungen nach Jurisdiktion

Gerichtshof der Europäischen Union

GR Real (Rechtssache C-351/23). Gerichtshof der Europäischen Union, Urteil vom 24. Juni 2025.

Der Gerichtshof entschied über ein Vorabentscheidungsersuchen eines slowakischen Gerichts, das aus einem außergerichtlichen Hypothekenvollstreckungsverfahren hervorging, in dem Verbraucher ihr Familienheim durch Versteigerung verloren, während gerichtliche Verfahren zur Überprüfung möglicherweise missbräuchlicher Vertragsklauseln noch anhängig waren. Der Gerichtshof entschied, dass die Verbraucherschutzgarantien der Richtlinie 93/13/EWG verlangen, dass die nationalen Verfahrensvorschriften eine wirksame gerichtliche Kontrolle der Missbräuchlichkeit von Vertragsklauseln ermöglichen, bevor eine mit einer Hypothek belastete Wohnung im Wege der außergerichtlichen Vollstreckung veräußert werden darf, und dass Verfahren, die eine unumkehrbare Veräußerung vor Abschluss dieser Kontrolle zulassen, mit der Richtlinie unvereinbar sind.

Bedeutung für Gutachter. Die Entscheidung ist unmittelbar keine Bewertungsentscheidung, sie verändert aber das verfahrensrechtliche Umfeld, in dem Beleihungsbewertungen beauftragt und verwendet werden. In mehreren Mitgliedstaaten mit außergerichtlichen Vollstreckungsverfahren (namentlich in der Slowakei, Tschechien, Ungarn und den baltischen Staaten) bewirkt sie praktisch eine Verlangsamung des Verfahrens der Zwangsversteigerung und eine Verlängerung des Zeitraums, in dem auf die Sicherheitenbewertung des Kreditgebers zurückgegriffen wird. Gutachter, die in diesen Jurisdiktionen Sicherheitenbewertungen für Vollstreckungszwecke erstellen, sollten damit rechnen, dass ihre Gutachten länger als bisher in Gebrauch bleiben und an Marktbewertungen zwischen dem ursprünglichen Wertermittlungsstichtag und der schließlich erfolgenden Veräußerung gemessen werden können.

Europäischer Gerichtshof für Menschenrechte

Gebäudeabriss in Baku (Aserbaidshan), Urteil vom Juli 2025. Europäischer Gerichtshof für Menschenrechte, Artikel 1 des Zusatzprotokolls Nr. 1.

Der Europäische Gerichtshof für Menschenrechte stellte eine Verletzung des Rechts auf Achtung des Eigentums im Zusammenhang mit dem Abriss und der Enteignung eines Gebäudes in Baku fest, die von den Behörden unter Berufung auf Erdbebengefahr angeordnet worden waren. Der Straßburger Gerichtshof befand, dass die nationalen Behörden bei der Festsetzung der Entschädigung keine hinreichenden Gründe für die Zurückweisung des von den Beschwerdeführern vorgelegten Wertgutachtens angeführt und ungerechtfertigt allein den Betrag aus dem vom Gericht selbst in Auftrag gegebenen Sachverständigengutachten übernommen hatten. Der Gerichtshof sprach insgesamt 75.000 € für materiellen Schaden, 3.000 € für immateriellen Schaden und 1.055 € für Kosten und Auslagen zu.

Bedeutung für Gutachter. Die Entscheidung bekräftigt eine Rechtsprechungslinie zu Artikel 1 des Zusatzprotokolls Nr. 1, wonach nationale Gerichte einander widersprechende Bewertungsnachweise in Enteignungsverfahren begründet würdigen müssen und die Zurückweisung des Gutachtens eines Beschwerdeführers zugunsten eines gerichtlich beauftragten Sachverständigengutachtens ihrerseits zu begründen ist. Für europäische Gutachter, die im Bereich Enteignung und angemessene Entschädigung tätig sind, unterstreicht der Fall, dass die dokumentarische Belastbarkeit eines Gutachtens — die ausdrückliche Begründung der Auswahl der Vergleichsobjekte, der Herleitung der Anpassungen und der abschließenden Wertabgleichung — selbst zu einer Grundrechtsfrage wird, sobald das

Gutachten zur Stützung eines Anspruchs gegen den Staat vorgelegt wird.

Frankreich

Cour de cassation, Chambre commerciale, 19. Juni 2024, n° 22-24.169. (In der Fachliteratur bis 2025 erörtert.)

Die Cour de cassation (französischer Kassationshof) entschied im Zusammenhang mit erbschaftsteuerlichen Streitigkeiten über die *valeur vénale* (Verkehrswert), dass ein nach dem steuerauslösenden Ereignis (dem Tod des *de cujus*) erfolgter Verkauf des geerbten Grundstücks von den Erben nicht dazu benutzt werden darf, die ursprünglich erklärte *valeur vénale* nachträglich herabzusetzen — auch dann nicht, wenn der nach dem Ereignis erzielte Kaufpreis deutlich niedriger ist und der Verkauf innerhalb weniger Monate nach dem Todesfall stattfindet. Die Entscheidung wendet den Grundsatz an, dass die *valeur vénale* für erbschaftsteuerliche Zwecke auf den Zeitpunkt des steuerauslösenden Ereignisses zu bestimmen und nicht anhand späterer Transaktionen mit demselben Grundstück zu korrigieren ist.

Bedeutung für Gutachter. Die Entscheidung bestätigt eine strenge Stichtagsdisziplin in der französischen erbschaftsteuerlichen Praxis und begrenzt die Bedeutung nachträglicher Verkäufe als rückschauenden Wertnachweis. Gutachter, die französische Erbschaftsteuerbewertungen erstellen, sollten umfassende Vergleichspreisnachweise auf der Grundlage von Transaktionen *am oder vor dem* Wertermittlungstichtag erarbeiten und Verkäufe des Bewertungsobjekts selbst, die nach dem Stichtag stattgefunden haben, nicht als bestätigenden Nachweis anführen. Dieselbe Argumentation dürfte auf die Schenkungsteuer (*droits de donation*) und auf die Vermögensteuer (*IFI*) zu übertragen sein.

Cour de cassation, 3^e Chambre civile, 8. Januar 2026, n° 23-22.803. Gütliche (außergerichtliche) Sachverständigengutachten.

Die Dritte Zivilkammer der Cour de cassation traf eine Entscheidung, die den rechtlichen Wert gütlicher (außergerichtlicher) Sachverständigengutachten im französischen Zivilverfahren neu bestimmt. Die Entscheidung folgt der weiter reichenden Verfahrensreform durch das *décret* vom 18. Juli 2025, das Mediation und einvernehmliche Streitbeilegung im Zivilverfahren zur Regel gemacht hat. Der Kassationshof stellte klar, dass gütliche Sachverständigengutachten, die auf vertragliche Veranlassung nur einer Partei erstellt wurden, gleichwohl als Beweismittel zugelassen und mit Beweiswert versehen werden können, sofern der Grundsatz des kontradiktorischen Verfahrens gewahrt bleibt und das Gericht die Unabhängigkeit und die Methodik des Sachverständigen würdigt.

Bedeutung für Gutachter. Die Entscheidung stärkt die beweisrechtliche Stellung von *expertises immobilières* (Immobilien Gutachten), die außerhalb des Rahmens gerichtlich bestellter Sachverständiger (*experts judiciaires*) erstellt werden, sofern das Gutachten den methodischen Standards des Berufsstands genügt und die Unabhängigkeit des Sachverständigen dokumentiert ist. Für französische Gutachter, die an streitigen Bewertungen in Erb-, Scheidungs- oder Handelssachen arbeiten, schafft die Entscheidung eine tragfähigere Grundlage dafür, dass auftraggeberseitig beauftragte Gutachten in späteren Rechtsstreitigkeiten zitiert und herangezogen werden. Entsprechend wichtiger wird die sorgfältige Dokumentation der Unabhängigkeit (Fehlen von Weisungsgebundenheit, wirtschaftlichem Interesse oder verwandtschaftlicher Beziehung zum Auftraggeber) sowie die Anwendung der von den Berufsstandards vorgeschriebenen Methoden.

Deutschland

Bundesfinanzhof, Urteile vom 12. November 2025 (II R 25/24, II R 31/24, II R 3/25). Die Trilogie zum *Grundsteuer Bundesmodell* — Verfassungsmäßigkeit des reformierten Bewertungssystems der Grundsteuer. Der II. Senat des Bundesfinanzhofs (BFH) entschied in drei parallelen Revisionsverfahren, dass die Vorschriften über das Ertragswertverfahren, das nach dem sogenannten *Bundesmodell* für die Bewertung von Wohngrundstücken in elf deutschen Bundesländern für Zwecke der zum 1. Januar 2025 reformierten Grundsteuer angewandt wird, verfassungsgemäß sind. Kläger waren Wohnungseigentümer in Nordrhein-Westfalen, Sachsen und Berlin, die ihre Feststellungen mit der Begründung angriffen, die schematischen Bewertungsregeln verstießen gegen den allgemeinen Gleichheitssatz (Artikel 3 Absatz 1 des Grundgesetzes) und führten zu systematisch unzutreffenden Werten.

Der BFH führte aus, der Gesetzgeber habe ein Bewertungssystem geschaffen, das *konzeptionell einer Verkehrswertorientierung folgt* und darauf angelegt ist, im Durchschnitt aller bewerteten Objekte den „objektiv-realen Grundstückswert“ innerhalb eines Korridors des *gemeinen Werts* abzubilden. Das Gericht befand, dass der Gesetzgeber den ihm zustehenden Spielraum bei der Abwägung zwischen den Zielen des Bewertungsschemas und den mit einem solchen Schema notwendig verbundenen Ungleichbehandlungen nicht überschritten habe und dass das erhebliche Gewicht, das der Vermeidung eines erneuten „Bewertungsstaus“ durch automatisierte periodische Fortschreibung beigemessen wurde, verfassungsrechtlich zulässig sei. Der BFH betonte ferner, dass den Steuerpflichtigen die durch frühere BFH-Beschlüsse vom 27. Mai 2024 (II B 78/23 und II B 79/23) eröffnete Möglichkeit erhalten bleibt, einen niedrigeren gemeinen Wert durch ein qualifiziertes Gutachten nach § 198 BewG nachzuweisen — mit der Maßgabe, dass der niedrigere Wert durch das Gutachten eines öffentlich bestellten Grundstückssachverständigen unter Anwendung anerkannter Bewertungsverfahren zu belegen ist.

Eine von *Bund der Steuerzahler Deutschland* und *Haus & Grund Deutschland* unterstützte *Verfassungsbeschwerde* ist nunmehr beim Bundesverfassungsgericht unter dem Aktenzeichen 1 BvR 472/26 anhängig.

Bedeutung für Gutachter. Die Trilogie ist die bedeutsamste deutsche Rechtsprechung zur Immobilienbewertung im Berichtszeitraum. Für Gutachter besteht die wichtigste praktische Folge in der gestiegenen Nachfrage nach Gutachten nach § 198 BewG (*Verkehrswertgutachten zum Nachweis des niedrigeren gemeinen Werts*) in Fällen, in denen der nach dem *Bundesmodell* ermittelte *Grundsteuerwert* den tatsächlichen Marktwert des Bewertungsobjekts zu übersteigen scheint. Die ausdrückliche Bestätigung dieses Wegs durch den BFH bedeutet in Verbindung mit seiner allgemeinen Zurückhaltung gegenüber dem *Bundesmodell*, dass das Wertgutachten nunmehr der maßgebliche Mechanismus ist, über den einzelne Steuerpflichtige Entlastung von systematischer Überbewertung erlangen. Gutachter sollten mit einem dauerhaften Anstieg entsprechender Aufträge rechnen und die Einhaltung der Methodik der ImmoWertV 2021 sowie der Anforderungen des § 198 BewG sicherstellen. Die anhängige Verfassungsbeschwerde beim BVerfG (1 BvR 472/26) ist auf weitere Änderungen dieses Rahmens hin zu beobachten und dürfte ein Hauptgegenstand der nächsten vierteljährlichen Folge sein.

Bundesfinanzhof, Urteil vom 7. Oktober 2025, IX R 26/24. Aufteilung des Kaufpreises auf Grund und Boden und Gebäude (*Kaufpreisaufteilung*). In einem einkommensteuerlichen Verfahren über die Bemessungsgrundlage der *Absetzung für Abnutzung* (AfA) eines vermieteten Wohngrundstücks befasste sich der IX. Senat des Bundesfinanzhofs mit der Methodik der Aufteilung

des vertraglichen Kaufpreises auf den nicht abschreibbaren Anteil für *Grund und Boden* und den abschreibbaren *Gebäudeanteil*. Die Kläger hatten die Heranziehung von *Bodenrichtwerten* durch die Vorinstanz mit der Begründung angegriffen, die zugrunde liegende Datenbasis sei für die konkrete Lage des Bewertungsobjekts nicht verlässlich dokumentiert.

Bedeutung für Gutachter. Die Entscheidung setzt eine Linie der BFH-Rechtsprechung fort (namentlich BFH II R 25/24 vom 12. November 2025, oben angeführt, und früher BFH IX R 12/21 vom 20. September 2022), die bestätigt, dass die ImmoWertV die Quelle der anerkannten Bewertungsgrundsätze für die Kaufpreisaufteilung ist. Der Fall zeigt zugleich, wie wichtig die Datengrundlage der *Bodenrichtwerte* ist: Verfügen die *Gutachterausschüsse* für die konkrete Lage nicht über ausreichende Vergleichspreisdaten, kann der *Bodenrichtwert* wegen Mängeln der Datenqualität angegriffen werden. Für Gutachter, die Kaufpreisaufteilungen oder Stellungnahmen zur AfA-Bemessungsgrundlage erstellen, bedeutet dies, dass die Dokumentation der Datengrundlage des *Bodenrichtwerts* — einschließlich der zugrunde liegenden *Kaufpreissammlung* — Teil der belastbaren Analyse wird.

Vereinigtes Königreich

Mehrunnisa Alikhan Ahmed Khan v Wolverhampton City Council. Upper Tribunal (Lands Chamber), 2025.

Das Upper Tribunal (Lands Chamber) bestätigte, dass maßgeblicher Wertermittlungstichtag für die Bemessung der Entschädigung bei einer hoheitlichen Grundstücksentziehung (*compulsory purchase*) der Tag ist, an dem das Eigentum aufgrund einer *general vesting declaration* (GVD) auf die erwerbende Stelle übergeht. Das Tribunal befasste sich zugleich mit dem Verhältnis zwischen zeitnahen und rückschauenden Bewertungen: Im entschiedenen Fall belief sich die unmittelbar nach der GVD erstellte erste Bewertung der Gebietskörperschaft auf 150.000 £, während eine zweite, im Jahr 2025 während des Tribunalverfahrens unter Zugrundelegung des angenommenen Übergangstichtags drei Monate nach der Anzeige des Eigentümers erstellte Bewertung 120.000 £ ergab. Ungeachtet der Bestätigung des rechtlich maßgeblichen Wertermittlungstichtags zog das Tribunal den zeitlich näher am Erlass der GVD ermittelten Betrag vor, was zu einer höheren Zahlung an den Anspruchsteller führte.

Bedeutung für Gutachter. Die Entscheidung ist für britische Gutachter, die im Bereich der hoheitlichen Grundstücksentziehung arbeiten, methodisch lehrreich. Zwar ist der rechtlich maßgebliche Wertermittlungstichtag der Tag des Eigentumsübergangs; die Bevorzugung der zeitnahen Bewertung gegenüber einer Jahre später auf denselben rechtlichen Stichtag erstellten rückschauenden Bewertung zeigt jedoch, dass zeitnahe Marktnachweise praktisches Gewicht haben, auch wenn sie am Übergangstichtag selbst formal nicht zulässig wären. Gutachter sollten dafür sorgen, dass zeitnah zu einer GVD erstellte Gutachten mit hinreichenden Belegen dokumentiert werden, um als maßgebliche Aufzeichnung dienen zu können, und damit rechnen, dass spätere Tribunalverfahren dem zeitnahen Gutachten den Vorzug vor einer rückschauenden Rekonstruktion geben.

Aktuelle Gerichtsentscheidungen — Lateinamerika

Aktuelle Gerichtsentscheidungen • The Valuation Engineer

William Bert Craytor • Jahrgang 1, Ausgabe 1 • Sommer 2026

DOI: 10.67330/ywghg47

Deutsche Übersetzung, mit KI-Unterstützung erstellt und redaktionell geprüft; die englische Fassung ist maßgeblich.

Umfang und Quellen

Dieser Überblick stützt sich, soweit zugänglich, auf öffentlich verfügbare Primärquellen — insbesondere auf die Entscheidungssammlung des Interamerikanischen Gerichtshofs für Menschenrechte (corteidh.or.cr), die Portale des brasilianischen Obersten Bundesgerichts (*Supremo Tribunal Federal*, STF) und des Obersten Gerichtshofs (*Superior Tribunal de Justiça*, STJ) (portal.stf.jus.br, stj.jus.br), das mexikanische *Semanario Judicial de la Federación* (sjfsemanal.scjn.gob.mx), die argentinische *Corte Suprema de Justicia de la Nación* (csjn.gov.ar), das Portal des chilenischen *Poder Judicial* (poderjudicial.cl) und *Diario Constitucional* sowie das kolumbianische *Corte Constitucional* (corteconstitucional.gov.co) — und auf Sekundärkommentierung in berufsständischen und wissenschaftlichen Publikationen. Entscheidungen in spanischer und portugiesischer Sprache wurden im Original gelesen und werden hier zusammenfassend beschrieben; die Verfahrensbezeichnungen behalten ihre ursprüngliche Schreibweise und Diakritika bei.

Kostenpflichtige juristische Recherchedatenbanken (vLex, La Ley, Thomson Reuters Proview Latin America) werden nicht genutzt. Entscheidungen nationaler Gerichte unterhalb der höchststrichterlichen Ebene werden selektiv erfasst, wobei Entscheidungen von klarer präjudizieller oder methodischer Bedeutung Vorrang haben. Die hier angeführten Fundstellen sollten vor jeder Verwendung in der Praxis oder für weitere Veröffentlichungen anhand der amtlichen Entscheidungssammlung des jeweiligen Gerichts überprüft werden.

Diese Bestandsaufnahme behandelt Argentinien, Brasilien, Chile, Kolumbien und Mexiko als die wichtigsten nationalen Jurisdiktionen. Peru, Venezuela, Ecuador, Bolivien, Paraguay, Uruguay sowie die zentralamerikanischen und karibischen Jurisdiktionen werden in dieser Folge nicht gesondert erfasst, sollen aber in nachfolgenden vierteljährlichen Folgen ergänzt werden, sofern der Stoff es rechtfertigt. Die venezolanische Eigentumsrestitution und sanktionsbezogene Bewertungsarbeit mit Kuba-Bezug werden wegen ihrer methodischen Besonderheit und ihrer Bedeutung für die grenzüberschreitende Praxis in einer künftigen Folge behandelt.

Regionaler Kontext

Anders als der Überblick zur Europäischen Union hat dieser lateinamerikanische Überblick keine unmittelbare Entsprechung zu jener supranationalen Regulierungsebene, die dort die KI-Verordnung der EU und die neu gefasste Gebäuderichtlinie bilden. Die Regulierung der Immobilienbewertung in Lateinamerika ist ganz überwiegend national; die wichtigsten supranationalen Einflüsse sind die grundrechtliche Rechtsprechung des Interamerikanischen Gerichtshofs für Menschenrechte zu Artikel 21 der Amerikanischen Menschenrechtskonvention (Recht auf Eigentum) sowie die Übernahme der International Financial Reporting Standards (insbesondere IFRS 13, Bemessung des beizulegenden Zeitwerts) durch die meisten Jurisdiktionen der Region für Zwecke der Rechnungslegung.

Zwei methodische Themen ziehen sich gleichwohl durch die Rechtsprechung der Region und seien hier erwähnt:

Bewertung indigenen Gemeinschaftseigentums. Die eigentümlichste methodische Herausforderung der lateinamerikanischen Bewertungspraxis ist die Bewertung indigenen und afrostämmigen Gemeinschaftseigentums in Restitutions-, Enteignungs- und Entschädigungszusammenhängen. Die Rechtsprechung des Interamerikanischen Gerichtshofs hat durchgängig entschieden, dass indigenes Gemeinschaftseigentum verfassungsrechtlich ebenso geschützt ist wie individuelles Privateigentum; die Methoden, mit denen solches Eigentum bewertet wird — wo es typischerweise keine marktüblichen Transaktionen zwischen unabhängigen Dritten gibt, wo das Eigentum gemeinschaftlich und unveräußerlich gehalten wird und wo die kulturelle und spirituelle Bedeutung des Gebiets integraler Bestandteil seines Werts ist — bleiben jedoch unterentwickelt. Ende 2025 waren vierundzwanzig Urteile des Interamerikanischen Gerichtshofs zugunsten indigener Völker noch offen oder nur teilweise umgesetzt, mehrere davon mit ausstehenden Entschädigungsberechnungen, die von ungeklärter Bewertungsmethodik abhängen.

Inflation, Währung und Indexierung bei der Enteignungsentschädigung. Mehrere Jurisdiktionen der Region, allen voran Argentinien, verlangen, dass Bewertungsmethodik und Entschädigungsberechnung mit hoher Inflation, Fremdwährungs-*tasaciones* und komplexen Indexierungsregimen zurechtkommen. Der argentinische Oberste Gerichtshof hat entschieden, dass eine ursprünglich in Fremdwährung ausgedrückte Enteignungsentschädigung nicht zu einem künstlichen Paritätskurs in Landeswährung umgerechnet werden darf und dass die Bewertung den realen wirtschaftlichen Wert des Grundstücks im Zeitpunkt der Entschädigung abbilden muss. Ähnliche Fragen stellen sich in Venezuela (im Zusammenhang mit Restitutions- und sanktionsbezogenen Bewertungen) und zeitweise in Brasilien in Phasen erhöhter Inflation. Für Gutachter, die über mehrere Jurisdiktionen der Region hinweg tätig sind, sind diese Inflations- und Währungserwägungen in einer Weise integraler Bestandteil der Bewertungsmethodik, für die es in Europa oder den Vereinigten Staaten keine Entsprechung gibt.

Hinweise für nachfolgende Folgen

Die auf diese Bestandsaufnahme folgenden vierteljährlichen Folgen konzentrieren sich auf Veränderungen: neue Entscheidungen des Interamerikanischen Gerichtshofs, die Erledigung der hier bezeichneten anhängigen Verfahren, Entwicklungen bei der Umsetzung der kolumbianischen *Jurisdicción Agraria y Rural* bis zu ihrem operativen Start 2027 sowie Entwicklungen in den Mitgliedstaaten zur Anwendung der Zeitwertbemessung nach IFRS 13. Peru, Venezuela, Ecuador und die zentralamerikanischen Jurisdiktionen werden ergänzt, sofern der Stoff es rechtfertigt. Besondere Aufmerksamkeit gilt der methodischen Fortentwicklung der Bewertung indigenen Gemeinschaftseigentums, wo die materiellen Entscheidungsgründe des Interamerikanischen Gerichtshofs der praktisch verfügbaren Bewertungsmethodik für Gutachter und Entschädigungsbehörden vorausgeeilt sind.

Fallzusammenfassungen nach Jurisdiktion

Interamerikanischer Gerichtshof für Menschenrechte

Maya Q'eqchi' Agua Caliente vs. Guatemala und *Comunidad Garífuna de San Juan y sus Miembros vs. Honduras*. Interamerikanischer Gerichtshof für Menschenrechte, Urteile von

2024, in der regionalen Rechtsprechung bis 2025 erörtert.

In zwei zeitgleichen Urteilen entschied der Interamerikanische Gerichtshof über die Pflichten Guatemalas und Honduras', angestammte Gebiete indigener und afrostämmiger Gemeinschaften zu demarkieren, mit Titeln zu versehen und vor Entzug zu schützen. Die Gemeinschaft Maya Q'eqchi' Agua Caliente hatte einen endgültigen Eigentumstitel an ihrem Gebiet erst nach fünfundvierzig Jahren behördlicher und gerichtlicher Bemühungen erlangt; die Garífuna-Gemeinschaft von San Juan in Honduras sah sich dem fortdauernden Entzug angestammter Küstenländereien ausgesetzt. Beide Urteile befassten sich mit den Pflichten des Staates aus Artikel 21 (Recht auf Eigentum) der Amerikanischen Menschenrechtskonvention, indigenes Gemeinschaftseigentum anzuerkennen, zu demarkieren und zu schützen. Kommentatoren aus der Region haben beobachtet, dass der Restitutionsrahmen des Interamerikanischen Gerichtshofs in diesen Entscheidungen etwas restriktiver ausfällt als in früheren Urteilen, insbesondere in der Behandlung „gutgläubiger Dritter“ (*terceros inocentes*), die ohne Kenntnis des indigenen Anspruchs überlappende Rechte erworben haben.

Bedeutung für Gutachter. Die Entscheidungen bekräftigen, dass indigenes und afrostämmiges Gemeinschaftseigentum im gesamten interamerikanischen System verfassungsrechtlich geschützt ist und dass sich die Bewertung solchen Eigentums in Restitutions- oder Entschädigungszusammenhängen nicht ohne methodische Ergänzung auf herkömmliche Vergleichspreise aus Transaktionen zwischen unabhängigen Dritten stützen kann — die es typischerweise gar nicht gibt. Gutachter, die in indigenen Restitutionsverfahren beauftragt werden, sollten damit rechnen, dass die Bewertung drei verschiedene Bestandteile zusammenführen muss: den zugrunde liegenden Marktwert von Land vergleichbarer physischer Beschaffenheit in der Region; die Opportunitätskosten und Verbesserungen, die der indigenen Nutzung zuzurechnen sind; und die nichtökonomische kulturelle und spirituelle Bedeutung des Gebiets, die der Interamerikanische Gerichtshof in seiner Rechtsprechung als entschädigungsfähige Dimension des Eigentumsrechts behandelt. Die Methodik für den dritten Bestandteil bleibt unterentwickelt und ist fortdauernder Gegenstand der Erwägungen des Gerichtshofs.

Mapuche / Consejo de Todas las Tierras vs. Chile. Interamerikanischer Gerichtshof für Menschenrechte, Urteil von 2024 (in der Rechtsprechung 2025 erörtert).

Der Interamerikanische Gerichtshof verurteilte den chilenischen Staat wegen Menschenrechtsverletzungen gegenüber 135 Angehörigen des Mapuche-Volkes im Zusammenhang mit den territorialen Rückforderungsaktionen des Consejo de Todas las Tierras (1989–1992) und der Anwendung der Antiterrorgesetzgebung auf indigene Gebietsansprüche. Das Urteil führt die Doktrin aus *Norín Catrimán und andere vs. Chile* (2014) fort und erkennt ausdrücklich das Recht indigener Völker auf Selbstbestimmung im Zusammenhang mit Gebietsansprüchen an.

Bedeutung für Gutachter. Zwar betreffen die tragenden Entscheidungsgründe strafverfahrensrechtliche Schutzrechte und nicht unmittelbar die Bewertungsmethodik; die Anerkennung indigener Selbstbestimmung im Zusammenhang mit dem Gebiet schafft jedoch eine verfahrensrechtliche Überlagerung, die Bewertungsarbeiten für die Restitution indigener Gebiete in Chile und verwandte Zusammenhänge berührt. Gutachter, die in solchen Zusammenhängen beauftragt werden, sollten sicherstellen, dass die Methodik angemessene Konsultationsverfahren einbezieht und die Selbstbestimmung der Gemeinschaft darüber anerkennt, wie ihre territorialen Werte charakterisiert werden.

Argentinien

Argentinische Doktrin zum Tribunal de Tasaciones de la Nación. Gefestigt durch die Rechtsprechung bis 2025.

Der argentinische Oberste Gerichtshof (*Corte Suprema de Justicia de la Nación*, CSJN) und die unteren Bundesgerichte haben die Doktrin zu Bewertungen des Tribunal de Tasaciones de la Nación (TTN) in Enteignungssachen und in Verfahren der *expropiación inversa* (umgekehrte Enteignung) fortentwickelt. Nach gefestigter CSJN-Doktrin (Fallos 313:82, in den Jahren 2024–2025 wiederholt angewandt) kommt TTN-Bewertungen wegen der fachlichen Qualifikation der Mitglieder des Tribunals, der Beweisgrundlage, auf der sie beruhen, und der Einheitlichkeit der Praxis eine Vermutungswirkung zu; ein Abweichen von einer TTN-Bewertung setzt den Nachweis eines offensichtlichen Fehlers oder einer Auslassung voraus. Bundesgerichte, die diese Doktrin in den Jahren 2024–2025 angewandt haben — darunter das Juzgado Federal de Oberá in *Sosa v. Entidad Binacional Yacyretá* (Juni 2024, *expropiación inversa* aufgrund des Yacyretá-Stausees) —, stützen sich weiterhin auf TTN-Bewertungen als maßgeblichen Anker der Entschädigung.

Bedeutung für Gutachter. Für argentinische Gutachter folgt daraus praktisch, dass private Gutachten, die gegen TTN-Bewertungen vorgelegt werden, eine erhebliche Beweislast tragen. Um eine TTN-Bewertung zu verdrängen oder abzuändern, muss das private Gutachten einen methodischen Fehler oder eine Auslassung in der Arbeit des TTN aufzeigen, nicht bloß eine abweichende Bewertungsschlussfolgerung. Gutachter, die solche Gutachten erstellen, sollten die methodische Kritik auf die Auswahl der Vergleichsobjekte, die Herleitung der Anpassungen und die Behandlung der örtlichen Marktverhältnisse durch das TTN richten und nicht allein auf das Ausmaß der Wertdifferenz.

Fremdwährungs-tasaciones und Entschädigungsparität. Die CSJN-Doktrin, dass eine ursprünglich in Fremdwährung ausgedrückte Enteignungsentschädigung nicht zu einer künstlichen Parität von eins zu eins in argentinische Pesos umgerechnet werden darf, gilt bis 2025 fort. Die Doktrin schützt enteignete Eigentümer vor Entschädigungen, die aufgrund von Inflation oder Währungspolitik den realen wirtschaftlichen Wert des Grundstücks im Zeitpunkt der Entschädigung verfehlen. Der Grundsatz ist in den Jahren 2024–2025 im Zusammenhang mit langlaufenden Enteignungsverfahren wiederholt bestätigt worden.

Bedeutung für Gutachter. Argentinische Gutachter, die Bewertungen für Enteignungssachen erstellen, die Zeiträume erheblicher Inflation oder währungspolitischer Änderungen umfassen, sollten die Währungsgrundlage der Bewertung ausdrücklich dokumentieren und die angemessene Indexierungsmethodik zwischen dem Wertermittlungstichtag und dem voraussichtlichen Zahlungszeitpunkt erwägen. Gutachten, die Inflations- und Währungsrisiken nicht behandeln, werden Gerichte, die die CSJN-Entschädigungsdoktrin anwenden, kaum überzeugen.

Brasilien

Supremo Tribunal Federal, RE 922.144 (Tema 865 der *repercussão geral*). Brasilianisches Oberstes Bundesgericht, Urteil von 2024 mit fortdauernder Anwendung im Jahr 2025.

Der STF entschied, dass die Entschädigung in Verfahren der *desapropriação* (Enteignung) verfassungsrechtlich mit dem durch Artikel 100 der Bundesverfassung eingerichteten *precatório*-Regime vereinbar ist. Konkret ist die Differenz zwischen dem von der öffentlichen Hand zunächst hinterlegten Betrag (zum Zweck der *imissão provisória na posse*, der vorläufigen Besitzeinweisung) und dem letztlich durch gerichtliches Sachverständigengutachten festgestellten Wert über das *precatório*-

System zu zahlen und nicht als sofortige Ergänzungshinterlegung. Die Entscheidung wägt das Verfassungsgebot einer *justa, prévia e em dinheiro* (gerechten, vorherigen und in Geld zu leistenden) Entschädigung (Artikel 5 Ziffer XXIV) gegen die Anforderungen an eine geordnete Zahlungsabwicklung im *precatório*-System ab.

Bedeutung für Gutachter. Die Entscheidung hat unmittelbare praktische Folgen für die Bewertung der Ersthinterlegung, die Gutachter im Auftrag enteignender Stellen erstellen. Da jede Unterschätzung bei der Ersthinterlegung später über das *precatório*-System ausgeglichen wird (was mit erheblichen Verzögerungen verbunden sein kann), besteht für Kläger ein struktureller Anreiz, Erstbewertungen als zu niedrig anzugreifen. Gutachter, die Bewertungen für die Ersthinterlegung erstellen, sollten die Methodik mit besonderer Sorgfalt dokumentieren, um die Bewertung gegen ein späteres gerichtliches Sachverständigengutachten zu verteidigen, und dabei berücksichtigen, dass die Lücke zwischen der Ersthinterlegung und der schließlichen gerichtlichen Festsetzung nach der Rechtsprechung des STF und dem Rahmen des Decreto-Lei 3.365/1941 der Ausgleichszinspflicht (*juros compensatórios*) unterliegt.

Superior Tribunal de Justiça, REsp 2.129.162 (*tema repetitivo*). Brazilianischer Oberster Gerichtshof, Urteil vom Mai 2025.

Der STJ entschied eine Serienstreitfrage zu Anwaltshonoraren in Fällen, in denen die enteignende Stelle von der *desapropriação*-Klage Abstand nimmt. Das Gericht stellte fest, dass sich die Honorare in Fällen der Klagerücknahme nach Artikel 27 § 1 des Decreto-Lei 3.365/1941 richten und nicht nach den allgemeinen Verfahrensvorschriften des *Código de Processo Civil*. Der Fall ging aus einer Klage von Cemig (Companhia Energética de Minas Gerais) hervor, mit der eine Verwaltungsdienstbarkeit für eine Stromverteilungsleitung erstrebt wurde.

Bedeutung für Gutachter. Die Entscheidung ist verfahrensrechtlicher und nicht methodischer Natur, hat aber Auswirkungen auf die Ausgestaltung der Beauftragung von Gutachtern in Verfahren über Verwaltungsdienstbarkeiten und *desapropriação*. Gutachter, die in solchen Verfahren als gerichtliche Sachverständige (*peritos judiciais*) herangezogen werden, sollten sicherstellen, dass Honorarvereinbarungen unter Bezugnahme auf den einschlägigen Rahmen des Decreto-Lei und nicht auf die allgemeinen Verfahrensvorschriften dokumentiert werden.

Superior Tribunal de Justiça, Doktrin zu den *juros compensatórios*. Fortdauernde Anwendung 2025.

Der STJ hält an seiner Doktrin fest, dass Ausgleichszinsen (*juros compensatórios*) in *desapropriação*-Sachen erst nach der gerichtlichen Entscheidung über das Eigentum am enteigneten Grundstück anfallen. Die Zinsen betreffen die Differenz zwischen der Ersthinterlegung und dem endgültigen Entschädigungsbetrag, und zwar zu dem historisch festgelegten Satz von zwölf Prozent jährlich (Súmula 618 des STF); sie dienen dazu, den enteigneten Eigentümer für den vorzeitigen Besitzverlust vor Zahlung der endgültigen Entschädigung zu entschädigen.

Bedeutung für Gutachter. Brasilianische Gutachter sollten damit rechnen, dass die von ihnen im gerichtlichen Sachverständigengutachten festgestellte Bewertungslücke — zwischen der Ersthinterlegung der öffentlichen Hand und dem ermittelten Marktwert — zur Bemessungsgrundlage der Ausgleichszinsen wird. Die wirtschaftliche Bedeutung dieser Lücke ist daher größer, als es die nominale Differenz nahelegt, insbesondere in langen Verfahren.

Chile

Corte Suprema, Rol N° 31.423-2025. Chilenischer Oberster Gerichtshof, Urteil vom September 2025.

Der chilenische Oberste Gerichtshof wies die Kassationsbeschwerde eines Immobilienentwicklers (*inmobiliaria*) gegen den Staatsfiskus (*Fisco de Chile*) in einem besonderen Enteignungsverfahren nach Artikel 12 des Decreto-Ley N° 2.186 zurück. Der Kläger hatte geltend gemacht, die von der Sachverständigenkommission (*Comisión de Peritos*) festgesetzte Bewertung bilde die höchste und beste Nutzung (*mayor y mejor uso*) des Grundstücks sowie den Wert vorhandener Bauten und Nebenanlagen nicht ab; zusätzlich hatte er eine Entschädigung für entgangenen Gewinn (*lucro cesante*) unter Berufung auf einen bestehenden Mietvertrag verlangt, der künftige Mieteinnahmen erbracht hätte. Das Gericht entschied, das Vorbringen des Klägers laufe auf eine Meinungsverschiedenheit über die Beweiswürdigung der Vorinstanz hinaus und nicht auf eine tatsächliche Verletzung der für die Bewertung maßgeblichen Rechtsnormen; die Begründung der Vorinstanz — die sich auf die Nachweise der Sachverständigenkommission in Verbindung mit Urkunden- und Zeugenbeweisen stützte — sei hinreichend fundiert.

Bedeutung für Gutachter. Die Entscheidung veranschaulicht die erhebliche Zurückhaltung, die die Doktrin des chilenischen Obersten Gerichtshofs der Bewertung der Sachverständigenkommission in Enteignungsverfahren entgegenbringt, sowie den begrenzten Raum für eine Kassationsüberprüfung von Bewertungsergebnissen allein wegen methodischer Meinungsverschiedenheiten. Gutachter, die in chilenischen Enteignungssachen gegenläufige Bewertungen erstellen, sollten sich auf nachweisbare methodische Fehler in der Arbeit der Kommission konzentrieren statt auf eine alternative Bewertungsargumentation, und sich bewusst sein, dass Ansprüche auf *lucro cesante* aus bestehenden Mietverhältnissen erhebliche Beweisanforderungen zu erfüllen haben.

Corte Suprema (Enteignung in Coquimbo). Chilenischer Oberster Gerichtshof, Urteil vom 17. Juli 2025.

Der chilenische Oberste Gerichtshof bestätigte eine Berufungsentscheidung, die die Entschädigung in einer Enteignungssache in Coquimbo auf UF 2.969,75 (auf Grundlage des Gutachtens eines privaten Gutachters) gegenüber dem vom Staat vertretenen Betrag von UF 2.349,07 (auf Grundlage eingetragener Verkäufe aus dem *Conservador de Bienes Raíces*) an hob. Das Gericht betonte, dass die Heranziehung des privaten Gutachtens durch das Berufungsgericht — gewürdigt zusammen mit Zeugenbeweisen und einem weiteren Wertgutachten zu einem nahe gelegenen Grundstück — eine zulässige Ausübung der *sana crítica* (begründeten Beweiswürdigung) darstelle und dass sich die Kassationsüberprüfung nicht auf eine Nachprüfung dieser Beweiswürdigung erstrecke.

Bedeutung für Gutachter. Die Coquimbo-Entscheidung ist das Gegenstück zu Rol N° 31.423-2025: Ebenso wie sich die Bewertung der Kommission durchsetzt, wenn ein Entwickler sie angreift, kann sich auch die höhere Bewertung eines privaten Gutachters durchsetzen, wenn sie von einer angemessenen methodischen Begründung und bestätigenden Nachweisen getragen wird. Die praktische Lehre für chilenische Gutachter lautet, dass Methodik und Belegunterlagen eines privaten Gutachtens ausschlaggebend sind und dass Nachweise aus eingetragenen Verkäufen allein nicht genügen, um ein gut begründetes privates Gutachten zu verdrängen, das auf zusätzliche Vergleichsobjekte und Kontextnachweise gestützt ist.

Tribunal Constitucional, Oktober 2025. *Tasación* von *bienes raíces* nach dem Steuergesetzbuch.

Das chilenische Tribunal Constitucional wies eine verfassungsrechtliche Rüge gegen die Vorschriften des *Código Tributario* über die *tasación* von Grundstücken (*bienes raíces*) für Zwecke des *impuesto territorial* (Grundsteuer) zurück. Abweichende Richter hielten die angegriffenen Normen für unzureichend, soweit es die Möglichkeiten der vollstreckten Partei betrifft, gegen die *tasación* Einwendungen zu erheben; dies beeinträchtigt die verfassungsrechtliche Garantie eines rationalen und gerechten Verfahrens.

Bedeutung für Gutachter. Die Entscheidung erhält den bestehenden administrativen Rahmen der *tasación* für die chilenische Grundsteuer, doch die abweichende Begründung benennt eine Schwachstelle, die in künftigen verfassungs- oder verwaltungsrechtlichen Angriffen erneut aufgegriffen werden kann. Chilenische Gutachter, die Bewertungen im Zusammenhang mit Grundsteuerstreitigkeiten erstellen, sollten das politische Umfeld nach der Entscheidung beobachten, da administrative Reformen zur Verbesserung der Einwendungsmöglichkeiten der Steuerpflichtigen plausibel erscheinen.

Kolumbien

Corte Constitucional, Billigung der *Jurisdicción Agraria y Rural*. Kolumbianisches Verfassungsgericht, Entscheidung von 2025.

Das kolumbianische Verfassungsgericht billigte in der Normenkontrolle die Gesetzgebung zur Schaffung der neuen *Jurisdicción Agraria y Rural*, einer spezialisierten Gerichtsbarkeit für ländlichen Grundbesitz innerhalb der kolumbianischen Judikative. Das Gericht billigte die Kernvorschriften über den Gegenstand der Reform, ihre Eingliederung in die Judikative, die Errichtung spezialisierter Gerichte und Kammern sowie die Haushaltsaspekte. Die neue Gerichtsbarkeit soll 2027 ihre Arbeit aufnehmen und ist darauf ausgelegt, Streitigkeiten über Landrestitution, agrarische Eigentumsstreitigkeiten und Fragen aus der Umsetzung des Friedensabkommens von 2016 zu entscheiden. Zum Zeitpunkt der Billigung waren im bestehenden Verwaltungs- und Gerichtssystem agrarische Streitigkeiten über mehr als 61.000 Hektar anhängig.

Bedeutung für Gutachter. Die neue Gerichtsbarkeit wird zum wichtigsten Forum für Bewertungstreitigkeiten aus Landrestitutionsverfahren und ländlichen Eigentumssachen in Kolumbien werden. Kolumbianische Gutachter, die im Restitutionsbereich tätig sind (einschließlich derjenigen, die für die *Unidad de Restitución de Tierras* und das *Instituto Geográfico Agustín Codazzi* arbeiten), sollten damit rechnen, dass die neuen Gerichte eine eigene Bewertungsrechtsprechung entwickeln und dass sich die methodischen Standards aus dem gegenwärtig verwaltungsgeprägten Rahmen fortentwickeln. Praktiker sollten den operativen Start 2027 und die Entwicklung des anfänglichen Geschäftsanfalls der Gerichte beobachten.

Bestehender IGAC-gestützter Bewertungsrahmen für die Landrestitution. Fortdauernde Anwendung 2025.

Die kolumbianische Restitutionspraxis stützt sich weiterhin auf den vom *Instituto Geográfico Agustín Codazzi* (IGAC) erstellten *avalúo comercial* (Verkehrswertgutachten) als wesentliche Grundlage der Entschädigung in Restitutions- und Ersatzkonstellationen, ergänzt, soweit erforderlich, durch *avalúos* der *Lonjas de Propiedad Raíz*. Der Rahmen ist in der Entscheidung SU-034 von 2018 des Verfassungsgerichts (zu den Verfahrensanforderungen der Umsetzung) sowie in nachfolgenden Entscheidungen einschließlich T-120 von 2024 (zum Verhältnis zwischen IGAC-Bewertung und individuellen Präferenzen der Begünstigten bei Entschädigung in Natur) niedergelegt.

Bedeutung für Gutachter. Für Gutachter, die in der kolumbianischen Landrestitution tätig sind,

rückt der institutionelle Rahmen die IGAC-Bewertungen in das operative Zentrum der Entschädigungsberechnung, wobei die *avalúos* der *Lonja de Propiedad Raíz* als ergänzender Mechanismus oder als Mittel der Streitbeilegung dienen. Die kommende *Jurisdicción Agraria y Rural* kann diese institutionelle Architektur verändern, und das relative Gewicht von IGAC- gegenüber Lonja-Bewertungen gehört zu den Fragen, die die neuen Gerichte voraussichtlich fortentwickeln werden.

Mexiko

Suprema Corte de Justicia de la Nación, ISR-Pflicht auf die Differenz zwischen *avalúo* und Kaufpreis. Mexikanischer Oberster Gerichtshof, Entscheidung Anfang 2026 bekannt gegeben.

Das Plenum des mexikanischen Obersten Gerichtshofs (SCJN) erklärte die Artikel 125, 130 Ziffer IV und 132 der *Ley del Impuesto sobre la Renta* (ISR, Einkommensteuergesetz) für verfassungsgemäß. Diese sehen vor, dass die Differenz als steuerpflichtiges Einkommen des Erwerbers zu behandeln und damit der ISR zu unterwerfen ist, wenn der *avalúo* beim Erwerb einer Immobilie den vertraglichen Kaufpreis um mehr als zehn Prozent übersteigt. Das Gericht entschied, dass die Regelung nicht gegen die verfassungsrechtlichen Grundsätze der Verhältnismäßigkeit und der Gleichmäßigkeit der Besteuerung verstößt, weil der Erwerb einer Immobilie zu einem Preis wesentlich unterhalb des Marktwerts eine objektive Vermehrung des Vermögens des Erwerbers darstellt und daher zu Recht einen steuerbaren wirtschaftlichen Vorteil bildet.

Bedeutung für Gutachter. Die Entscheidung hat erhebliche praktische Folgen für mexikanische Immobilientransaktionen und für die sie begleitende Bewertungspraxis. *Avalúos*, die von *corredores públicos* oder anderen dazu befugten Berufsträgern erstellt werden, haben nunmehr unmittelbare steuerliche Folgen für Erwerber, wenn der *avalúo* den vertraglichen Preis um mehr als zehn Prozent übersteigt. Gutachter sollten mit einer verschärften Prüfung der Methodik in Transaktions-*avalúos* und mit einer steigenden Nachfrage nach *avalúos* rechnen, die die Grundlage der Vergleichspreise besonders sorgfältig dokumentieren. Die Entscheidung erhöht zudem mittelbar das Gewicht der methodischen Wahl zwischen alternativen Bewertungsverfahren, da diese Wahl nunmehr die Steuerlast des Erwerbers beeinflussen kann.

Suprema Corte de Justicia de la Nación, Amparo en revisión 391/2024. Zweite Kammer des mexikanischen Obersten Gerichtshofs, Urteil vom 25. Juni 2025.

Die Zweite Kammer des mexikanischen Obersten Gerichtshofs entschied über eine Verfassungsfrage eines Bergbauunternehmens gegen ein Dekret, das den rechtlichen Rahmen für die Enteignung von Bergbaufeldern und damit verbundene Nebenrechte betrifft (vorübergehende Inanspruchnahme, Dienstbarkeiten, Vorrechte bei Wasserkonzessionen und Regelungen zur Verlängerung von Pachtverhältnissen). Das Gericht entschied, dass das Dekret nicht gegen das Rückwirkungsverbot verstößt, weil die betroffenen Vorschriften bloße Anwartschaften auf künftige Rechte und nicht erworbene Rechte betreffen.

Bedeutung für Gutachter. Die Entscheidung ist für die Bewertung mineralischer Rechte und von Bergbauimmobilien in Mexiko bedeutsam. Die Einordnung von Rechten auf Verlängerung von Bergbaufeldern und verwandter Regelungen als „bloße Anwartschaften“ statt als erworbene Rechte wirkt sich auf die Bewertung von Bergbaukonzessionen und auf die Einschätzung des regulatorischen Risikos in deren Bewertung aus. Gutachter, die Bewertungen von Bergbauobjekten oder verwandten Rechten in Mexiko erstellen, sollten die regulatorische Risikodimension der Bewertung ausdrücklich behandeln und es vermeiden, bedingte oder von einer Verlängerung abhängige Werte so zu

behandeln, als seien sie voll erworben.