

残差制約アプローチ (RCA): 枠組みとプロトコル

2025 年 3 月 5 日に Zenodo で最初に公表された論文の改訂版の日本語訳。人工知能の支援を得て翻訳し、編集部が校閲した。英語版が正本である。

William Bert Craytor, SRA

第 1 巻第 1 号 · 2026 年夏号

発行日: 2026 年 7 月 1 日

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

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

本論文はオープンアクセス論文であり、[クリエイティブ・コモンズ表示 4.0 国際ライセンス \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) の条件に従って配布される。

要旨. 従来の鑑定評価手法の重大な限界は、ペア売買分析のような旧来の手作業による技法に依存している点にある。これらの技法は、不動産の特性の価値寄与 (value contribution) を、あたかも存在しないかのように素通りしてしまう。従来の取引事例比較表には価値寄与の欄がなく、補正額の欄しかないのである。しかし実際には、価値寄与こそが、従来型の鑑定評価者による過大評価および過小評価を防ぐ数学的制約に不可欠な基礎を与える。

残差制約アプローチ (Residual Constraint Approach, RCA) は、多変量適応回帰スプライン (Multivariate Adaptive Regression Splines, MARS) と、特性の価値寄与に対する厳密な数学的制約とを併せて用いる多段階の評価過程を統合することにより、従来の取引事例比較法 (Sales Comparison Approach, SCA) を発展させる。とりわけ問題となるのは、測定されていない特性、すなわち美観、デザイン、状態、品質といった潜在変数であり、その価値は通常、従来型の鑑定評価者の主観的判断とバイアスに左右される。

RCA における数学的制約は、測定可能な不動産の特性が売却価格に及ぼす価値寄与を推定するために、MARS 回帰を専門的に適用することに依拠する。しかしこの推定は通常、実際の売却価格の約 80% を説明するにとどまり、デザイン、機能的効用、状態、品質のように主観的に評価される特性は含まれない。サンフランシスコ・ベイエリアでは、これらの主観的な構成要素を表す残差が不動産の総価値の平均およそ 20% を占める。この 80/20 という数値は、典型的なデータ条件下におけるサンフランシスコ・ベイエリアでの著者の経験を反映したものであり、市場を問わず不変の定数と読むべきではない。達成可能な R^2 がより低い市場において本枠組みが示す対応は、その限界を受け入れてより広い誤差範囲を報告することではなく、測定されていない価値決定要因を特定して測定することである — 沿岸地域における海までの距離、丘陵地の住宅地における標高、そして MLS データが提供しない場合に外部の情報源から収集する同種の変数がその例である。他の文脈では MARS の残差はしばしば推定誤差として扱われるが、不動産評価においては潜在変数の価値の間接的な尺度として機能する。当初は実用的でないように見えるかもしれないが、各構成要素の寄与の合計が残差に一致するかぎり、残差は最終的な評価結果に影響を与えることなく、意味のある記述的構成要素へ分解できることが証明によって示される。

この方法論は、著者による 20 年にわたる実証的な適用から生まれ、R および Python 環境での MARS の反復的な実装を通じて発展してきた。

キーワード: 不動産評価, 住宅鑑定評価, 残差制約アプローチ, MARS, 多変量適応回帰スプライン, 取引事例比較法, 残差分析, 潜在変数, 自動評価

目次

1 序論

2 背景	6
3 評価エンジニア	8
4 前提	9
5 RCA の作業手順	11
5.1 MARS 分析の種類	11
5.2 処理の作業手順	11
5.3 データの作業手順	12
5.3.1 MARS のためのデータ要件	12
5.3.2 プロトコル: 成約リスティング	13
5.4 プロジェクトのフォルダ構成	15
5.4.1 プロトコル: 複数回の MARS(earth) 実行	16
5.5 中核となる Excel ワークブック	18
5.6 バックアップ	18
6 ワークブックの表記法	19
6.1 記述次元	19
6.1.1 プロジェクト次元 (“a”)	20
6.1.2 実行バージョン次元 (“r”)	21
6.1.3 作業段階 (“w”)	21
6.1.4 Excel ワークブックのシート (“s”)	22
6.1.5 不動産リスティング (“p”) — MLSData の行	22
6.1.6 MLS 変数 (“v”) — MLSData の列	22
6.1.7 最終的な表記の例	23
6.1.8 正当化	23
6.1.9 行列は使わないのか	23
6.1.10 第一の単純化	24
6.1.11 第二の単純化	24
7 RCA 取引事例比較表	26
8 変数	29
8.1 変数の由来オントロジー	29
8.2 変数の用途オントロジー	30
8.3 2 次および 3 次の項	31
8.3.1 補正変数	32
8.3.2 集約変数	32

8.3.3 特殊な変数と表記法	33
8.3.4 残差の分解	33
9 RCA の中核となる証明	35
10 MARS モデル	36
10.1 モデル変数と交互作用の「 g 」関数	36
10.2 g 関数の添字づけ	36
10.3 グラフの次元	37
10.4 最終的なモデル式	37
10.5 モデルをグラフ化するのに必要な図の数	38
11 データの特性	41
11.1 残差の目的	41
11.2 近似的な回帰変数としての状態評価と品質評価	41
11.3 データの誤り	42
11.4 正確性とは何か	43
11.5 異常値としての対象不動産	45
11.6 予期しないデータに対処するうえでの CVR2 の重要性	45
12 残差	47
12.1 順位づけと採点	47
12.2 CQA-残差曲線	50
12.3 CQA 曲線の特性	50
12.3.1 CQA 曲線の構成要素の分析	50
12.3.2 市場の力学と社会的含意	51
12.4 対象不動産の残差	52
12.5 RCA の価値結論	52
12.5.1 一般的な考慮事項	52
12.5.2 検証、監査、苦情	52
13 結論	54

1 序論

本稿は、住宅その他の不動産の市場価値を正確に推定するための現代的アプローチについて、その論理的かつ数学的な基礎を提示する。この手法は、北カリフォルニアの住宅 — とりわけサンフランシスコ・ベイエリア周辺の地域社会にある住宅 — に多変量適応回帰スプライン (Multivariate Adaptive Regression Splines, MARS) ソフトウェアを適用してきた 20 年にわたる経験の中から生まれたものであり、評価の精度において大きな前進をもたらす。MARS は、

不動産の特性と価値との間の非線形な関係を巧みに捉え、同時に最も影響力の大きい変数を効率的に発見して優先順位づけすることにより、評価の質を高める。この能力により、鑑定評価者は新たな市場地域における価値の動態を、従来の手法で可能な速度をはるかに超える速さで解き明かすことができる。MARS は、本稿が提案する残差制約アプローチ (Residual Constraint Approach, RCA) の根幹をなす。RCA という名称は、その厳密な枠組みに由来する。

- RCA は、測定された特性と測定されていない特性の双方について、価値寄与 (value contribution) に厳格な制約を課す。
 1. MARS から導かれるすべての価値寄与は、残差を含めて、各取引事例の純売却価格に合計一致しなければならない。
 2. 残差は、それらの合計が当該取引事例の残差に等しいという条件の下で、値を割り当てた下位構成要素へと分解してよい。
- RCA は、美的な魅力、状態、品質といった主観的特性を推定する際のバイアスを、それらの合計価値を残差に結びつけることによって最小化する。この残差は、有能な評価エンジニア (Valuation Engineer) が MARS モデルを構築したときに正確に定量化される。

RCA は主として、主観的特性が市場価値に大きく影響する住宅不動産を対象として設計されているが、より広い資産評価への応用可能性も有している。米国の住宅不動産市場が推定 45 兆ドルの規模を有することを踏まえれば、正確な鑑定評価は決定的に重要である。ファニーメイ (Fannie Mae) は、標準的な鑑定評価が実際の売却価格から約 5% 乖離すると報告している。もっともこの数値はバイアスを覆い隠している。鑑定評価者はしばしば売却価格に合わせるよう圧力を受けるからである。さらに、鑑定評価の 60%-70% は再融資を目的とするもので — 比較可能な売却価格が存在しない —、RCA が提供するような精度の必要性は明白である。連邦準備制度理事会の 2012 年の分析は、2008 年の金融危機以前に行われた鑑定評価の 15%-20% が売却価格から $\pm 10\%$ を超えて乖離していたと推定した。この範囲は、不動産の種類と市場によって差はあるものの、今日の再融資目的の鑑定評価にはなお当てはまる可能性が高い。なぜ正確性が問題となるのか。ローン・トゥ・バリュー (LTV) の低い取引 — たとえば、信用力の高い買主が 100 万ドルの住宅に 50 万ドルの頭金を入れる場合 — では、貸手は差押えのリスクが小さいとみて鑑定評価を免除することがある。しかし鑑定評価を省いた買主は、過大な支払いをする危険を負い、購入後に初めて瑕疵に気づくことになりかねない。このことは、RCA のような堅牢な評価手法の価値を際立たせる。鑑定評価の正確性を定量化することは依然として困難である。研究者は「市場価値」からの乖離を直接測る指標を持たず、間接的な代理指標に頼らざるを得ない。主観的判断に根ざし、鑑定評価者の能力によって左右される従来の鑑定評価手法は、分析をいっそう複雑にする。用いられてきた方策には次のようなものがある。

1. Tzioumis (2016) は、11 州の 4,782 名の鑑定評価者による 259,436 件の鑑定評価を調査し、3 分の 1 が契約価格の 1% 以内に収まり、半数が契約価格を 1%-3% 上回り、少数が平均して 5% 上回っていることを見出した — 上方バイアスの存在を示唆する結果である。同研究は、鑑定評価者を「市場価値どおり」型 (契約価格に近い評価を行う型) と「市場価値超過」型 (一貫して高い評価を行う型) の二類型に分類した。
2. Krivorotov and LaCour-Little (2020) は、購入目的の鑑定評価の 90% が契約価格に一致するかこれを上回ることを見出し、「アンカーバイアス」 — 望ましいとされる価値に合わせようとする傾向 —、とりわけ再融資の文脈におけるそれを指摘した。

これらの研究は鑑定評価実務におけるバイアスを明らかにしており、そのバイアスは従来のツールの限界によって増幅されている。仮に鑑定評価者が一様に有能かつ中立であったならば、その評価額は収束するはずである。特定の集団による過大評価が持続していること自体が、構造的な問題の兆候である。

上に引用したバイアス研究と、それが指し示す構造的なパターンは、RCA が対処しようとしている従来の取引事例比較実務の、より深いところにある特徴を反映している。従来の方法論は、認識論的な地位を大きく異にする二つの経路の上で機能している。測定可能な特性の経路 — GLA、部屋数、敷地面積、築年数、位置 — においては、鑑定評価者はペア売買分析や取引事例に対する局所線形回帰を通じて、補正額を客観的に導出することが期待される。有能な鑑定評価者はこの作業を行っており、そうして得られる補正額は、少なくともデータに照らしておおむね防御可能である。もっともこの経路においてすら、従来の取引事例比較表が記録するのは補正額だけであって、価値寄与の欄はそもそも存在しない。また従来の補正額は価値寄与の差として計算されるわけでもない — それらはペア売買その他の手続から直接に推定されるのであって、その基礎にある価値寄与が計算されることはない。これに対し RCA では、あらゆる補正額が、当該特性についての対象不動産と取引事例の価値寄与の差である — その帰結を証明 II が厳密に述べる構造である。主観的特性の経路 — 状態、施工の品質、デザイン、機能的効用、美観 — においては、利用可能な裏づけ手続は乏しいものから貧弱なものまでにとどまる。主観的特性の補正は、直接の取引事例集合の外にある不動産のペア売買分析によって、あるいは同業者の一般的な実務への援用によって、あるいは当該市場における鑑定評価者自身の記録された経験への言及によって裏づけられうる。いずれも専門職基準には形式上適合している。しかしそのどれ一つとして、測定可能な経路が日常的に達成している統計的厳密さには及ばない。

その結果生じるのが、基準が要求するものと方法論が提供しうるものとの間の手続的な非対称性である。基準はあらゆる補正について、測定可能であれ主観的であれ、裏づけを要求する。方法論は一方の経路では堅牢な裏づけを提供するが、他方では、標本数 1 の逸話的なペア売買から、他の鑑定評価者が慣行的に用いている値の引用に至るまでの、より弱い手続の連続体を提供するにとどまる。眺望補正について入念なペア売買分析を行う誠実で有能な鑑定評価者は、根拠なく同業者の実務を引用する者よりも実質的に優れた裏づけを生み出す。しかし標準様式の上で二つの鑑定評価書を比較する審査者は、両者を容易には区別できない。様式が記録するのは補正額であって、その導出の厳密さではないからである。主観的特性についてより強力な手続が実務家に一般には利用可能でなかったがゆえに、方法論はこの非対称性を許容してきた。

RCA は双方の経路を改善するが、その改善の性質は根本的に異なる。測定可能な経路では、MARS 回帰が、ペア売買分析や通常の線形回帰では表現できない非線形の関係、変数間の交互作用、基底関数の構造を捉える。従来の取引事例比較表の 3 件から 9 件の取引事例ではなく、市場地域における適格な売買の全体 — 通常は数百件 — に対して当てはめられることにより、MARS から導かれる価値寄与は、客観的であると同時に、それが置き換える方法論よりも実質的に正確である。主観的な経路では、RCA は従来のいずれの裏づけ手続とも種類の異なることを行う。RCA は個々の主観的特性についての補正額を推定しようとはまったくしない。すべての主観的特性 — 潜在変数 — の寄与は、まとめて残差として現れる。すなわち、各取引事例の実際の純売却価格と MARS による予測価格との差であり、鑑定評価者の入力なしにデータから計算できる量である。その残差を、デザイン、機能的効用、眺望、状態、品質などの名前のついた構成要素へ分解する作業は、合計が残差に一致するよう制約されており、証明 II(第 9 節) により、この制約を満たす分解はいずれも同一の価値結論をもたらす。作業手順に残る判断は、取引事例の順位づけされた残差分布のなかで対象不動産をどこに位置づけるかであり、これは検証可能な順位表に照らして行われ、報告書に明示的に記録される。弱い RCA 鑑定評価はなおありうる — 経験の浅い VE は過適合した MARS モデルを作りうるし、対象不動産をありえない位置に置いたり、データ中の異常を見落としたりしうる — が、その弱さは、モデルと順位表を検分することによって審査者が特定できる種類の弱さであって、標準様式による審査では表に出てこない種類の弱さではない。

再融資の問題を離れても、鑑定評価者は所有者に対して正確な評価額を提供する義務を負う。誤りが数万ドルないし数十万ドルの損失を招きうる高額の購入においてはなおさらである。それにもかかわらず、多くの人は鑑定評価に 600–1,000 ドルを支出することをためらい、その付加価値を疑う。RCA のような先進的手法は、当初は資源集約的であるものの、より高い

精度を約束する。改良と経験の蓄積が進めば、その効率性 — そして採用 — は高まるはずである。

2 背景

著者の学問的背景は数学、計算機科学、統計学にある。職業上はソフトウェアエンジニアであり、鑑定評価と会計に副次的な経験を有する。2002 年以来断続的に鑑定評価を行い、2004 年ごろからは住宅不動産における補正額と価格趨勢を求めるために MARS を適用してきた。

MARS を用いて鑑定評価を行うことは、鑑定評価における「深い経験」にほかならない。MARS は、市場価値に寄与する変数と、それらの値の変化が売却価格に及ぼす影響との双方を、他のいかなる統計的手法よりも詳細に反映したモデルを生成する。したがって、MARS を用いて鑑定評価を行う者は、MARS モデルが生成する変数値と売却価格との因果関係を絶えず探究することになる。

透明性の観点から述べれば、RCA の方法論は次の特徴を備えた住宅地域に対して最もよく適合する。

1. 相互にかなり異なる不動産から成る複雑な近隣地域
 - ・ 古い住宅
 - ・ 改修の品質にばらつきがある
 - ・ 建築様式とデザインが多岐にわたる
 - ・ GLA、敷地面積、標高、眺望といった不動産の属性に変動がある。
2. 複雑な市場条件
 - ・ 急速に変化する市場条件
 - ・ 多様な種類の買主
3. 良質なデータ源: RCA のプロトコルはデータマイニングに基づいており、質の高いデータを豊富に必要とする。データの乏しい地域に用いることもできるが、それに要する時間と費用に見合わない場合がある。
4. 市場価値の正確な推定が求められること。RCA は多くの場合、 $\pm 1-2\%$ の精度を提供できるはずである。評価エンジニアの報告書は、最終的な価値結論の精度に関する条件について所見を述べるべきである。データや情報が不足している場合、あるいはデータマイニングによって市場の一貫したパターンが見出されていない場合、精度は損なわれるので、依頼者にその状況を伝えなければならない。具体的には、高い R^2 および CVR^2 (交差検証 R^2 。MLS データ集合を訓練用と検証用の区分に無作為に分割する操作を繰り返して算出する)、意味のある MARS モデル、そして順位づけされた残差における不動産の CQA(condition-quality-appeal、状態・品質・魅力) 特性の強いパターンが、価値結論の正確性と信頼性を評価する主たる手段である。

過去 20 年余にわたるこれらの RCA プロトコルの開発は、MLSListings Inc.(カリフォルニア州サニーバール) およびその他の MLS の利用可能性に依拠している。これらは現在 MLSListings を通じて利用でき、著者のデータ分析経験の 95% 以上を占める。著者の専門知識は、北カリフォルニアの主要 15 郡にわたる広範な鑑定評価業務に由来し、さらに副次的な 10 郡での経験がこれに加わる。

RCA の方法論は、次の二つの主要な基準を満たす市場において最も効果的である。

1. 質の高い鑑定評価データが豊富に利用できること

2. 次の特徴をもつ複雑で異質な住宅地域であること

- 多様な建築様式
- 程度の異なる改修
- 数十年にわたる開発の履歴

カリフォルニア州各郡における経験

本稿における記述は、以下の北カリフォルニア各郡において不動産を評価してきた経験に基づく。

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

上記の地域は、本質的にシリコンバレーとその周辺の郡を含み、さらに外側へ広がってセントラルバレーやシエラ山麓を含むより農村的な地域にまで及ぶ。付言すべきは、著者が、米国内のあまり開発の進んでいない地域の多くの鑑定評価者が手にしうるものと比べて、常に豊富で相対的に正確なデータにアクセスできてきたということである。

用いている MLS データは RESO 認証を受けており、RCF(RESO Common Format) に準拠している。RCF 認証を受けているにもかかわらず、通常は副次的な重要性しかもたないデータが、一部の不動産について欠落していたり誤っていたりすることがある。これは当初の税務評価官の入力誤り、あるいはその後の販売エージェントによる MLS 入力誤りに起因する。こうした誤りはしばしば一定のパターンを示す。たとえば、当該特性が存在しない場合に単に入力しない、というものである。具体的には、カーポートがないときに「carport spaces」に 0 を入力せず、何も記入しないのである。カーポートが稀な地域では、分析者は、カーポートについて有用な補正額を得るために、データなしを 0 で置き換えることがある。R/earth その他の MARS 実装は欠測データを扱えるが、それをどう扱うかを決めるのは評価エンジニア (Valuation Engineer, VE) である。

MLSListings のデータは、フロリダからハワイに至る他の州の MLS から入手できる。これは、プロトコルとコードが異なる州でどの程度うまく機能するかを検証するのに役立つ。

3 評価エンジニア

定義 1: 評価エンジニア

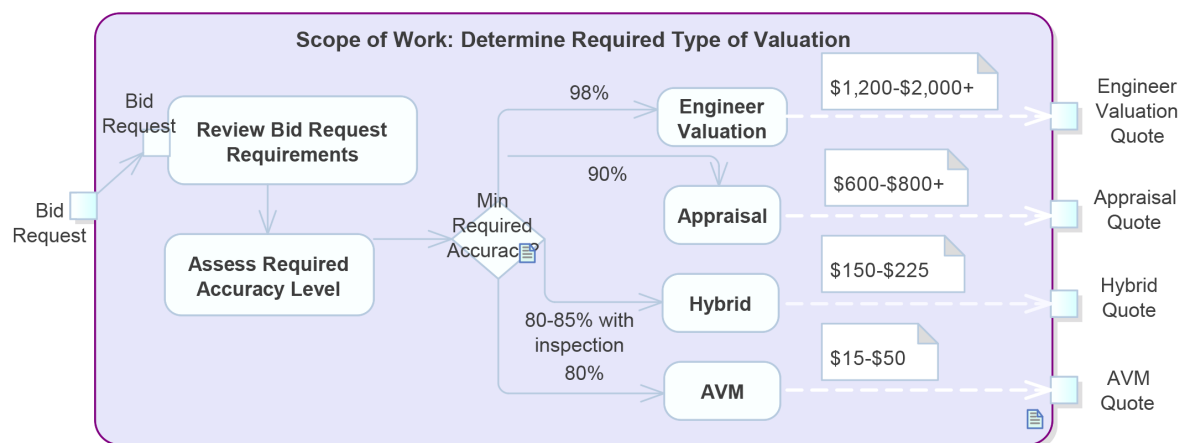
曖昧さを避けるため、「評価エンジニア」(Valuation Engineer、略して「VE」)という用語は、免許を有する鑑定評価者の専門的能力を備えるとともに、複雑な不動産の市場価値を確定するためにデータマイニング、統計学、プログラミング、詳細なプロトコル、および人工知能を活用する高度な知識、技能、経験を併せもつ個人を指すものとして用いる。また、評価は詐欺の可能性が確認されている以上、何らかの形の免許を必要とする業務であることも認識しておく必要がある。

貸手が通常提示する評価の選択肢は次のとおりである。

1. 自動評価モデル (Automated Valuation Models, AVM) — 最も費用の低い選択肢
2. ハイブリッド鑑定評価 — 免許を有する鑑定評価者の監督の下で無免許の調査員を用いるもの
3. 標準的な鑑定評価 — 免許を有する鑑定評価者が行うもの

鑑定評価業務の選択肢として通常考慮されていないのが、エンジニアレベルの鑑定評価である。これを行う技能を備えた鑑定評価者は単純に多くない。しかしそれには需要があつてしかるべきである。エンジニアレベルの鑑定評価は、RCA の方法論 (MARS と制約) を用いることにより、著しく高い精度を提供し、かなり良好な客観性をほぼ保証することができる。

新しい分譲地にある規格的な不動産については、標準的な鑑定評価で概ね足りる。しかし、30 年以上にわたる相当の改修履歴をもつ地域の古い住宅については、エンジニアによる評価が推奨される。両手法の選択は、不動産の特性と市場条件に基づく構造化された意思決定の手順に従うべきである。



4 前提

前提 1: 地理的範囲と不動産の種類

本稿の範囲は、主としてカリフォルニア州の大都市圏および都市部における戸建住宅とコンドミニアムに焦点を当てる。取引事例比較法は他の種類の不動産や他の地域にも適用されるが、農地、集合住宅、商業用不動産の評価においては収益還元法と原価法がより支配的な役割を果たすのが通常である。そうした場合、RCA のアプローチは修正を要するか、適用範囲が限られることがある。

前提 2: データの保存

CSV と Excel は MLS システムで広く用いられており、RCA のプログラムにデータを提供する主たる手段と想定する。ただし、ブローカーフィードを用いて SQL データベースへ継続的にデータを流し込むこともできる。バックアップと中間的な保存には SQL データベースがしばしば用いられ、なかでも大きな役割を果たすのが無償かつオープンソースの SQLite と PostgreSQL である。報告書の断片にはテキストファイルや RTF ファイルが、図表には PNG、JPG、TIFF ファイルが用いられる。なお、Microsoft の SQL Server も多くの利用者を有するが、無償版の「SQL Server Express」は制約が非常に大きく (たとえば CPU コアを 1 つしか使用できない)、他の版は商用利用に高額ライセンス料を要する。カリフォルニアの不動産データベースは非常に大きくなりうるし、結合 (join) には非常に時間がかかりうるので、大規模な結合には 8 コア以上、RAM 128GB の計算機が有用である。SQL データベースは、任意に境界を定めた地域内のすべての不動産を検索する必要がある場合にも重要である。したがって、たとえばクラスター分析によって近隣地域を定義したのであれば、何らかの SQL データベースを用いることになるだろう。しかし、データベースごとに地理情報を扱う組込み関数は異なるため、相互に互換性はない。RCA に関する一連の論考では、複雑な SQL 処理には PostgreSQL を、単純な処理には SQLite を前提とする。

前提 3: MARS 処理には「R/earth」を用いる。

2017 年まで、Salford Systems 社の MARS ソフトウェアは年間 270 ドルという手頃な解であった。しかし Minitab 社による Salford Systems 社の買収後、MARS は同社の SPM パッケージに組み込まれ、年間 16,000 ドルとなった — 多くの小規模事業者にとって手の届かない価格である。幸い、Stephen Milborrow が保守する R 言語の「earth」パッケージ (Milborrow 2024a) が発展し、Minitab の SPM MARS に匹敵するか、それを上回る機能を備えるに至った。earth パッケージにはさらに次の利点がある。無償で入手でき、R の包括的な統計エコシステムと切れ目なく統合され、Pandas を介して Python から利用できるため、さまざまな作業手順に対して優れた柔軟性をもたらす。MARS に関する議論は、R/earth の使用を具体的に前提とする。

前提 4: 表記法

CSV ファイルや Excel スプレッドシートといった外部の情報源に由来する生データは、標準的なラテン文字を用いて表す。これには、マルチプル・リスティング・サービス (MLS) の不動産データ、R/earth パッケージを用いた MARS(多変量適応回帰スプライン) の設定、およびプロジェクト固有のデータが含まれる。この表記法は、R または Python における対応するデータフレームの表現にも及ぶ。

これに対し、不動産の特性およびその種類や部分集合を論じる際には、精度と分析上の明晰さを高めるため、表記はギリシャ文字に移る。この数学的表記はより厳格に見えるかもしれないが、重要な目的に資する。すなわち、抽象的な不動産の特性を、その生データ表現から区別することである。約束事として、ギリシャ文字は一貫して不動産の特性を表し、ラテン文字はスプレッドシート上のデータ構造を示す。たとえば記号 M は、行に不動産を、列に対応する特性を配した MLS データ行列を表す。

前提 5: プログラミング言語

この分野では R と Python(Pandas 併用) が主要な言語であり、速度が重要な場合には C または C++ が用いられる。R/earth 自体は R と C で書かれている。

前提 6: 添字づけ

添字の付け方はプログラミング言語や数学によって異なり、混乱の元となりうる。

- 数学と R では添字は 1 から始まる。
- C、C#、C++、Python では 0 から始まる。

本稿では、一貫性と簡明さを保つため、すべての添字を (Python や C と同様に) 0 から始める。その仕組みは次のとおりである。

- リスト、行列、データフレームについて
 - 添字 0 は「対象不動産」(調査の中心となる不動産) のために予約する。
 - 例: 行が不動産である表において、行 0 は対象不動産であり、列 0 はすべての不動産の一意な識別子を保持する。
 - 実務上の助言: 対象不動産を先頭 (行 0) に置き、続いて売却日の新しいものから古いものへ順に並べ、通し番号 (0、1、2 など) を付す。
- 対象不動産が存在しない場合
 - データ構造が不動産を含まない場合 (たとえば変数や関数のみの場合)、添字 0 は用いず、添字は 1 から始める。
 - R 利用者への注意: R は 1 始まりの添字を用いるので、本稿の数式を R で実装する際には、0 始まりから 1 始まりへ調整するためにすべての添字に 1 を加えること。

5 RCA の作業手順

効果的な分析のためには、RCA の作業手順を理解することが不可欠である。この過程は、分析上の作業と意思決定点の構造化された連なりに従う。

5.1 MARS 分析の類型

MARS 分析の作業手順は、市場分析のみが必要なのか、それとも特定の不動産の評価ないし鑑定評価が求められるのかによって異なる。なお、特定の不動産の鑑定評価には、通常その市場地域の分析が必要となる。

- 市場分析: 郡、市、あるいは画定された地域といった広い範囲を対象とする。このアプローチは特定の対象不動産を扱わないため、価値結論を要しない。作業手順は、モデルの構築、検証、および裏づけとなる分析結果の作成を中心とする。最適なモデルが特定され記録された時点で過程は終了する。
- 単一不動産分析は、市場分析のすべての要素に加えて、特定の不動産の価値の決定を含む。これは従来の鑑定評価手法とは根本的に異なる。従来のアプローチでは、価値は通常 3 件から 9 件の取引事例を補正して加重平均を計算することによって導かれる。これに対し RCA は、主として次の二つの構成要素を通じて価値を決定する。
 1. MARS モデルによる予測価格
 2. 当該不動産に固有の残差の推定値 (および対応する CQA スコア)

RCA における取引事例比較表は、主として、変数ごとの寄与を示すことによって対象不動産と取引事例との価値の差を説明する役割を果たす。RCA の主要な利点はその安定性にある。依頼者が取引事例の選定について議論することはあっても、提示のためにどの取引事例を選ぶかにかかわらず価値結論は一貫している。すべての売買 (15 件であれ 600 件であれ) が同一のモデルによって補正されるからである。古い売買も現在の市場条件に自動的に補正されるが、一般には最近の売買のほうが望ましい。期間が長くなるほど測定されていない市場特性の重みが増すため、時間的な隔たりが大きくなるにつれてモデルが価格要因を捉える力は低下する。

5.2 処理の作業手順

全体的な処理の作業手順は、次の各段階として記述できる。

1. 業務範囲 (Scope of Work, SOW) の確定。
2. プロジェクトの初期フォルダとファイルの設定。
3. MLS データのダウンロード。必要な不動産レコード数と日付範囲の制限については、プロトコル 1.1–1.3 を参照。
4. 設定シートの記入: 回帰変数、交互作用、集約、計算。
5. MARS(R/earth) 回帰を実行して価格モデルを作成。
6. MARS モデルから不動産の価値寄与および残差 (純残差および残差/SF) を生成。
7. 残差/SF により不動産を順位づけ。
8. 各取引事例について CQA スコアを作成。

9. 不動産の全体的な魅力を真に表す品質順位が作成されているという前提の下で、順位表における全体的なパターンの傾向を検討し、対象不動産の最適な位置を見出して CQA スコアを付与する。
10. 対応する残差/SF を、残差/SF · (対象不動産の)GLA として対象不動産に割り当てる。
11. MARS による価格推定値と割り当てた残差を加算し、対象不動産の価値結論を得る。
12. 対象不動産の残差を、記述的な構成要素とその価値寄与へ分解する。
13. 取引事例についてモデルが生成した価値寄与および補正額、ならびにすべての小計と補正後売却価格を計算する。
14. 補正後売却価格は対象不動産の最終的な価値結論に一致する。
15. 取引事例比較表に載せる最も類似した 6–12 件の不動産を選定する。
16. 取引事例の残差総額を残差構成要素ごとの価値寄与に分解し、分解が完了した時点で残差の合計値を 0 とする。
17. すべての小計と取引事例比較表における取引事例の補正後売却価格を再計算する。値を確認する。
18. 取引事例比較表を報告書に取り込む。

RCA において、最終的な価値結論が取引事例比較表上の取引事例の価値から直接に導かれるのではないことに注意されたい。RCA の手法において、取引事例比較表は、対象不動産に類似して見える不動産が最終的な価値結論ないし対象不動産の評価額より高く、あるいは低く売れたのはなぜかを説明するためだけの道具となる。これは別の見方をしなければならない。すなわち、対象不動産の価値結論は MARS モデルによる価格推定値と、評価エンジニアが推定した残差に依存する。しかしこの二つの量は、R/earth 回帰に投入されたすべての不動産に大きく依存しており、その回帰が不動産の残差の順位表を作り、その順位表が対象不動産の CQA スコアを決め、評価エンジニアによる順位表上の位置づけを通じてその残差を決定したのである。取引事例比較表における補正額は、実のところ RCA が課す数学的制約によって決まる。残差の分解はある程度 VE の裁量に委ねられるものの、分解した価値寄与の合計は常に対応する残差に一致しなければならない。証明 2 から分かるように、残差構成要素の個々の値は VE の判断によって変わりうるが、ある不動産についてのそれらの合計は完全に制約されており、したがって VE が残差の分解をどのように変えようとも、当該不動産の補正後売却価格は変化しない。

5.3 データの作業手順

本稿では RCA 手法について高い水準の作業手順のみを提示することに留意されたい。RCA 分析の各段階および下位段階については、別の論考で立ち入って論じる。

5.3.1 MARS のためのデータ要件

売買取引の開始日と、鑑定評価または評価の価格時点に基づく終了日を指定する必要がある。これは通常、1 を超える各交互作用回数について必要となる回帰変数と交互作用の推定個数を考慮して決める。

5.3.2 プロトコル: 成約リスティング

プロトコル 1.1: 必要な MLS 成約リスティング件数の決定

本プロトコルは、MARS 回帰に投入する成約リスティングの推奨最小件数をどのように算出するかを述べる。まず概算を立て、1 次、2 次、場合によっては 3 次の実行によって MARS を走らせて初期モデルを生成し、次に以下の規則に照らして行数を検討し、さらに行数が必要かどうかを判断することになる。本プロトコルは説明変数が 100% 独立であることを前提としているが、不動産においてそのようなことはめったにない。たとえば GLA と部屋数は通常強く相関しており、GLA は敷地面積とも相関を示す可能性が高い。ここで問題となるのが「自由度 (Degrees of Freedom, DOF)」の計算である。真に独立な変数はそれぞれ自由度を 1 つもつ。しかし、いわゆる独立変数の間に従属関係がある場合、DOF を正確に計算するには本稿の範囲を超える分析が必要となる。変数間に共線性が存在する可能性が高いことから、以下の推定値はおそらく多めであることを念頭に置かれたい。

1. 1 次の項 (交互作用なし) については、変数 1 つあたり ~ 15 行という規則がよく機能する
2. 許容する 2 元交互作用 1 つにつき、およそ 15-30 件の不動産を加えるべきである
3. 高次の交互作用 (3 変数以上) については、それぞれ 30 件以上の不動産が必要である。

もっとも、これは厳格な規則ではない。実際に必要な件数は次の要素に左右されうる。

- データの信号対雑音比
- モデル化しようとする関係の複雑さ
- 応答変数のばらつき
- データの品質

例:

次のような場合を考える。

- 主要な変数が 10 個
- 許容する 2 元交互作用が 5 個
- 許容する 3 元交互作用が 2 個

このとき、必要な売買件数の最小値は次のように計算できる。

$$\begin{aligned} N &= (10 \text{ 変数} * 15 \text{ 行/変数}) + \\ &\quad (5 \text{ 2 元交互作用} * 22 \text{ 行/2 元交互作用}) + \\ &\quad (2 \text{ 3 元交互作用} * 30 \text{ 行/3 元交互作用}) \\ &= 320 \text{ 行} \end{aligned}$$

あるいは、次の方法によってもよい。あるデータ集合に対して R/earth を実行した後、モデル中の基底関数の個数に 1 を加えた値を自由度の推定値として用い、これに 20-30 を乗じて必要な入力レコード数を得る。

プロトコル 1.2: 売買期間は買主の嗜好の変化が最小である範囲にとどめること

過去の不動産データを収集する際、市場の選好の変化を適切な変数によって捉えられるのであれば、厳格な期間の制限はない。市場における建築様式の変遷を扱うにあたっては、次の要点を考慮されたい。支配的な建築様式の変化 (たとえばカリフォルニア・ランチ様式から地中海様式への移行) は、次のものを用いてモデル化できる。

- 分譲地の識別子
- MLS エリアコード
- 建築時期の区分

しかし、こうした様式の変化を特定の変数によって十分に捉えられない場合には、重要な制約に直面する。すなわち、売買日は通常は市場条件の補正に用いられるが、これを同時に建築上の選好の統制にも用いると、この二つの異なる効果が交絡してしまう。そのような場合には、次のいずれかを採るのが望ましい。

1. データ収集の対象期間を、新しい建築様式が導入された後に限定する
2. 取引事例の選定を、同一の建築様式を共有する分譲地に限定する

これにより、時間の経過に伴う市場条件の適切な補正を可能にしつつ、不動産の特性の一貫性を保つことができる。あるいは、少なくともこの場合には、データから建築様式を抽出できるという前提の下で、この二つの変数の交互作用を考えることもできる。もっとも、そのような抽出が常に可能とは限らない。

プロトコル 1.3: 売買件数の少ない地域では最善を尽くすこと

売買データに対して MARS によるモデル化を行うには、意味のある分析のためにおよそ 15 件の取引が必要である。それより売買件数が少ない場合 (たとえば 5 件の不動産) には、各レコードを 2 回複製して 15 件という閾値に達させることができる。この方法における重要な留意点は次のとおりである。

1. 過適合を抑えるため、1 次の回帰のみを用いる
2. データ集合が限られすぎているため、交差検証は無視ないし省略する
3. このような状況では MARS がペア売買や標準的な線形回帰の手法を上回る可能性が高いものの、データの複製が不確実性の測定と統計的有意性の妥当性に影響することに留意する

長期的に最善の解決は、実際の売買を追加してデータ集合を拡張することである。しかし限られたデータで進めざるを得ない場合、データの複製は、統計上の限界を認識したうえでの実用的な暫定策となる。

5.4 プロジェクトのフォルダ構成

プロジェクトのフォルダ構成について論じておくのがよいだろう。市場分析と単一不動産鑑定評価とで別々のフォルダを設けるのが望ましい。この二つのフォルダはいずれも同様の下位フォルダ構成をもつ。

```

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

```

第 1 段階の処理は、バージョンフォルダとその下位フォルダを作成する。この処理は MLS_Original.xlsx を複製する。このスプレッドシートは本来変更されるべきものではなく、特定の日時に MLS からダウンロードした内容をそのまま反映するものであるが、期間の長いプロジェクトでは新たな MLS ダウンロードが必要になる可能性があるからである。次にこの段階は、設定パラメータを備えた拡張版 MLS.xlsx ワークブックの複製をバージョンフォルダに作成する。MARS の完全な実行ごとに作成されるバージョンフォルダは、いくつも存在しうる。

続く各段階は、ワークブックに対する列の追加、並べ替えその他の変更を生成し、鑑定評価が完了した時点で更新版を `MLS_Stage(N)` としてプロジェクトのデータフォルダに書き出す。繰り返すと、大本の `MLS.xlsx` は絶えず変化しており、バージョン複製は、変数の削除・追加その他の変更を伴う、元の `MLS` ワークブックの継続的な変異である。処理を行う R または Python のコードは、多くの場合 Data 配下の `MLS.xlsx` ワークブックからの入力で行われる。こうして、RCA 実行の各バージョンの背後にあるすべてのデータの複製を常に保持することになる。

バージョンは、これ以上の改善が不可能となるまで必要なだけ作成し、その時点で最良のバージョンを最終版として選択する。フォルダ名の末尾に「_Final」を付けて名称を変更すれば、ディレクトリを見たときに一目で分かる。興味深い、あるいは競合するモデルをもつ古いバージョンのいくつかは、通常は保存しておくべきである。各バージョンフォルダには、報告書に埋め込むことのできるテキスト、図、グラフの各ファイルが含まれる。最終的な `MLS_Stage5.xlsx` ワークブックには、報告書に必要なだけのシートが含まれる。とりわけ、ファニーメイの取引事例比較表に転記するためのスプレッドシートを備えることがある。また、集約の背後にある計算ないし内訳も必然的に含む。ここが、RCA のアプローチで最も時間を要する部分である — 難しい市場地域で完璧なモデルを見出そうとする格闘である。

取引事例比較表を再現するために必要なものはすべてバージョンフォルダに複製する必要がある。コードが頻繁に変更される場合には、使用した R または Python のコードもこれに含まれる。

5.4.1 プロトコル: 複数回の MARS(earth) 実行

プロトコル 2.1: 市場地域の価格要因を精密に調べるための過適合実行

earth の初期実行では、基礎にある価格要因を発見するために過適合を戦略的に用いるべきである。たとえば 800 件の売買で代表される市場地域に、作業場をもつ不動産が 3 件しか含まれていないことがあるが、そうした孤立した事例では作業場が最終的な売却価格に大きな影響を与えている可能性がある。前進パスのみの実行は過適合したモデルを生み、これらの作業場のおおよその価値寄与を示す。もっとも、800 件中 3 件しかない以上、作業場の項は後退パスで、また交差検証の下では確実に除去される。前進パスのみの実行が作業場に付与した値は書き留めておき、潜在変数分析に転記して、取引事例比較表に現れる作業場付きの不動産に用いることができる。このような過適合モデルは、後退パスの枝刈りや交差検証を伴う実行では除去されてしまう他の多くのデータの乏しい特性についても、詳細な推定値を与えてくれる。

こうした探索的な実行はそれぞれ独自のバージョンフォルダとして保存すること。記録された価値寄与は、後に残差を分解する際に VE が引用したくなる証拠である。もちろん、VE が最終的には枝刈りと交差検証に依拠して、回帰が一度も見えていない不動産 — たとえば後から現れる対象不動産 — に対してモデルを頑健にしなければならないことは理解しておく必要がある。過適合した実行はあくまで研究のための道具であって、決して最終モデルではない。

プロトコル 2.2: 異なる earth() パラメータによる複数回の交差検証実行

市場地域は構造も規模も異なるため、VE は通常、一連の実行を試みる必要がある。どの変数が価格に最も影響するかを理解するため、測定済みの変数の集合が価値の推定に十分な働きをしているか、それとも新たなデータ (水辺までの距離、標高、ショッピングセンターまでの距離など) を加える必要があるかを判断するため、そして非効率あるいは誤導的な交互作用を排しつつ新たな種類の交互作用を発見するためである。最終的に VE が求めるのは、訓練データにおける良好な R^2 と、検証データに対する可能な限り最良の CVR^2 である。そこに至るには、新たなデータの収集、既存データの精度の改善、あるいは市場地域の規模と性格に合わせた earth() パラメータの調整 (たとえば利用可能な自由度に応じた調整 — プロトコル 1.1 を参照) が必要となりうる。上に述べたプロジェクト構成が多数の実行バージョンフォルダを支えているのは、まさにこのためである。候補となる各モデルはそのパラメータと R^2/CVR^2 とともに保存されるので、実行を通じた軌跡を比較できる。 R^2 が上昇する一方で CVR^2 が停滞ないし低下するバージョンは、市場の信号ではなく雑音に当てはめているのであり、それを可視化するのがバージョンの履歴である。

プロトコル 2.3: 備考欄および記述的特性列の Prolog 展開

この機能はなお幾分実験的な段階にあるが、earthUI は現在、組込みの Prolog エンジン (vProlog) を備え、確定節文法 (Definite Clause Grammars, DCG) を用いて MLS の自由記述列 — とりわけ Public Remarks (公開備考)、さらにはエージェント備考や、部屋の特徴・既存構築物といったコンマ区切りの特性一覧 — を解析する。この展開は、「owned solar」「3 car garage」「40x50 workshop」といった語句を、回帰に投入できる通常の実行列へと変える。こうして回収されたデータの乏しい特性は、潜在変数分析に直接供される (後述の、近似的な回帰変数としての状態・品質評価に関する節を参照。この機能自体は本号の earthUI 姉妹論文で解説している)。

テキストの展開は候補となる予測変数の集合を増大させ、そのために検証を目的とした複数回の earth/RCA 実行が必要となる。展開された列はモデルに入れたり外したりして試し、共線的である場合には導出された件数やブール値へ統合し、雑音ではなく市場の信号を担っていることを確認するために CVR^2 に照らして検証しなければならない。そうした実験の一つひとつを、プロトコル 2.1 および 2.2 とまったく同様に、独自の実行バージョンフォルダとして扱うこと。また、解析文法自体が開発途上にあるので、VE は抽出された列を原文に照らして点検したうえでモデルに用いるべきである。

5.5 中核となる Excel ワークブック

ほとんどの MLS サービスは不動産データを Excel の CSV ワークシートにダウンロードするので、データと各種の設定パラメータを保存する手段としては Excel がおそらく最も便利である。プロジェクトごとに MLS.xlsx という中核の Excel ワークブックが一つあり、その中に MLS からダウンロードした成約リスティングを格納する「MLSDData」というシートと、いくつかの設定・パラメータシートが置かれる。

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

元のデータは MLS_Original.xlsx といった名称のワークブックに保存する。これを別名で複製し、たとえば「MLS.xlsx」として、R または Python の RCA プログラムに読み込ませる。この二つのワークブックはいずれも Data フォルダの下に置く。さて Data/MLS.xlsx ワークブックは、変数の組合せ、MARS のパラメータ、その他の設定の変更を試しながら、おそらく何度も実行の反復を経ることになる。反復のたびに、VER_YYYYMMDDHHMMSS という形式で日時を名称とした新しい入出力フォルダが作られる。各フォルダには入力となる MLS.xlsx ワークブックの複製が置かれ、四つの処理段階 (第 2-5 段階) それぞれについて生成された Excel ワークブックのほか、出力されたテキスト文書、図、グラフが格納される。なお最終段階である第 5 段階には、依頼者の種別と SOW に応じて、報告書用の他の多くのシートが含まれうる。

5.6 バックアップ

新規のあるいは変更した Excel スプレッドシートをプロジェクトのバージョンフォルダに書き出す際、プロジェクトフォルダ内の SQLite または PostgreSQL データベースへバックアップを取ることもできる。データベースファイルは、その後の期間に検討や参照のために用いられる過程で意図せず破損する可能性が相対的に低い。

6 ワークブックの表記法

6.1 記述次元

ワークブック「M」には、おおよそ六つの記述次元がある。

1. 鑑定評価案件 “a”
2. 実行バージョン “r”
3. 作業段階 “w”
4. シート “s”
5. 不動産リスティング “p”
6. 変数集合 “v”

したがって、あるワークブック M の個々のセルは次のように記述できる。

$$r_* \underset{w_*}{W} \overset{a_*}{s_*} p_* v_* \quad (1)$$

(注: 「*」は何らかの有効な添字を意味する)

あるいは次のように書く。

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

ここで:

- n_p = 行数 (シートによる)
- n_v = 列数 (シートによる)
- n_r = 実行バージョン数
- n_m = 作業段階数
- n_g = 分析プロジェクト数
- n_h = ワークブックのシート数

「W」は、六つの中核データシートから成る Excel ワークブック全体を指すことを理解されたい。各シートはそれぞれ独自の表をもち、対応する R または Python のプログラムを実行すると別々の内部データフレームに読み込まれる。R にも Python にも、名前を受け取ってそれがデータフレームのどの行あるいは列に属するかを見つける内部関数が備わっている。たとえば添字 v_j において $v = (\text{“GLA”, “LotSize”, “BathRms”, “BedRms”, ...})$ であり $j=0$ であれば、 $v_j = \text{“GLA”}$ となる。R でも Python でも、内部的に「GLA」をそれが格納されている列 — たとえば対応するデータフレームの第 15 列 — へ解決できる。したがって本稿では名前の配列に添字を付し、実行中のプログラムに名前を渡すと、プログラムがその名前を内部データフレームの実際の列番号または行番号へと再び索引づけする。やや複雑ではある。しかし厳密を期すためには、これが必要であり、また最も便利でもある。これは分析者の役割と開発者の役割との「責務の分離」である。もっとも実際には、評価エンジニア (VE) が両方の役割を担うこともありうる。なお、R でいう「data frame」は Python では「dataframe」と呼ばれる。

また、MARS の実行時パラメータシートは W の次元ではないことにも注意されたい。W は MARS に入力されるデータのバージョンに関わるものであり、MARS/earth がそのデータをどう扱うかを定めるパラメータとは別だからである。

明確を期するため、ワークブックについては、データワークブックの各シートが読み込まれる先である R または Python の内部データフレームにより直接対応する、複雑な多次元表記法を提示する。これは「重い」表記法である。しかし証明などに入る際には、扱う対象について前提を置くことで多くの荷物を捨て去ることになる。それでもこの表記法は、当初から作業手順とデータの全体像をよりよく把握するのに役立つ。

処理の対象となるワークブックの集合の表記法から始めよう。

6.1.1 プロジェクト次元 (“a”)

業務範囲 (SOW) が承認されたら、設定パラメータ、出力されるグラフ、図、スプレッドシート、報告書の断片を含む入出力データを保存するためのプロジェクトフォルダを設定する必要がある。VE はすべてのプロジェクトを整数のような単純な連番 ID で識別すべきである。そうすれば報告書が紛失した場合にもそれが明らかになる。プロジェクトが中止された場合には、当該プロジェクトフォルダ内で中止された旨を示す一貫した方法を設けるべきであり、決して削除してはならない。ここではプロジェクト ID が 0 から始まり 1 ずつ増加すると仮定するが、他の方式もありうる。さらに、有効なプロジェクト ID はすべて、処理プログラムによって作成される順序集合「a」に格納されるものと仮定する。

鑑定評価プロジェクトのデータはワークブックのプロジェクトワークシートに置かれるが、何らかの中央の SQL データベースに保存することもできる。

推奨されるプロジェクトの項目は次のとおりである。

1. ローカル ID(Local ID): 会社内で連番として生成される短い整数の ID。保存用の主キーとして好適であり、プロジェクトが増えるたびにデータベースが自動生成することもできる。
2. グローバル GUID(Global GUID): 必要であれば、次のサイトを用いて GUID(全世界的に一意な識別子) を作成できる。

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

3. 不動産の所在地 (市場分析の場合) または住所
4. 鑑定評価の価格時点
5. 依頼日
6. 現地調査日
7. 当初報告書の日付
8. 第 1 回更新報告書の日付
9. 第 2 回更新報告書の日付
10. 第 3 回更新報告書の日付
11. 依頼者名
12. 依頼者住所
13. 貸手名
14. 貸手住所
15. 所有者名
16. 所有者住所
17.

プロジェクト ID の一覧の例:

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

6.1.2 実行バージョン次元 (“r”)

作業手順の第二段階として、入力に対して MARS 回帰を実行する。モデルを改善するために各種のパラメータを変えながら、数回から多数回の実行を行うことになるだろう。実行のたびに、実行日時に基づく名称の新しいバージョンフォルダが出力用に作られる。たとえば次のようになる。

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

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

競合しうるモデルと出力をもつ有望なバージョンは、最終版が決まるまで、場合によっては何らかの有用性があるならそれ以降も保存しておくべきである。名称の末尾に「Final」を置いて、最終版であることを示すようにバージョンフォルダ名を変更してもよい。

6.1.3 作業段階 (“w”)

RCA の作業手順は通常、以下に列挙する五つの段階に分けられる。各段階は次の段階を始める前に正常に完了していなければならない。VE は、先へ進む前に結果を検討できるよう、ある段階の終わりで処理を停止できるべきである。市場地域はかなり複雑で、しばしば非常に個性的であるため、これが必要となる。新たな手法を要する異例の問題に遭遇することもある。VE は必要な変更を加えるため、さらには新たな手法を試すために、R や Python のコードに踏み込めなければならない。またそうした場合、取り組んでいる段階を何度も実行して、さまざまな手法を試したくなることもある。もっとも、VE は一つの命令ですべての段階を実行できるようであるべきでもある — ただしいくつかの但し書きが付く (たとえば 3 次の集約は、第 2 段階の先行実行を通じてすでに決定されていなければならない)。これは、各段階が完了した時点で、その内部データをすべてファイルないしデータベースに保存し、次の段階が前段階のデータをメモリに再読込することで、あたかも前段階がまだメモリ上にあるかのように開始できるようにすべきだ、ということを意味する。

各段階は次のとおりである。

```
w = ("stage1", — 設定
     "stage2", — MARS 処理
     "stage3", — 最終取引事例比較表
     "stage4", — 報告書用の断片
     "stage5") — 検討・整理
```

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

6.1.4 Excel ワークブックのシート (“s”)

RCA の中核 Excel ワークブックには、それぞれが内部のデータフレームに対応づけられる次のシートがある。

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

2 次の集約を別のシートに置く理由は、それが対称表によって最も容易に定義できるからである。この表は交互作用のあらゆる組合せを示し、各組に対応する価値寄与と補正額を集約(加算)する集約変数を対ごとに指定できるようにする。3 次の集約についてはこれができない。3 次元の対称表が必要となるが、Excel ではそれが不可能だからである。さらに、20x20x20 の表は 8,000 セルを扱うことになる。3 次の項の名称は、1 次の集約を載せた表に単に列挙するほうが容易である。たとえば次のようにする。

変数	集約先
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 不動産リスティング (“p”) — MLSDData の行

不動産リスティングは MLSDData シートの行を構成する。第 4 段階に至るまで、他のシートには現れない。第 4 段階では、最終的に選定した取引事例をもって報告書用の断片が作成される。不動産 ID の一覧の例を示す。

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

6.1.6 MLS 変数 (“v”) — MLSDData の列

変数の扱いについては後により詳しく述べる。ここでは、MLSDData シートには通常さまざまな型の変数が 30–40 個あるが、MARS 回帰の説明変数として選ばれるのは通常 10–20 個にとどまり、と述べるだけで足りる。これらの変数は他のシートにも現れるが、多くの場合は MARS に投入しうるすべての変数の一覧としてである。変数一覧の例は次のとおりである。

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

6.1.7 最終的な表記の例

以上の表記を最後の例にまとめてみよう。上記の各一覧を用いて、次のようにおく。

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

ここでは 0 始まりの添字を用いているので、各一覧の先頭要素の添字は 0 であると想定されることに注意されたい。Python も 0 始まりの添字を用いるが、R は 1 始まりである。

したがって上式は、次のセルの値を指していることになる。

- 変数 v_0 すなわち列「GLA」
- 不動産 p_2 すなわち ID「MLS823542」
- シート s_0 すなわち「MLSDData」
- 作業段階 w_1 すなわち「stage2」
- バージョン r_1 すなわち
「ver20241209_151130」
- プロジェクト a_{141} すなわち「142」

上記の表記は、次のようにも書ける。

$$\begin{matrix} \text{"ver20241209_151130"} \\ \text{"stage2"} \end{matrix} W_{\text{"MLSDData"}}^{\text{"142"}} \begin{matrix} \text{"MLS823542"} \\ \text{"GLA"} \end{matrix} \quad (4)$$

6.1.8 正当化

「なぜ表記に単純な添字や数字を使わないのか」。この表記法には二つの要件がある。

1. 部分集合。たとえば分析者は一覧の部分集合、とりわけ変数の部分集合をしばしば選択する。MLSDData シートには 35–40 個の変数があるかもしれないが、ある実行で用いられるのはそのうち 10–20 個にすぎない。実行ごとに、VE は MARS/earth へ投入する変数を頻繁に変更する。
2. 反復。たとえば $i = 1, 2, \dots, n_1$
アルゴリズム、プロトコル、証明は通常、順序づけられた一覧のすべての項目を先頭から末尾まで反復できること、あるいは特定の項目を取り出せることを必要とする。

したがって、項目の部分集合ないし一覧を表す文字が必要であり、その文字は、一覧の項目にわたって反復し、データフレームの行や列に索引づけするための文字列を与えてくれる別の添字を伴っていなければならない。

6.1.9 行列は使わないのか

では線形代数の行列はどうか。行列は整数または実数のみを扱う。データ構造としての行列は R でも Python でも扱える。しかし MLS データで扱う変数の多くは「文字列」すなわち文字データである。数値のみの変数であっても、実数ではなく単なる名称であることが多い。

たとえばある都市の MLS エリアコードは (“660”, “661”, “662”, “663”, ..., “672”) かもしれない。earth の回帰では変数を「因子 (factor)」変数として指定でき、その場合、数値は順序をもたない単なる名称として扱われる。そうした場合、生成されるモデルは数列上を走る関数ではなく、各数値についての価値寄与を与えるにとどまる。他の例としては、状態についての “C1”、..., “C6”、品質についての “Q1”、..., “Q6” がある。

6.1.10 第一の単純化

中核となる Excel プロジェクトワークブックについての上記の表記は、全体を結びつけるうえでは有用だが、アルゴリズムや数学的証明を説明するために持ち歩くには重すぎる。込み入った作業を行う際には、所与のプロジェクト、バージョン、段階、シートの内部にいと仮定してよい。まだ暗黙になっていない場合にのみ、何を扱っているかを述べればよい。何を扱っているかを一度示してしまえば、上記の表記から実際に必要なのは次のものだけである。

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

ここで: n_p = 不動産の数
 n_v = 入力変数の数
 n_s = 入力シートの数

6.1.11 第二の単純化

添字を用いるよりも、各シートを、それが担う具体的な種類のデータフレームの名で呼び直すほうが便利である。シートごとに行と列の機能は異なる。そこで、W というデータフレームの集合を、次のように M、V、A、I、C の各データフレームへ分割する。

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

ここで: n_p = 不動産の数
 n_v = 入力変数の数

注: 対象不動産は既定により常に不動産一覧の先頭に置かれる

「MLSData」という Excel スプレッドシートおよび対応するデータフレームでは、第 1 行は常に対象不動産のために予約されている。この行が、その下にある他の MLS 掲載不動産とともに並べ替えられることはない。またこの行が MARS 回帰に説明データとして投入されることもなく、投入されるとすれば目的データ、すなわち被説明変数としてである。同じことが不動産の一覧にも当てはまる。特に断らないかぎり、すべての不動産一覧において添字 0 の先頭要素は常に対象不動産の ID であると想定する。

対象不動産のない市場地域についてモデルを作成することが課題である場合でも、第 1 行にはやはり何らかのダミーの不動産データが読み込まれる。こうすることで、プログラムを書くにも、アルゴリズムを定めるにも、証明を行うにも、はるかに便利になる。したがって、不動産の一覧を c で表すとすれば、 c_0 は対象不動産であり、 c_i 、 $i = 1, \dots, n_p$ は MARS/earth 回帰に投入される取引事例のリスティングである。

◇ 回帰変数の仕様

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

ここで: n_v = 利用可能な変数の数
 n_s = 仕様数の数

なお $n_s \leq n_v$ である。回帰への入力には通常、利用可能な変数の部分集合を選ぶからである。

◇ 計算

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

ここで: n_v = 変数の数
 n_c = 計算仕様数の数

◇ 交互作用

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

ここで: n_f = 起点となる変数の数
 n_t = 終点となる変数の数

データフレーム I は対称的なデータフレームであり、「起点 (from)」の行と「終点 (to)」の列に変数がアルファベット順に並ぶことに注意されたい。セルに「1」を記すと、対応する変数の組の間の交互作用が許容されることを示す。「0」であれば交互作用は許されない。2 次の交互作用については、可能な交互作用が一つしかないので比較的単純である。3 次の回帰では、3 変数すべての間に三つの対の交互作用が生じ、3 変数の交互作用項 (すなわち変数) を生成するには、その三つの対の交互作用がすべて許容されていなければならない。

◇ 集約

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

ここで: n_v = 変数の数

7 RCA 取引事例比較表

RCA のプロトコルが、R/earth にアクセスして詳細な手順のすべてを実行できる R または Python のプログラムを必要とすることを念頭に置かれたい。この作業はデータ集約的であり、手作業でこなすことはとうてい不可能である。同時に、プログラムに不具合がないという前提の下で、必要な作業が正確に、人為的誤りなく行われることを担保するためにも、ソフトウェアが必要である。

また、この作業を行う VE にとっては、必要に応じてそうしたプログラムを改変できることが不可欠である。予期しない新たな市場地域、近隣地域、不動産、あるいはその他の新しい要求や問題に対処する場合はとりわけそうである。

RCA の処理においては、三種類の取引事例比較表がある。

1. RCA 処理用取引事例比較表。本稿では「MLSDData」と呼ぶ。これは非常に大きなデータフレームおよび対応するスプレッドシートであり、多くの中間的な詳細を含みうる。それらは最終的に集約の過程を経て、下記の RCA 取引事例比較表に至る。例を示すには大きすぎてここには載せられない。評価エンジニアは多くの場合、このスプレッドシートを自らの好みに合わせて調整することになる。そうした表の作成については、いずれ別稿を書く予定である。
2. RCA 最終取引事例比較表。GSE 以外の報告書で用いられるもので、表 2 のような体裁をとる。これはたとえばファニーメイの様式を模しており、読み手にある程度の親しみやすさを与える程度には似せてある。しかし、取引事例 1 件あたりの列数がファニーメイの表より多いことに注意されたい。各取引事例には「測定値」「補正額」の各列に並んで「価値寄与」の列があり、対象不動産にも同様の列がある — いずれもファニーメイの様式にはない列である。より簡略な形式としては表 1 もある。
3. GSE アップロード用取引事例比較表。Alamode のようなソフトウェアに取引事例比較表を取り込むために用いる。もっとも、鑑定評価様式ソフトウェアのすべてがスプレッドシートから取引事例比較表のデータを取り込む機能を備えているわけではないことに注意されたい。必要上、こうした表には価値寄与のための独立した列を設けることができない。ファニーメイのような GSE には価値寄与という概念自体がないからである。方策としては、GSE には要求どおりの様式を渡しつつ、それを RCA 取引事例比較表と、変数の集約に関する詳細などの裏づけ資料によって補強することである。

RCA 処理用取引事例比較表はファニーメイの取引事例比較表とは似ておらず、おおよそその転置、すなわち行と列を入れ替えたものである。各行はある不動産のデータを含み、各列はある不動産変数のデータを含む。不動産の行が数百から数千、列が 200 以上になることも珍しくない。第 1 行は変数名のために予約され、第 2 行は、市場分析のみを行う場合であっても対象不動産のために予約される。スプレッドシートをデータフレームに読み込むと、シートの第 1 行は names 属性となり、対象不動産の第 2 行がデータフレームの第 1 行となる。

RCA 取引事例比較表の列には、識別のための列があり、続いて一次変数、外部変数、合成変数が組み合わされて説明変数として MARS に投入される。MARS はこれらの変数のうち、売却価格 — 被説明変数ないし目的変数とも呼ばれる — に有意な影響を与えるとみなしたものを出力する。2 次の回帰では、これらの変数の対が新たな変数として得られることがある。3 次の回帰では、いわゆる説明変数として働く変数の三つ組が得られることもあるが、もちろんそれらは基礎となる変数と強く相関している。

表 1: RCA スプレッドシートの例: 縦方向レイアウト

対象不動産				事例 1			事例 2		
	測定値	価値寄与		測定値	価値寄与	補正額	測定値	価値寄与	補正額
売却価格				\$1,150,000			\$1,500,000		
基礎価値		\$500,000			\$500,000	\$0.00		\$500,000	\$0
		測定された特性			測定された特性			測定された特性	
居住面積	1,530	\$463,500		1,350	\$382,500	\$81,000	1,700	\$540,000	-\$76,500
敷地面積	6,000	\$200,000		6,500	\$220,000	-\$20,000	7,000	\$240,000	-\$40,000
築年数	53	\$63,600		57	\$68,400	-\$4,800	51	\$61,200	\$2,400
浴室数	2	\$15,000		3	\$30,000	-\$15,000	3	\$30,000	-\$15,000
寝室数	3	\$6,000		4	\$9,000	-\$3,000	4	\$9,000	-\$3,000
測定分合計		\$1,248,100.00			\$1,209,900.00	\$38,200		\$1,380,200.00	-\$132,100
		残差に係る特性			残差に係る特性			残差に係る特性	
残差	CQA 5.8	\$30,000		CQA 2.5	-\$59,900	\$89,900	CQA 8.2	\$119,800	-\$89,800
	残差の分解			残差の分解			残差の分解		
デザイン		7,000			-\$15,000	\$22,000		\$35,000	-\$28,000
状態		5,000			-\$8,000	\$13,000		\$20,000	-\$15,000
品質		8,000			-\$10,000	\$18,000		\$30,000	-\$22,000
機能的効用		5,000			-\$12,000	\$17,000		\$15,000	-\$10,000
付帯設備		5,000			-\$14,900	\$19,900		\$19,800	-\$14,800
小計		30,000			-\$59,900	\$89,900		119,800	-\$89,800
合計				合計			合計		
価値寄与合計		\$1,278,100.00			\$1,150,000.00			\$1,500,000.00	
補正後売却価格						\$1,278,100.00			\$1,278,100.00

8 変数

MARS を実行する入力側にあるのは、次の単純な変数だけである。

1. MLS スプレッドシートから読み込んだ不動産の特性。
2. MLS 変数を精緻化し、あるいは既存変数の変換を加えるために VE が作成した導出変数。たとえば VE は、売却日を「DaysOffMarket」(鑑定評価の価格時点から売却日を差し引いた値) へ変換することを好むかもしれない。
3. VE は、住宅ローン金利のような第三者の情報源からデータを取り込むことがある。あるいは、緯度、経度、標高のデータが MLS データに欠けている場合、他の情報源から取り込むことがある。これらの変数は「外部変数」と呼ぶ。

MARS の出力側では、次のいずれかの項の組合せから成る価格モデルを扱う。

1. 入力変数を一つだけ含み、その変数の価値寄与を与える式となっている項。これらは 1 次の項ないし (出力) 変数と呼ばれる。項そのものが価値寄与の値として扱われ、名前を与えられる。
2. 入力変数を二つ含み、その二変数の交互作用の価値寄与を与える式となっている項。これらは 2 次の項ないし (出力) 変数と呼ぶ。
3. 入力変数を三つ含み、その三変数間の交互作用の価値寄与を与える式となっている項。本稿ではこれらを 3 次の項ないし (出力) 変数と呼ぶことがある。

MARS の出力変数は、残差の価値寄与を含めて、いずれも価値寄与である。それらはさらに、補正額の計算やグラフの作成など、取引事例比較表の計算における変数として扱われる。

変数には、二つの基本的なオントロジーに分類されるさまざまな型がありうる。すなわち (1) 変数の由来、および (2) 変数の用途、つまり非残差か残差か、である。

8.1 変数の由来オントロジー

1. 1 次変数

- 1.1 MLS 回帰変数
- 1.2 外部回帰変数
- 1.3 導出回帰変数

2. 因子変数:

因子変数を扱うには、それを識別できる関数が必要である。因子変数は Excel のプロジェクト設定ワークブックで指定されるので、渡された変数が因子型に分類されている場合に真を返す関数を作成する。

3. 交互作用変数: 次数が 2 以上に設定されている場合、モデルの生成過程で MARS が作り出す変数である。ただし、解釈の難しさ、過適合の可能性、自由度への影響、そして頑健なモデルを保証するために必要となる不動産レコード数の増加を考えると、交互作用する変数は 3 個までが推奨の上限である。

- 3.1 2 次変数: よく見られるもので、GLA:BathRms、GLA:LotSize、DateOfSale:BathRms、Latitude_Longitude が好例である。

3.2 3 次変数: あまり見られない交互作用であり、慎重な扱いと過適合の可能性への注意を要する。

4. 残差分解変数

8.2 変数の用途オントロジー

1. MARS が生成する価値寄与

1.1 1 次変数

- 1.1.1. MLS 回帰変数
- 1.1.2. 外部回帰変数
- 1.1.3. 導出回帰変数

1.2 交互作用変数

- 1.2.1. 2 次変数
- 1.2.2. 3 次変数

1.3 残差分解変数

1.4 集約変数

2. エンジニアが生成する変数および値

2.1 残差分解変数

- 2.1.1. デザイン・美観
- 2.1.2. 造園
- 2.1.3. 機能的効用
- 2.1.4. 眺望
- 2.1.5. 状態
- 2.1.6. 品質
- 2.1.7. ...

これらの変数群には次節で表記法を与え、より詳しく説明する。

プロトコル 3.1: 因子変数

因子変数、あるいは単に因子は、交互作用を含むあらゆる項に現れうる。MARS 回帰とグラフ化にあたって特別な扱いを要するため、因子変数は重要である。

文字値をとる変数は必然的に因子であるが、数値変数については、MARS に因子として認識させるために因子変数として指定しなければならない。たとえば MLS のエリアはしばしば番号で識別される。とりわけサンフランシスコ・ベイエリアの近隣地域は、しばしば一意な整数で識別される。これらの値を因子として MARS に渡すと、MARS は、他の地域との関係を顧慮することなく、特定の地域がその境界内の不動産売買に対して平均してどれだけの価値を寄与しているかを見出そうとする。これに対し通常の変数では、変数値の変化に伴って価値寄与に何らかの連続性があると想定する。これは重要な区別である。さらに、これらの因子変数は非因子変数とともにモデルの項に現れることがあり、それがグラフ化を難しくする。具体的には、因子を含む単一変数の項をグラフ化する際には、そうした因子をすべてまとめ、その寄与が正か負かを示す棒グラフを描くにとどめる。

項の中で因子が他の変数とともに現れる場合、因子は常に一つの取りうる値とともに現れる。ある不動産について因子変数はただ一つの値しかとりえないからである。すなわち、ある地域の値が 662 であれば、定義上その地域が他の値をとることはありえない。これが分かりにくいとすれば、事態をさらに厄介にすることに、MARS が生成するモデルの中には因子に対応する単一の変数名が存在しない。MARS は入力として与えられた因子変数の語幹をとり、取りうる値を末尾に付した変数名で新しい列を作る。したがって、Excel スプレッドシートの単一の列、たとえば MlsArea に対して、取りうる値ごとに MlsArea660、MlsArea661 といった複数のデータフレーム列が作られる。MlsArea661 のような変数は、エリア 661 にある不動産については 1、それ以外については 0 という値をとる。そのうえで MARS は、それらの列をブール変数であるかのように回帰にかける。MARS が提供するモデルでは、この変形された名称が用いられる。

おそらく、これらのモデル上の名称を元の名称に戻したくなるだろう。簡潔で分かりやすいものを好む依頼者 (少なくとも可能なかぎり簡潔なもの) にとって、グラフがより理解しやすくなるからである。MlsArea661=1 と表示する代わりに、MlsArea=661 と表示することになる。そうすればグラフの見出しに「MLS エリア = 662 について」などと記すだけで項をグラフ化できる。こうして、M 個の因子変数を含む N 次の項は (N-M) 次の項としてグラフ化され、因子の次元は表題の中に消える。

一つの項に二つ以上の因子変数が存在しうるので、取りうる値の組合せごとにグラフが必要になる可能性が高いことに注意されたい。異なる値を三つとる変数 x と、異なる値を二つとる変数 y が、非因子変数 z とともに同じ項に現れる場合、 $2 \cdot 3 = 6$ 枚のグラフが必要となる。例としては「MlsArea」と「建築様式」の組合せがある。たとえばサンフランシスコのヴィクトリア様式住宅は、市内のどの地域にあるかによって価値が高くも低くもなりうる。幸い、ほとんどの地域では、価値の寄与要因としてモデルに用いるほど重要な建築様式は通常 1-3 種類にとどまる (建築様式自体は 6 種類以上あるかもしれないが)。

8.3 2 次および 3 次の項

2 次または 3 次の回帰では、出力モデルの項の一部に、それぞれ二つあるいは三つの変数が含まれることがある。たとえば次のようになる。

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

これらの項は、変数名をアルファベット順に並べ替え、必要に応じてアンダースコアで連結

して命名することに注意されたい。モデルが生成した変数名の三つ組についても、3 次の項として同様に扱う。たとえば「BathRms__GLA__LotSz」である。

これらの変数は第 3 段階で一次入力変数に加えられる。この段階で RCA のプログラムは、交互作用変数を含むすべての変数について、不動産ごとの寄与を計算する。

8.3.1 補正変数

RCA のプログラムは第 3 段階で、価値寄与の計算後、ある特性についての対象不動産の寄与から対応する取引事例の価値寄与を差し引くことにより、すべての変数と不動産についての補正額も計算する。

8.3.2 集約変数

集約変数は、変数の部分集合を置き換えるものとして定義され、それらの価値寄与と補正額をすべて加算して、報告書で用いる共通の変数にまとめる。たとえば、仕上げ済みおよび未仕上の居住面積を、より細かな変数に分割していることがある。

1. 地上階の合法的な仕上げ済み面積
2. 地下階の合法的な仕上げ済み面積
3. 地上階の非合法的な仕上げ済み面積
4. 地下階の非合法的な仕上げ済み面積
5. 地上階の未仕上げ面積
6. 地下階の未仕上げ面積

そう、これが実際に意味をもつ地域が存在するのである。

MARS 回帰はこれらすべての変数を回帰にかけ、それらについての寄与と補正額を与えるモデルを生成する。しかし様式に従う報告書のためには、これらの異なる寄与をおそらくより少数の変数へまとめる、すなわち集約する必要がある。集約の詳細は附属資料に記載すべきである。より頻繁な例としては、交互作用の補正額を一次変数のいずれかの補正額へ集約することが挙げられる。たとえば、GLA_Lotsize の補正額を GLA に、あるいは LotSize に集約し、さらには集約シートの定めに従って両者に 50:50 で分割することもありうる。

8.3.3 特殊な変数と表記法

π_i	不動産 $i = 1, 2, \dots, n_p$ の実際の純売却価格。対象不動産については π_0 が価値結論を表す (証明 I を参照)
α_i	不動産 $i = 1, 2, \dots, n_p$ の補正後純売却価格
ρ_i	不動産 $i = 0, 1, \dots, n_p$ の MARS による推定純売却価格
β	モデルの基礎価値
$\varphi_{i,j}$	不動産 $i = 0, 1, \dots, n_p$ における回帰変数 $j = 1, 2, \dots, n_r$ の価値寄与
$\delta_{i,j}$	不動産 $i = 1, 2, \dots, n_p$ における回帰変数 $j = 1, 2, \dots, n_r$ の補正額
ϵ_i	不動産 $i = 0, 1, \dots, n_p$ の残差合計
$\epsilon_{i,c}$	不動産 $i = 0, 1, \dots, n_p$ における残差構成要素 c 、 $c = 1, 2, \dots, n_c$ の価値寄与
$\xi_{i,c}$	不動産 $i = 0, 1, \dots, n_p$ における残差構成要素 c 、 $c = 1, 2, \dots, n_c$ の補正額

ここで

n_p = 取引事例の数

n_r = 回帰変数の数

n_c = 残差分解変数の数

8.3.4 残差の分解

MARS 回帰がモデルを作成すると、MARS が選択した入力変数から推定売却価格が生成される。この値を対応する純売却価格から差し引いて、不動産ごとの残差を得る。さて RCA のプロトコルは、VE が最終的な取引事例 — 通常は 6–12 件のリスティング — を決めたなら、各リスティングを順に検討し、無作為なデータ誤りの可能性は措いたうえで、算出された残差を、価値に影響を与えると VE が考える最も可能性の高い特性 (潜在変数) へ分解しなければならない、と定める。たとえばある不動産について、次のような分解変数を立てることが考えられる。

1. デザイン
2. 機能的効用
3. 市場水準を超える／下回る売買

4. 状態
5. 品質

不動産ごとに残差分解変数は幾分異なるのが通常であり、それらを一つの順序集合にまとめる必要がある。

稀少な特性の寄与の推定。 価値をもつ特性のなかには、回帰の経路には稀すぎるものがある。780 件の売買のうち 3 件にしか存在しない作業場は防御可能な係数を得られず、前進パスは通常そうした変数をまったく素通りする。これらは残差に属する特性であり、その寄与は残差空間において推定できる。すなわち、当該特性を除いてモデルを当てはめ、次にモデルの残差を、除外した列に対して通常最小二乗法で回帰するのである。二値のフラグであれば係数は一括の寄与額 (同義であるが、フラグの立った売買における残差の平均的なずれ) であり、数値列であれば単位当たりの寄与額である。相関する複数の付帯設備を同時に推定する場合 (たとえば附属建物の床面積と作業場フラグを併せて推定する場合)、回帰は同じ金額を二つの分解行で二重に数えるのではなく、両者に配分する。おおむね 10 件未満の観測に支えられた推定値は示唆的なものにとどまり、減価償却後の原価やより広い標本 (たとえば公開備考から回収したフラグ) によって裏づけるべきである。最終的に割り当てられる値は、残差制約それ自体によって上限が画される。earthUI(v0.11 以降) はこの方法を *Residual Contribution*(残差寄与) タブとして直接実装しており、同時推定、売買ごとの残差の証拠、作業ファイル用の記録メモを備えている。

表記: 上表の表記を用いると、次の式が得られる。

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

ここで n_c は残差分解の構成要素の数である。

VE は自らの判断を用いて残差を所与の構成要素へ分解しなければならない。これでは最終的な価値結論にバイアスが入り込むと思われるかもしれない。しかしそうはならない。分解による補正額の合計が残差に一致するなら、それらが価値結論に影響しないことについて数学的な証明があるからである。分解の目的は価値を確定することではなく、他の不動産売買の残差と比較して当該残差がなぜその値なのかを説明することにある。エンジニアのバイアスが最終的な価値結論に影響しうる唯一の箇所は、対象不動産の残差の推定である。当然ながら、対象不動産には残差を計算するのに必要な実際の売却価格が存在しないからである。

不動産にはそれぞれ残差の価値寄与の集合があるが、それらをすべての不動産にわたる一つの集合 C へ統合する必要がある。

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

ここで $[\epsilon_{i,c}]$ は値 $\epsilon_{i,c}$ の変数名であり、 A_i はそのような変数の集合、 n_c は各 ϵ_i に含まれる値の数である。

したがって、対象不動産と取引事例のデータを別々の列に配置すると、各不動産についての回帰モデルの全変数の価値寄与の下に、対応する n_c 個の残差の価値寄与から成る集合 C が現れるはずである。

9 RCA の中核となる証明

次のようにおく。

$$\begin{aligned} n_p &= \text{不動産の数} \\ n_r &= \text{回帰変数の数} \\ n_c &= \text{残差構成要素変数の数} \end{aligned}$$

証明 I: いずれの取引事例の補正後売却価格も、推定された対象不動産の売却価格に等しい。

任意の取引事例 i について、補正後売却価格 α_i は、純売却価格 π_i に補正額を加えたものである。補正額とは、基礎価値、回帰変数 $j = 1, 2, \dots, n_r$ 、および残差について、対象不動産と取引事例の価値寄与の差である。実際の売買のない対象不動産については、純売却価格 π_0 を価値結論 $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$ として定義する。このとき次が成り立つ。

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

証明 II: 残差構成要素の価値寄与の合計が当該取引事例の残差に一致するという制約の下で残差の分解を変更しても、RCA プロトコルの最終的な価値結論には影響しない。

これはほとんど自明である。それでも、数式に思いを致さなければ見落としやすい。従来の鑑定評価者は価値寄与ではなく補正額に注目するからである。ここで打ち消し合うのは価値寄与のほうである — 数十年にわたる見落としであった。ある取引事例 k 、 $k > 0$ について、残差の補正額の合計は n_c 個の分解された残差構成要素の和に等しく、それはさらに価値寄与の差の和に等しく、それは対象不動産の残差合計から当該取引事例の残差合計を差し引いたものに等しい。しかもこれは、作られた残差構成要素の個数や、それらが何を表すかとは無関係である。残差構成要素の値の合計が対象不動産についても選ばれた取引事例についても残差合計に等しいかぎり、 ξ_k は変化しない。

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

10 MARS モデル

10.1 モデル変数と交互作用の「 g 」関数

RCA 手法の中核をなす原則は、「説明的推論」がこのうえなく重要だということである。しかし残念ながら、さまざまな価値寄与の式について意味のあるグラフを提供することは、とりわけ交互作用や因子変数が用いられている場合には容易でない。したがって、これらのグラフを作成するための明確で厳格なプロトコルを示すことが不可欠と考えられる。以下でそれを行うが、このプロトコルはグラフを生成する R または Python のコードを書く開発者に向けたものであり、コーディングに関与しない読者にとっては表面的な関心にとどまるかもしれないことを了解されたい。

MARS が生成する価格モデルは、基礎定数に、それぞれ 1 個から 3 個の変数を含む項を加えた和である。これを分析するには、まずモデルを個々の項に分解する。次に、同一の変数の組を共有し、単一の交互作用を表す複数の項を、一つの交互作用の式にまとめる。回帰の最大次数を 3 と仮定する。基礎定数、 n_1 個の 1 次の式、 n_2 個の 2 次の交互作用の式、 n_3 個の 3 次の交互作用の式から成る 3 次の回帰モデルを考える。こうして、単一変数あるいは複数変数の交互作用から成る四つの異なる式の群が得られ、それぞれ次の二つの要素に応じて異なる方法でグラフ化される。(1) 因子変数 (カテゴリーカル変数) を含むかどうか、および (2) 含まれる相異なる変数の個数である。具体的には次のとおりである。

- 1 次の式 (変数 1 個) は 2 次元でグラフ化する。
- 2 次の式 (変数 2 個) は 3 次元でグラフ化する。
- 3 次の式 (変数 3 個) は 3 次元グラフの集合としてグラフ化する。
- 式が因子変数を含む場合、相異なる因子ごとにグラフの次元が一つ下がる (たとえば 1 変数が因子であれば 3 次元グラフは 2 次元になる)。

各関数 g の添字づけは、相異なる変数の個数と因子変数の有無を反映して、その関数をどうグラフ化すべきかを定めるのに必要な情報を与える。

10.2 g 関数の添字づけ

各関数 g は $f_{j,k} g_k^j$ と添字づけられる。ここで、

- $f_{j,k}$ は第一の添字 (g の前に置く上付き) である。
 - j は 0 から 3 までの値をとり、群を識別する。
 - * $j = 0$: 基礎定数 (通常の添字づけの例外)。
 - * $j = 1$: 1 次の式。
 - * $j = 2$: 2 次の式。
 - * $j = 3$: 3 次の式。
 - k は群内における特定の関数の添字である (たとえば群 j の最初の関数であれば $k = 1$)。
 - $f_{j,k}$ の値は、その関数に含まれる相異なる因子変数の個数を示す。
 - * $f_{j,k} = 0$: 因子変数なし。
 - * $f_{j,k} > 0$: 因子変数の個数 (たとえば 1、2、3)。
- g の後続の添字。

- g の後の上付き j (たとえば g^j) は、明確を期して群番号を繰り返す。
- g の後の下付き k (たとえば g_k) は $f_{j,k}$ の k と一致し、群内における関数を識別する。

f と g の添字は数値としては一致するが、その役割は異なる。 $f_{j,k}$ は因子変数の個数を数えるもので、グラフ化の判断に決定的であるのに対し、 g に付された j と k はその群と位置を定める。

10.3 グラフの次元

関数 $f_{j,k} g_k^j$ についてのグラフの次元 d は次のように計算される。

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

ここで、

- j は相異なる変数の個数 (定数の場合を除けば群番号) である。
- $f_{j,k}$ は因子変数の個数である。
- d はグラフ化すべき連続 (非因子) 変数の個数であり、最大次数が 3 であることから上限は 3 である。連続変数を d 個もつ関数は $d+1$ 次元で描かれる。すなわち $d=1$ なら 2 次元グラフ、 $d=2$ なら 3 次元グラフ (または 2 次元ヒートマップ)、 $d=3$ なら 3 次元グラフの集合となる。

たとえば次のようになる。

- $j=3$ (変数 3 個) かつ $f_{j,k}=0$ (因子変数なし) であれば $d=3-0=3$ となり、変数の一つを固定した 3 次元グラフの集合が必要となる。
- $j=3$ かつ $f_{j,k}=1$ (因子変数 1 個) であれば $d=3-1=2$ となり、モデルに現れる因子変数の値ごとに 3 次元グラフ (または 2 次元ヒートマップ) が必要となる。
- $j=0$ (定数) であれば単一の値にすぎないので、通常グラフ化は不要である。

$d=3$ の場合 (たとえば因子変数を含まない 3 次の式)、関数の全体を可視化するには複数の 3 次元グラフが必要となりうる。各グラフは一つの変数を異なる値に固定して、関数の挙動の一部を示す。

10.4 最終的なモデル式

MARS モデルは四つの式の群 (定数、1 次、2 次、3 次) に分解される。添字づけ $f_{j,k} g_k^j$ は次を教えてくれる。

- j : 群と変数の個数。
- $f_{j,k}$: 因子変数の個数。
- $d = j - f_{j,k}$: グラフの次元。

この構造は各 g 関数をどう可視化するかを導き、グラフ化の方法が式の複雑さと変数の型に対応することを保証する。

MARS モデルが与える売却価格の推定値 $\hat{P}$ は、次のように表せる。

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

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

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

$n_1 = 1$ 次の項の数
 $n_2 = 2$ 次の項の数
 $n_3 = 3$ 次の項の数

10.5 モデルをグラフ化するのに必要な図の数

モデル中のすべての関数を可視化するには何枚のグラフが必要か。その答えは、変数、それらの交互作用、そしてそれらが因子変数か非因子変数かによって決まる。

VE は R/earth 用の Excel 設定ファイルにおいて変数と交互作用を制限でき、それがどの交互作用を許容するかに影響する。実務上有用な方法は、R/earth にあらゆる交互作用を検討させ、売却価格の推定に有意な寄与をするものを選ばせたうえで、過適合や共線性といった問題がないかを点検することである。一部の交互作用は許容されず、それが高次の交互作用にも影響する (たとえば 3 次の交互作用において一つの対が除外されれば、その組合せ全体が除外される)。

因子変数の場合 (たとえば 12 個の値をもつ MLS エリアが GLA と交互作用する場合)、グラフ化はいつそう複雑になる。この場合、エリアごとに 1 枚、計 12 枚のグラフが必要となりうる。さらに建築様式のような別の因子 (6 個の値) が MLS エリアおよび GLA と交互作用すれば、最大 $12 \cdot 6 = 60$ 枚のグラフが必要になりかねない。もっとも R/earth がすべての組合せを有意と判定することはまれである。通常はモデルが特定するいくつかの主要な交互作用 (たとえば特定のエリアと様式) だけがグラフ化に値する。

必要なグラフ枚数の簡略な指針を示す。

1. 単一変数の関数 (非因子): 1 枚。
2. 単一変数の関数 (因子): すべての因子の値をまとめた棒グラフ 1 枚。
3. 基礎関数: グラフは不要。
4. 2 次の交互作用 (因子なし): 1 枚 (たとえば売却価格を色で表す 2 次元ヒートマップ)。
5. 2 次の交互作用 (因子 1 個): モデルに現れる因子の値ごとに 1 枚 (たとえば値が 2 個なら 2 枚)。
6. 2 次の交互作用 (因子 2 個): 因子の組合せの行列を示すグラフ 1 枚 (たとえば 3 エリア $\times$ 2 様式で 6 セル)。
7. 3 次の交互作用 (因子なし): 複数枚 (たとえば浴室数 = 1, 2, 3, 4 に対して 4 枚)。
8. 3 次の交互作用 (因子 1 個): モデルに現れる因子の値ごとに 1 枚。
9. 3 次の交互作用 (因子 2 個): モデルに現れる因子の値の組合せごとに 1 枚。
10. 3 次の交互作用 (因子 3 個): 相異なる値の数が最も少ない因子の値ごとに 1 枚とし、残る二つの因子については色の行列で価値寄与を表す。

より複雑な場合には、高度なグラフ化の技法が必要となりうる。

他にも、より洗練されたグラフを要する可能性がある。

MARS が生成した完全なモデルの例を [図 1](#) に示す。

11 データの特性

以下では MARS について論じる。MARS は受け取るデータに全面的に依存しており、そのデータの大半は何らかの MLS(マルチプル・リスティング・サービス) ないし税務評価官に由来する。ただし、データを保守・改善して利益を上乗せして販売する他のデータ提供者を経由することもある。地域によってはかなり信頼できるが、そうでない地域もある。重要な場面では、不動産の調査員が自ら観察と測定を行わなければならない、その正確性が問われる。評価エンジニアは、データが何を表しているのか、そして表すべきものをどれだけよく表しているのかを理解する必要がある。ここから一連の問題が生じる。

- ある変数が測定値であるかどうか分類するのか。
- 変数が測定値でないとすれば、他に何でありうるのか。
- 受け取るデータが正確でないとすれば、回帰はどうして有用でありうるのか。

MLS のリスティングサービスや税務評価官から回帰のために得るデータの大半は、測定値ないし個数を意味する数値データである。一部は真偽の論理値である。

11.1 残差の目的

残差は、回帰モデルに含まれていない変数がもたらす価値の間接的な尺度として機能する。ヘドニック分析の文献の用語でいえば、これら測定されていない特性は潜在変数である。すなわち、売却価格に影響を与えるが直接には測定されない変数であり、多くの場合それは客観的に測定する実際的な手段がないためである。したがってその値は推論されなければならない、実務上、潜在変数の値はしばしば主観的判断によって推定される — RCA においては、残差の順位表に係留された判断である。不動産を残差ないし平方フィート当たり残差によって並べることで、魅力の最も乏しいものから最も豊かなものまでの客観的な順位づけが得られる。この順位づけの手法は、地域市場を深く理解した熟練の分析者が MARS(多変量適応回帰スプライン) モデルを構築するかぎりにおいて有効である。これに加えて次の要件がある。

- データとモデルの品質: よく構成された回帰モデルと質の高いデータは、信頼できる残差の前提条件である。サンフランシスコ・ベイエリアにおいて、著者は R^2 が約 75–80%、交差検証 $R^2(CV R^2)$ が 55–70% のモデルを構築することを目標としている。
- 過適合を避け、 R^2 に上限を置く: サンフランシスコ・ベイエリアのような価格水準の高い地域の成熟した市場では、評価エンジニアは R^2 の上限をおよそ 80% に置くべきである。住宅の競争上の価値の多くはその主観的な魅力にあり、それは定量化できないため回帰モデルに直接組み込むことができないからである。 R^2 が 80% を大きく超える場合は過適合したモデルの兆候でありうるので、さらに精査すべきである。もっとも一言注意しておくと、過適合なしに R^2 が 90% を超える市場地域も存在するし、鑑定評価の難しい地域では 50–60% かそれ以下が精一杯ということもある。
- 特定の地域における高い R^2 : ただし、分譲地その他の均質な住宅地域のように、不動産の同質性ゆえに高い R^2 が驚くほど普通に見られる地域もある。

11.2 近似的な回帰変数としての状態評価と品質評価

ファニーメイの状態評価および品質評価の尺度 (C1–C6 および Q1–Q6) については特に述べておく必要がある。これらの尺度は粗すぎて真に測定された特性として扱うことはでき

ず、また鑑定評価者が付与するものである以上、実質的には主観的な評価、すなわちそれ自体が潜在変数である。ここでは二つの水準を明示的に区別しておくのが有用である。回帰変数として用いる鑑定評価者の粗い評価、すなわち粗粒度の状態 (condition-coarse-grain、condition-1) および粗粒度の品質 (quality-coarse-grain、quality-1) と、残差から回収される細粒度の潜在変数、すなわち細粒度の状態 (condition-fine-grain、condition-2) および細粒度の品質 (quality-fine-grain、quality-2) である。VE は鑑定評価者の粗粒度の評価を近似として受け入れ、それらを回帰にかけて近似的な価値寄与を得ることができる。この価値寄与は、細粒度の潜在変数の初期値として有用である。これを行う場合のプロトコルは、この二つの粗粒度変数について回帰が推定した価値寄与をモデルの推定値から差し引き — 取引事例比較表の回帰部分にある測定された属性から外し —、それを残差に戻して加えることである。その後、それらは残差から取り除かれ、対応する細粒度の潜在変数 `res_condition1` および `res_quality1` に割り当てられる。次いで、MLS の公開備考 (public_remarks) および関連する文字列一覧から抽出した追加情報を用いて、これらの潜在変数の値に微細な調整を加える。最終的には、細粒度の価値寄与が粗粒度の価値寄与を置き換える。粗い評価は近似モデルのための足場にすぎず、状態と品質について最終的に報告される値は、残差から導かれた細粒度のものである。

11.3 データの誤り

ソーシャルメディア上の意見によれば、カリフォルニアのマルチプル・リスティング・サービス (MLS) のデータは、米国の他の多くの地域よりおおむね質が高い。一般に、MLS データは成熟した大都市圏のほとんどでは良質であることが期待でき、農村部ではより劣ると考えてよい。

さまざまな理由による誤りがときおりデータに含まれるが、十分なデータが存在するかぎり、これらの誤りは概ね相殺され、価格モデルで用いる平均値への影響は最小限にとどまるはずである。

しかし、特定の地域では系統的な偏りが生じうる。たとえば、多くの住宅が当初は完全な 2 階建てとして建てられたが、一部の買主には 2 階の部屋を取り払って 1 階の一室以上を吹き抜け天井にする選択肢が与えられ、その結果として居住面積 (GLA) が大幅に減少した、という近隣地域を考えてみよう。この変更は何らかの理由でその後も評価官の記録に反映されなかった。その結果、実測値が 2,000 平方フィートであるのに MLS には 2,400 平方フィートと表示されることがある。このような場合、MARS(多変量適応回帰スプライン) 回帰を行う前に MLS データを訂正すべきである。なお、取引事例についてこの情報を得るには、ある程度の調査の労を要することがある点に注意されたい。齟齬は将来の参照のために記録しておくべきである。

以上を踏まえれば、質の高い MLS 地域においては、残差に流れ込む誤りは、デザイン、状態、品質といった主観的で測定されていない特性の影響に比べればごくわずかであると考えて差し支えない。これは RCA 手法の精度を見積もるうえで重要な考慮点である。たとえば、ある市場地域の R^2 の値が 80% であれば、価格のおよそ 20% が CQA 残差に帰属する。したがって、残差の順位表における対象不動産の位置づけに $\pm 10\%$ の誤差があれば、それは最終的な価値結論において $\pm 10\% \cdot 20\% = \pm 2\%$ の誤差として伝播する。精度はこうして高めるのである。多くの場合、 $\pm 1\%$ を大きく下回る正確な価値推定が可能である。

したがって、MARS 回帰においてモデル化されていない変数の集合的な価値の間接的な指標として残差を扱うことは、系統的な測定の偏りがどこになぜ存在するかを見極めるだけの地域市場の経験の評価エンジニアが備えているかぎり、合理的である。この方法は、不動産データの限界と機微を認識しつつ、分析が実際の不動産価値を正確に反映することを確認なものにする。

適用例: シーランチにおける未測定の価値決定要因の測定

ある市場地域で達成可能な R^2 が、その市場の見かけの同質性から予想される水準を下回るとき、本枠組みの対応は、その限界を受け入れることではなく、欠けている変数を特定して測定することである。ソノマ郡のシーランチ (Sea Ranch) 地区がこれを端的に示している。

これは「受け入れるのではなく測定する」という対応の具体例である。ソノマ郡のシーランチ地区は沿岸の開発地であり、海までの距離が主要な価値決定要因である。他の特性を一定とすれば、断崖から数百フィートの住宅は、内陸へ 4 分の 1 マイル入った住宅とは大きく異なる価格をつける。この地区の MLS データには海までの距離が変数として記録されていない。入手可能な MLS 変数で当てはめた最初の MARS モデルは、市場の見かけの同質性から支えられるはずの水準より説明力が低く、残差の構造もデータ上に見える特性によって不動産を十分に順序づけていなかった — これは、市場が本当に雑音に満ちているのではなく、重要な決定要因がモデルから欠落していることの兆候である。

処方は直接の測定であった。著者は海岸線に対する不動産の配置を示す地区の地図を入手し、数百戸の住宅について海までの距離を測定した。測定した距離を変数として MLS データ集に加え、MARS 回帰をやり直した。新しい変数はモデルによって有意なものとして選択され、説明力は大幅に向上し、残差の順位表はその後、それまで順位表を汚染していた潜在的な海距離の信号ではなく、残る主観的特性 — 眺望、状態、品質 — に整合するパターンによって不動産を順序づけるようになった。標高についても、MLS 変数が欠けているか信頼できない丘陵地の近隣地域で同様の対処がなされてきた。

この方法論上の原則は一般的なものである。達成可能な R^2 が市場の見かけの同質性から支えられるべき水準を下回るとき、第一の仮説は、重要な変数がモデルから欠落しているというものであり、適切な対応は、それが何であるかを特定して測定値を直接収集することである。情報源としては、地区の地図、GIS データ、税務評価官の記録、建築部局の記録、衛星画像、現地調査がある。こうした測定作業の費用は現実のものだが限度がある。ある市場地域について一度その変数を収集すれば、その地域における将来の業務へ引き継ぐことができる。RCA が新しい変数をすっきりと取り込めるのは、どの変数が MARS 回帰に入るかについて本枠組みが中立だからである。すなわち、新しい変数が実際の信号を捉えているかどうかは残差の構造が VE に教えてくれ、価値寄与はそれに応じて調整される。

もう一方の対応 — 低い R^2 を受け入れてより広い誤差範囲を報告すること — は、欠けている決定要因が真に測定不能である場合 (買主固有の選好、一時的な市場の心理、個別の取引事情) には防御可能であるが、単に未収集であるにすぎない場合には防御できない。モデルが弱いとき、VE が最初になすべきは、欠けている信号が後者ではないかを問うことである。

11.4 正確性とは何か

評価が完全に正確であるためには、次のことが必要である。

1. 市場価値への適合: 対象不動産の仮定上の売買取引が、所与の市場価値の定義に 100% 従っていること。
2. 正確で頑健なモデルの構築: R^2 がおよそ 75–80%、 CVR^2 が立地に応じておよそ 55–70% の質の高い MARS モデルが構築されていること。
3. 良好なデータ誤り分析: 測定された変数について起こりうるデータ誤りがすべて把握され、抑え込まれており、残差への流出がほとんどないこと。
4. 系統的偏りの考慮: データにおける系統的な偏りが把握され、補正されていること。
5. 残差順位の分析: 残差の順位に並ぶ不動産が、最下位から最上位に向けて特性の合理的なパターンをたどることを確認すること。相続裁判所の売却、短期売却、競売

といった異常値は調査し、説明を加え、場合によっては除外すべきである。

6. 対象不動産の位置づけ: 残差の順位づけを十分に理解したうえで、対象不動産を正確に位置づけること。魅力が明らかに高い不動産と明らかに低い不動産を特定し、その間に対象不動産を置いて採点する。次いで、上位と下位の不動産の間の順位差を可能なかぎり狭める。
7. 説明的推論: 正確性のために説明的推論を犠牲にしてはならない。鑑定評価者は、対象不動産と取引事例との価値の差を、少なくとも通常の理解力をもつ相手に説明できなければならない。たとえば、最適な R2 と CVR2 をもつ良好な MARS モデルは測定された特性の価値寄与を説明するために用いることができ、評価エンジニアによる残差の分解はその他の特性の価値寄与を説明するために用いることができる。
8. 過適合: 単独であれ組合せであれ、一件ないし数件の不動産を特定できてしまう特性を用いないことによって、モデルの過適合を避けること。とりわけ、その値がごく少数の不動産を指し示す因子変数の使用は避けるべきである。相当規模のデータ集合において近隣地域を識別する因子変数は適切だが、プールがごくわずかな地域でプール付き不動産を識別する因子変数は慎重に扱わなければならない。正確な GIS 座標、地番、住所のように個々の不動産を一意に特定する特性は、絶対に許されない。

回帰の R2 が 80% であれば、対象不動産の価値のおよそ 20% が残差に由来すると見込まれる。もっともこれは概算にすぎない。CQA-残差曲線から分かるように、残差の分布の大部分は CQA の下端と上端に集まる。CQA-残差曲線と CQA の低い側・高い側における傾きの急峻さからすれば、これらの領域では残差の推定の精度が落ちると予想されるかもしれない。しかしそれは、これらの領域における不動産の相対的な希少さと、隣り合う不動産の間に見られる魅力の差の大きさによって相殺される。したがって、中間域では $\pm 5\%$ の幅の影響を語りうるとしても、両端ではむしろ $\pm 2 - \pm 3\%$ というやや狭い幅を問題にすることになる。

RCA の精度を計算することは、CQA-残差曲線の非線形性のために一般にはやや難しい。しかし、妥当な残差の順位表があれば、個別の事案ごとに計算することは可能である。CQA スコアの範囲は 0.0 から 10.0 である。良好な MARS モデルがあれば、残差の順位表は魅力の最も乏しいものから最も豊かなものまでの明瞭なパターンを示すはずであり、全範囲の $\pm 5\%$ 以内、すなわち CQA スコアで ± 0.5 以内に評定することは難しくないはずである。回帰に 100 件の不動産が含まれているなら、それは対象不動産を 10 件の集団の中で正確に順位づけることに相当する。そこで、表 3 のバーリングゲームの売買データを参照し、VE が推定した CQA が 7.5、5.1、1.9 の各場合について、対象不動産の価値結論を 300 万ドルとして、およそその精度 (すなわち誤差) を計算してみよう。なお、表 3 は用いた数百件の不動産の一部の抜粋にすぎないので、必要に応じて (線形性を仮定して) 残差を補間する必要がある。念のため述べておくと、ここでは VE が各場合において、わずか 100 件の残差順位のうち ± 5 件の範囲内に対象不動産を順位づけられていると強く確信している、と仮定している。

CQA = 7.5 ± 0.5					
CQA	残差	絶対差	期待誤差	想定売却価格	誤差率
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	残差	絶対差	期待誤差	想定売却価格	誤差率
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	残差	絶対差	期待誤差	想定売却価格	誤差率
2.40	-\$228,411				
1.90	-\$276,381	\$47,970			
1.40	-\$326,938	\$50,557	±\$49,264	\$3,000,000	±1.64%

最後に、正確性を客観的に検証する唯一の方法は、厳格なプロトコル（本稿はその大枠を部分的に示唆するにとどまる）を確立し、その遵守を検証することである。実際の売却価格は市場の不完全性ゆえに市場価値から乖離しうるので、これが必要となる。鑑定評価とその後の売買を比較すれば、正確性についてある程度の示唆は得られる。しかし不動産市場の複雑さと不安定さゆえに、確立された基準とプロトコルに基づいて価値の推定を行うには、鑑定評価者ないし高度な資格を備えた評価エンジニアの専門的能力が必要となる。それらの基準とプロトコルは、有能な評価エンジニアが独立に作業してもほぼ同一の価値結論に至ることを保証するだけの厳格さを備えているべきである。良質なデータが得られる場合の大半において、これは達成可能な目標であるはずだ。

11.5 異常値としての対象不動産

不動産の市場価値を推定するには最近の売買取引の分析が必要である以上、最近の売買取引が、近代化や補修を経ていない古い対象不動産のさまざまな特性をうまく表しているのかを問うのは当然である。これはサンフランシスコ・ベイエリアにおいてとりわけ当てはまる。同地域では、最近の売買取引は通常、売却価格を高めるために改修を経た住宅だからである。したがって、取引事例の品質と状態の平均は、相当期間市場に出ていない古い住宅よりも一般に優れている。この価値の差は、対象不動産の残差の推定、言い換えればその CQA スコアによって捉えられるべきである。これは、対象不動産の評価におけるデータの品質と正確性に関わる重要な考慮点である。

11.6 予期しないデータに対処するうえでの CVR2 の重要性

MARS 回帰は、無作為に選ばれた不動産の部分集合 — 通常は MLS データ集合の 80% から 90% を占め、訓練集合と呼ばれる — に R/earth を適用する処理を何度も反復することに基づいている。残る 10%–20% の不動産は検証集合となる。この段階では、訓練データ集合から MARS モデルを作成し、それを用いて検証データ集合に含まれる不動産の売却価格を予測し、実際の売却価格と比較する。この方法論は、まだ観測されていない不動産の売却価格を予測できる頑健なモデルを構築するための最も効果的な方策として、データマイニングの実務家に認められている。

もっとも、CVR2 という指標の主たる有用性は、実際の市場地域の売買に関して回帰モデルが事実としてどれだけ正確かを反映する点にあるのではなく、指定された価格時点における仮定上の不動産売買についてモデルの予測能力を最適化する点にある。そうした不動産は訓練データ集合に含まれておらず、訓練集合にも検証集合にも見られない新奇あるいは予期しない特性の組合せを示すことがある。

こうした文脈のもとで、たとえば価格時点の 1 か月前に売り出されたとして、過去 15 年間改修されていない不動産の売却価格をどのように推定すればよいのか。強い CVR2 を示すモデルのほうが、この反復的な交差検証を経ていないモデルよりも、最も正確な推定を与える可能性が高い。また、データ集合に、改修・補修・近代化の程度が異なる取引事例が十分な件数含まれていることも決定的に重要である。

究極的には、取引事例のデータ集合においてさまざまな不動産の特性の価値寄与を正確に予測する頑健なモデルを手に行っているならば、そのモデルによる残差の順位表の中に新たな対

象不動産を効果的に位置づけることができる。この位置づけによって、当該不動産に適切な CQA スコアと残差を割り当てることが可能になる。

12 残差

「残差制約アプローチ」という名称が示すとおり、残差はこの方法論の中心にある。ここで残差とは、ある不動産の純売却価格と、回帰分析によって予測された売却価格との差を指す。

正の残差は、居住面積 (GLA) や敷地面積といった典型的な測定可能な属性に基づく推定価格よりも高く売れたことを示し、平均以上の魅力を備えている可能性を示唆する。負の残差は、より低く売れたことを示し、魅力が乏しいか、捉えられていない負の属性が存在する可能性を示唆する。

規模の大きい住宅ほど残差も大きくなりやすいことから、より標準化された比較のために GLA に基づく平方フィート当たり残差を用いることが考えられる。この問題のより立ち入った議論は別稿に譲る。

12.1 順位づけと採点

測定された変数について MARS(多変量適応回帰スプライン) モデルを作成した後、不動産ごとに残差を計算する。これらの残差は、観測された純売却価格とモデルによる予測売却価格との差を表す。残差/SF、すなわち残差を住宅の居住面積で除した値の列を追加すべきである。一般に、個々の不動産の残差は居住面積が大きいほど大きくなるという理解に基づく措置である。近隣地域によっては他の統計量のほうがうまく機能することもあるが、それは別の機会に論ずべき主題である。

前提 7: 「残差による順位づけ」

本稿における「残差による順位づけ」への言及は、残差による順位づけ、残差/SF による順位づけ、あるいは市場地域に最も適切と評価エンジニアが判断する残差の他の関数による順位づけを意味するものと解されたい。サンフランシスコ・ベイエリアにおける著者の経験によれば、とりわけ居住面積 (GLA) が異例に小さいか大きい対象不動産を扱う場合、残差/SF が一般に最も信頼できる。

次に、不動産を残差の最も小さい (負の大きい) ものから最も大きいものへと並べる。そのうえで各不動産について、0.0 から 10.0 の範囲をとる状態・品質・魅力 (Condition-Quality-Appeal, CQA) スコアを生成する。このスコアは次のように計算する。

CQA スコアは、全不動産の中における当該不動産の残差のパーセンタイル順位に対応する。具体的には、当該不動産より残差が低い不動産の割合を求め、それを 10 で除して算出する。たとえば、ある不動産の残差が全体の 50% の不動産の残差より高ければ、そのスコアは 5.0 となる。

端数処理の結果、残差が最も高い不動産が 10.0 のスコアを得ることがある。しかし論理的には、自らを含む 100% の不動産より高い残差をもつ不動産は存在しえないので、10.0 という満点は理論上ありえない。他方、最低点の 0.0 はありうる。少なくとも 1 件は、それより残差の低い不動産が存在しないからである。なお、比較対象の不動産が多い場合、端数処理により複数の不動産が 0.0 のスコアをとることもある。

したがって CQA スコアは、ある不動産の属性がモデルの想定にどれだけ合致しているかを直観的に測る尺度を与える。スコアが高いほどモデルの予測との整合がよいことを示す。

もっとも、CQA スコアが取引事例の魅力の間接的な指標たりうるのは、売買の条件がその市場価値に見合っている場合に限られることを認識しておく必要がある。残差の値に市場水準を超える、あるいは下回る要素が含まれていないかどうかを、当該売買取引について包括的な説明と徹底した調査を行って見極めるのは、評価エンジニアの責任である。市場水準を大きく超える、あるいは下回る売買は頻繁ではないが、強制売却、相続裁判所の手続、当事者

の無能力、通謀といった状況では生じうる。

表 3 は、MLS データの半ば現実的な抜粋を、残差および CQA スコアとともに、総額と平方フィート当たりの双方について示したものである。本研究で分析した不動産の総数はおよそ 600 件であった。回帰にかけた変数のいくつかは表示されていない点に留意されたい。言い換えれば、ここに示す残差は、売却日その他の変数の影響を MARS 回帰が考慮した後のものである。

表 3: 残差で並べ替えた取引事例

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-残差曲線

ある近隣地域について CQA を残差/SF に対してプロットすると、必ず 図 2 に示す CQA-残差特性曲線が得られる。この図は 2022 年頃のカリフォルニア州バーリングゲームのデータを表している。CQA スコアを残差の値に対してプロットしても一般に同様の曲線が観察されるが、著者は残差/SF を用いることを好む。対象不動産の居住面積 (GLA) との関係の中で CQA を位置づけるうえで適切だからである。残差/SF に GLA を乗ずれば、ある不動産の CQA を対応する残差の値へ変換できる。

評価エンジニアは、不動産の規模、具体的には居住面積 (GLA) が残差の大きさに大きく影響するかどうかを慎重に評価しなければならない。この判断は、不動産価値を計算する最も正確な方法を決めうるという点で決定的である。たとえば、高級な厨房設備や豪華な浴室設備は住宅の残差の価値を大きく高めうる。しかし、厨房と浴室の広さや配置が概ね一様な近隣地域では、こうした改良は GLA と強くは相関しないかもしれない。残差の価値への影響は、それらが占める空間よりも、それらが体现する品質によるところが大きいのである。

逆に、不動産の外観の印象 (curb appeal) を著しく高める高価な外構の特徴や資材は、不動産の規模とより直接的な関係をもちうる。規模の大きい住宅ほど、広範な造園、上質な屋根材、広々としたパティオといった手の込んだ外構を備えることができ、それが自然に不動産全体の魅力を、ひいては残差の価値を高める。

したがって評価エンジニアは、残差を分析する際にこうした力学を考慮することが不可欠である。質の高い設備や特徴が付加する内在的な価値を評価するだけでなく、それらの改良が不動産の物理的な規模とどのように相互作用するかを理解しなければならない。この包括的な方法によって、評価は不動産の質的属性と量的属性の双方を正確に反映することになる。

そこで、CQA-残差曲線と CQA-残差/SF 曲線の双方を描いて差異を検討することが推奨される。それは市場地域におけるパターンを示唆しうる。評価エンジニアは自らのプログラムに両方の残差を計算させ、何らかの加重平均をとって対象不動産に割り当てる残差を算出することもできる。それが大きな違いを生むこともあれば、生まないこともある。

12.3 CQA 曲線の特性

状態・品質・魅力 (CQA) 曲線は、ある地域における不動産価値が魅力、状態、市場の力学に応じてどのように変動するかを理解するための精緻なモデルを与える。このモデルは、不動産市場に影響を及ぼす経済的要因と物理的要因の相互作用を捉え、さまざまな市場区分にわたる不動産評価の機微に富んだ現実を映し出す。

12.3.1 CQA 曲線の構成要素の分析

1. 曲線の左側における指数的減衰:

CQA スコアがおおよそ 2 から 0 へ下がるにつれて、不動産は急速な劣化の特徴を示す。曲線のこの部分は指数的減衰関数によってよく表され、適切な維持管理と改修がなければ不動産がいかに急速に魅力と価値を失いうるかを浮き彫りにする。木部の腐朽、色褪せ、金属の錆といった物理的な劣化がこの低下を加速させ、とりわけ定期的な維持管理がなされていない住宅で顕著である。

2. 中間域における線形の上昇:

市場の大半は通常この区分に収まり、スコアがおおよそ 2 から 8 まで、魅力はほぼ線形に増加する。この部分は、おおむね標準的な水準で維持されているが、必ずしも例外的な品質や立地を備えているわけではない不動産を表す。標準的な仕様と全体的な状態に基づいて競争的に価格づけされる住宅が大半を占める、住宅市場の広い中間層を表すものである。

3. 曲線の右側における指数的增长:

ここではスコア 8 から 10 にかけて曲線が指数的に上昇し、市場の魅力の高い側を映し出す。この区分の不動産は、優れた維持管理に投資するだけでなく、不動産の魅力と価値を大きく高める改良にも投資する富裕層に属することが多い。指数的な成長は、そうした高品質の不動産の希少性と、富裕な買主の間の強い需要を反映している。また、価格の低い住宅が魅力の点で非常に高い評価を得ることもあれば、大規模で高額な住宅が低い評価にとどまることもある点を理解されたい。こうした力学は MARS の熟達した使用によって明らかになるものであり、多くの鑑定評価者は通常これを見落とす。

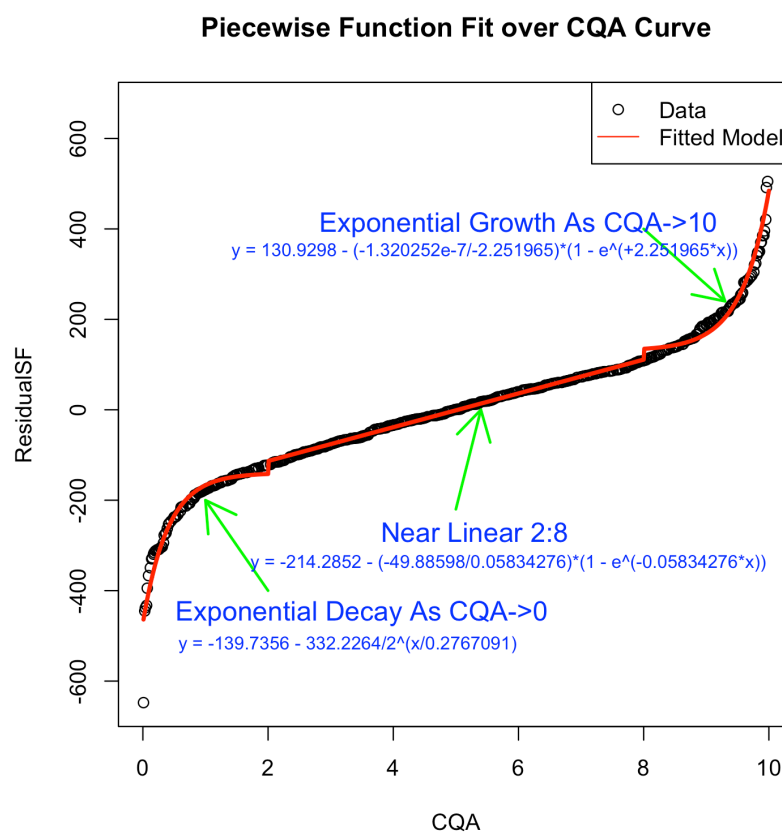
12.3.2 市場の力学と社会的含意

CQA 曲線は、地域ごとの相対的な富と経済的格差の物語を暗黙のうちに語っている。サンメイトオ郡のような地域では、バーリングゲームと、ヒルズボロ、ウッドサイド、アサートンといったより富裕な近隣地域との間の不動産価値の差が、著しい社会経済的な分断を示している。この格差は、不動産の維持管理から市場評価に至るまであらゆることに影響し、ひいては買主の属性や不動産市場の趨勢を左右する。

さらに、比較的富裕な個人が、自らが心地よく感じ、さまざまな理由で住みたいと考える中位層の近隣地域において高品質の住宅に投資する、という逸話は、不動産経済のもう一つの重要な側面 — 「過剰投資 (overbuilt)」の不動産という概念 — を際立たせる。そうした不動産は非常に魅力的ではあるものの、価格水準のより高い市場地域であれば得られたであろう価格を支払える買主を、地元では見つけられないことがある。

CQA-残差曲線は市場地域ごとに固有のものであり、分析にあたって見過ごしてはならない。

図 2: CQA-残差特性関数



12.4 対象不動産の残差

取引事例とは異なり、対象不動産には客観的に定まる残差が存在しない。それは評価エンジニア (VE) が推定しなければならない。この推定は残差制約アプローチ (RCA) のプロトコルにある程度の主観を持ち込むものであり、個人の判断が重要な役割を果たす決定的な局面を画する。

状態・品質・魅力 (CQA) スコアを推定するため、VE は順位づけされた不動産を丹念に分析し、順位表の中に識別可能なパターンを探す。この分析にはしばしば、各不動産に付された MLS の写真の詳細な検討が伴う。600 件といった大きなデータ集合について網羅的に検討することは現実的でない。そこで VE は実務的な推論を用いて、全体の順位表の中で対象不動産が占めるであろうおおよその位置を見定める。大規模な補修を要する不動産、いわゆる「要修繕物件 (fixers)」は概して下位に位置し、優れた改修と魅力を備えたものは上位に位置する傾向がある。

次に VE はより絞り込んだ手法を取り、順位表の中を「跳び移り」ながら、対象不動産の魅力に近い取引事例を特定する。範囲が狭まるにつれて、VE は不動産どうしの類似性、とりわけ品質、状態、意匠上の美観 (塗装、木工、石工、窓などの要素を含む)、機能的効用その他の関連する特徴により注意を払う。

重要なのは、VE の判断が根本において市場が形成した順位表に導かれているという点である。これにより、この過程の主観的要素は市場の現実に係留され、対象不動産の公正かつ正確な評価が目指される。

12.5 RCA の価値結論

12.5.1 一般的な考慮事項

CQA スコアの決定は、不動産の残差の推定値を確定するうえで重要である。両者は本質的に結びついているからである。CQA なしに不動産を評価することも可能ではあるが、この指標は、当該不動産が市場地域の他の不動産と比べてどこに位置するかを示すことで有益な文脈を与える。たとえば CQA スコア 3.2 は、その不動産が周辺の取引事例の 32% より魅力が高いことを示す。付与されたスコアに異議がある場合、異議を唱える者は順位表を参照し、写真による証拠その他の実質的なデータに基づいて提案する修正を正当化しなければならない。順位づけの仕組みが頑健であり、評価エンジニア (VE) がその枠組みの中に不動産を正確に位置づけているなら、大きな齟齬が生じることは考えにくい。

報告の段階では、通常、価値結論を裏づけるために最近の関連性の高い取引事例が 6 件から 12 件選定される。これらの取引事例と対象不動産との差異が分析され、変数ごとに補正が行われる。次にこれらの補正額を取引事例の純売却価格に適用して補正後売却価格を導く。数学的に示したとおり、これらの補正後価格は対象不動産の推定売却価格に収束するはずである。

12.5.2 検証、監査、苦情

対象不動産の推定残差が定まれば、すべての取引事例について残差の補正額を計算できる。これらの補正額は、関連すると判断された個別の CQA 変数に応じて VE がさらに手作業で分解を行う際の制約となる。強調しておくべきは、証明 II が確認するとおり、これらの詳細な分解は説明のためのものであり、最終的な評価には影響しないという点である。

鑑定評価の価値結論に不満をもつ利用者からの苦情に関しては、従来型の鑑定評価者は通常、価格の高いあるいは低いこの物件やあの物件がなぜ利用者の期待する価値への影響をもたないのかを説明するのに時間を空費しなければならない。RCA では、ありうるすべての取引事例 (通常 100-600 件) が価格結論とまさに同一の価値へ評価されるので、任意の不動産について

て補正額を示すのにほとんど労力を要しない。RCA の過程は、対象不動産の残差の採点を除いて、全体が自動化され客観的である。したがって RCA のもう一つの利点は、VE の主観的判断の不完全さに対する批判が主としてその一点に限られ、しかもそれを効果的かつ効率的に管理できることにある。また、 R^2 を約 80% と仮定すれば、対象不動産の残差の推定における $\pm 10\%$ の誤差が最終的な価値推定に及ぼす影響は $\pm 2\%$ の誤差にとどまる。

多くの場合、対象不動産は不動産の残差順位の中に非常にうまく収まるため、精度は $\pm 1\%$ 以内と見込むことができる。

推定残差と対象不動産の MARS による推定売却価格との和が、残差制約アプローチ (RCA) の価値結論と呼ばれる最終的な価値結論を与える。この手順は不動産評価に包括性と透明性をもたらし、評価の正確さと公正さに対する信頼を育む。

13 結論

残差制約アプローチは、三つの実質的な貢献をもつ枠組みである。第一に、少数の取引事例の集合ではなく市場全体に対してモデルを当てはめることから導かれる、構造的なアンカリング抑止の性質である (証明 I)。第二に、残差の分解が価値結論を乱すことなく説明のための層として機能することを可能にする不変性の結果である (証明 II およびそれに続く注記)。第三に、経験的な規則性 — CQA-残差特性曲線 — であり、これは市場地域を越えて、指数的・線形・指数的という特徴的な形で繰り返し現れるように見える。この枠組みの限界もまた同じく実質的である。残差の分解は不変ではあるが識別的ではなく、CQA の位置づけの手順は VE の判断に依拠しており、その判断はいまだ正式な評価者間一致の研究によって検証されておらず、精度に関する主張も代替手法との比較検証を待っている段階にある。

本稿が提示するのは枠組みとプロトコルである。本稿がいまだ提示していないが、現在進行中のより大きな研究計画が提示しつつあるのは、防御可能な既定値を備えたソフトウェアとしてこのプロトコルを実装する、実行可能で監査追跡が可能なパイプラインである。そのパイプライン — MARS の基底関数の段階を担う CRAN の earthUI パッケージ、それらの基底の上に正則化された構造モデルを構築する glmnetUI パッケージ、そして空間平滑化を担う mgcvUI パッケージ — は、本号の姉妹論文が扱う主題であり、方法としての RCA から成果物としての RCA への架け橋である。鑑定評価規制における教義と実務の乖離 (Craytor, 2026、本号) は、その架け橋がなぜ重要かを示している。すなわち、鑑定評価の産物は経済的には将来を見据えたリスク情報として消費されており、現在価値についての意見を監査可能かつ再現可能にする枠組みは、その将来志向の用途に誠実に向き合うための前提条件である。RCA は、このより大きな企ての一手にすぎない。それが企ての全体ではないし、本稿もそう主張するものではない。

参考文献

- Ambrose, Brent W. et al. (2025). “Do Appraiser and Borrower Race Affect Mortgage Collateral Valuation?” In: *Review of Finance*. DOI: [10.1093/rof/rfaf046](https://doi.org/10.1093/rof/rfaf046).
- Craytor, William Bert (2026). “The Doctrine–Practice Gap in Real Estate Appraisal: A Structured Account of Functions, Boundaries, and Tensions”. In: *Valuation Engineer Journal* 1.1.
- Friedman, Jerome H. (1991). “Multivariate Adaptive Regression Splines”. In: *The Annals of Statistics* 19.1, pp. 1–141.
- Hastie, Trevor, Robert Tibshirani, and Jerome H. Friedman (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second. Springer Series in Statistics. New York, NY: Springer. ISBN: 978-0-387-84857-0.
- James, Gareth et al. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics 103. New York: Springer. ISBN: 978-1-4614-7137-0.
- Krivorotov, George and Michael LaCour-Little (2020). “AVM versus Appraisal-Based Underwriting in Refinance Mortgages: The Trade-off Between Noise and Bias”. In: *SSRN Electronic Journal*. DOI: [10.2139/ssrn.3603327](https://doi.org/10.2139/ssrn.3603327).
- Kuhn, Max and Kjell Johnson (2016). *Applied Predictive Modeling*. Corrected at 5th printing. New York: Springer. ISBN: 978-1-4614-6848-6.
- Milborrow, Stephen (2024a). *Earth: Multivariate Adaptive Regression Splines*. R Package. URL: <https://CRAN.R-project.org/package=earth>.
- (2024b). *Notes on the Earth Package*. Tech. rep. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (2024c). *Variance Models in Earth*. Tech. rep. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.

- Perry, Andre M., Jonathan Rothwell, and David Harshbarger (2018). *The Devaluation of Assets in Black Neighborhoods*. Tech. rep. Brookings Institution.
- Perry, Andre M., Jonathan Rothwell, J. Williamson, et al. (2024). *How Racial Bias in Appraisals Affects the Devaluation of Homes in Majority-Black Neighborhoods*. Tech. rep. Brookings Institution.
- Property Appraisal and Valuation Equity Task Force (2022). *Action Plan to Advance Property Appraisal and Valuation Equity*. Tech. rep. U.S. Department of Housing and Urban Development.
- Tzioumis, Konstantinos (2016). “Appraisers and Valuation Bias: An Empirical Analysis”. In: *Real Estate Economics* 45.3, pp. 679–712. DOI: [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- Williamson, Jake and Mark Palim (2022). *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Tech. rep. Fannie Mae.
- Zhang, Wengang (2020). *MARS Applications in Geotechnical Engineering Systems: Multi-Dimension with Big Data*. Singapore: Springer. ISBN: 978-981-13-7421-0.