

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

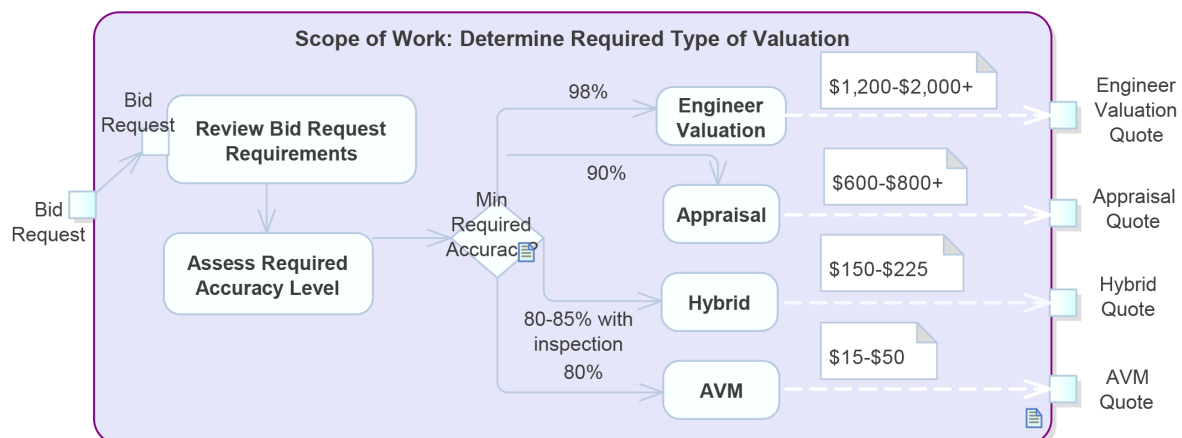
Zur Vermeidung von Mehrdeutigkeit wird der Begriff „Valuation Engineer“ (kurz „VE“) verwendet, um eine Person zu bezeichnen, die über die Fachkompetenz eines lizenzierten Gutachters verfügt, verbunden mit fortgeschrittenen Kenntnissen, Fähigkeiten und Erfahrungen im Einsatz von Data Mining, Statistik, 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 R2 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 R2/CVR2 aufbewahrt, sodass der Verlauf über die Läufe hinweg verglichen werden kann. Eine Version, deren R2 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:

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

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 n_c 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 $f_{j,k}^j g_k^j$ 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.

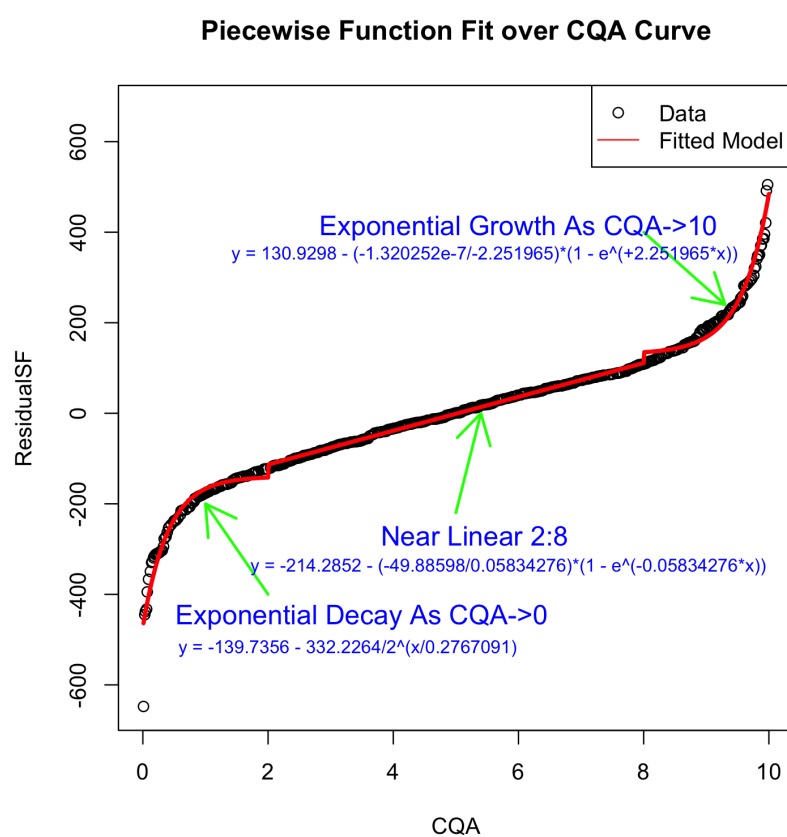
Darüber hinaus verdeutlicht die Anekdote über eine relativ wohlhabende Person, die in ein hochwertiges Haus in einem Wohngebiet mittlerer Preisklasse investiert, in dem sie sich wohlfühlt und aus verschiedenen Gründen bevorzugt leben möchte, einen weiteren zentralen Aspekt der Immobilienökonomie — das 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.