

条件付き期待値 (Conditional Expectation)

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

DOI: [10.67330/99hhbc69](https://doi.org/10.67330/99hhbc69)

形式的定義。 一連の予測変数（測定された特徴量）が与えられた場合の、結果（ここでは売買価格）の条件付き期待値とは、予測変数の各固定値における結果の分布の平均値である。予測変数の関数として見なすと、これは条件付き期待関数（CEF）と呼ばれ、回帰分析が近似を目指す関数である。普通最小二乗法（OLS）回帰は、CEF に対する最良の線形近似を与え、CEF 自体が線形である場合には CEF を正確に再現する。

直観的な説明。 回帰分析から導き出される補正は、ある「単価」を起点としています。つまり、市場が 1 平方フィートあたり、バスルーム 1 つあたり、築年数 1 年あたりに支払う金額です。この記事では、その単価がどこから導き出されるのか、そしてそれが何を基準としたものなのかについて解説します。簡単に言えば、ある特徴が変化し、他の特徴が一定に保たれた場合の市場平均価格の傾きであり、回帰分析とは、販売データからその傾きを測定する方法である。OLS 回帰は、その特徴量の全範囲にわたって単一の傾きを測定します。一方、MARS（多変量適応回帰スプライン、Friedman 1991）は、市場での支払額が異なる各範囲ごとに傾きを測定し、範囲ごとに 1 つの単価を報告します。どちらの場合も、その単価は同じもの、すなわち平均価格がその特徴量に応じてどのように変動するかを測定したものであり、この項の残りの部分では、それが何であるか、なぜ販売比較表でもそれが測定されるのか、そして同じ販売データに対する 2 つの回帰分析が 2 つの異なる単価を示す場合、それが何を意味するのかについて解説します。

仕組みを知りたい読者のために。「特徴量が増加するにつれての平均価格」には「条件付き期待関数」という名称があり、この記事では以降、この用語を使用する。これは特徴量の空間上の曲面であり、ある方向におけるその傾きが単価となる。OLS は、その曲面を最もよく近似する平面を当てはめ、曲面が平面である場合にのみそれを正確に再現します。一方、MARS は、検出した区切り点で曲がる平らな断片からなる曲面を当てはめるため、真の曲面が曲線を描いている場合にはより正確な近似を得ることができます。第 1 号に掲載された 8 件のパシフィカ比較事例が、その実例として挙げられています。

この記事は法律というよりは統計学に関するものです。回帰分析によって何が推定できるかについて扱っており、回帰分析を行う鑑定評価者や、販売事例比較表（これも同様に同じことを推定するものです）を使用する鑑定評価者を対象としています。

第 1 号で実施されたパシフィカの比較事例に関する回帰分析を用いて、その係数（回帰分析によって各特徴に割り当てられた比率）から潜在価格を導出した。潜在価格とは、市場が特徴の 1 単位に対して支払う金額のことである。しかし、回帰分析で何が推定されたのかについては明確に述べられておらず、この点に言及しておく価値がある。というのも、本号で続くバイアス、内生性、識別性に関するほぼすべての論点は、推定された内容と望まれていた内容との相違に関連する問題だからである。

回帰分析は通常、残差の二乗和（残差とは、実際の売買価格から回帰式で算出された価格を差し引いた値のこと）を最小化して直線を当てはめ、その傾きを算出するという形で説明されます。この説明は、関わる算術的な処理に関しては正しいものの、その目的という点では誤解を招く恐れがあります。計算はあくまで手段に過ぎず、真の目的は、手元にある売上データだけでなく、市場全体（母集団）を記述する量を見出すことにある。

記号の読み方。

P 売買価格。

$\mathbf{z}$	不動産の測定値の一覧。このような一覧は「ベクトル」と呼ばれる。
$\mathbf{z}$	そのリストに対する特定の値の組み合わせ。例えば、1,650 平方フィート、7,200 平方フィートの敷地、眺望なし、状態 7.00 など。
$E[\cdot]$	期待値：すべての取り得る値の平均値であり、各値はその確率に応じて重み付けされる。
$ $	「与えられた」。
$E[P \mathbf{Z} = \mathbf{z}]$	$\mathbf{z}$ に等しい特徴を持つ不動産の価格の期待値。
ε	その期待値からの販売実績の乖離。
u	販売における構造的な誤り：その売買価格から、因果関係に基づいて想定されるべき価格を差し引いたもの。
$\hat{p}$	当てはめ式から得られる値。
$z_{\text{GLA}}, \dots$	各特徴の値：GLA および敷地面積（平方フィート）；眺望とリフォームの有無（ある場合は 1、ない場合は 0）；スコアは 0.00～10.00 の範囲で設定する。

特性ベクトル $\mathbf{z}$ を固定する——つまり、1,650 平方フィート、7,200 平方フィートの敷地、眺望なし、そして状態は「良好」（『ジャーナル』の状態評価尺度で 7.00、解説例で説明済み）とする。市場に出回っている不動産のうち、まさにこの特性ベクトルを持つものについては、価格にばらつきがあります。価格が同じにならないのは、不動産ごとにベクトルでは捉えきれない違いがあること、買い手と売り手が異なること、そして交渉の結果、必ずしも同じ価格で決着するとは限らないためです。その市場には、 $\mathbf{z}$ に対応する価格の完全な分布が存在する。この分布の平均は、 $\mathbf{z}$ が与えられた場合の価格の条件付期待値であり、 $E[P | \mathbf{Z} = \mathbf{z}]$ と表記される。この一般的な定義はコルモゴロフ（1933）によるものである。成約価格の平均は、この期待値そのものではなく、その推定値である。

不動産の特性が特性空間全体にわたって分布しているため、条件付き期待値は曲面を描き出します。各点には数値があり、それはその特性を持つ不動産の平均価格です。この曲面は、条件付き期待値関数（、CEF）として知られています。これは市場の特徴であり、特定のデータセットに属するものではありません。たとえこれまで一度も不動産の評価が行われていなかったとしても、この曲面は存在していたでしょう。図1は、本例におけるその曲面を示しています。

2 つの事実により、CEF は推定の自然な対象となります。第一に、 $\mathbf{z}$ のすべての関数の中で、CEF は予測誤差の二乗の期待値を最小化するものです。 $\mathbf{z}$ のみに基づいて価格を推測し、その誤差の二乗によって評価される場合、 $E[P | \mathbf{Z} = \mathbf{z}]$ が最善の推測となります。第二に、あらゆる結果は

$$P = E[P | \mathbf{Z}] + \varepsilon, \quad E[\varepsilon | \mathbf{Z}] = 0,$$

の形で表すことができる。この 2 番目の式は仮定ではなく恒等式である。なぜなら、条件付き平均からの偏差は、定義上、条件付き平均がゼロとなるからである（Angrist & Pischke 2009, ch. 3）。

この項目は、項目 008（省略変数バイアス）から 010（パラメータ同定）よりも前に位置づけられている。なぜなら、これらの各項目は、回帰分析によって推定される値が、対象とする CEF であるかどうかを問うものだからである。 $E[\varepsilon | \mathbf{Z}] = 0$ という恒等式自体は常に成り立つ。なぜなら、 ε は条件付き平均からの距離として定義されているからである。成立しなくなる可能性があるのは、別の誤差（構造的誤差）について、同じ条件、すなわち条件付き平均がゼロであるという条件が満たされない場合である。構造的誤差は、項目 009（内生性）で u と表記されるものであり、価格と、鑑定評価者が測定しようとしている因果関係との差である。この誤差は、特徴量とともに変動する可能性があります。これは、例えば、価格に影響を与える特徴量が除外された場合に発生します。価格が特徴量に逆影響を及ぼす場合——通常、価格が上昇すると市場滞在日数（DOM）も増加するため、DOM を用いて価格を説明する回帰分析では、価格自体が DOM に及ぼす影響の一部が混入してしまう。これが、売買価格の回帰分析に DOM を含めるアナリストがほとんどいない理由である；あるいは、特徴量の測定に誤差が含まれている場合などである。省略変数バイアス（項目 008）と 内生性（項目 009）はこうしたメカニズムに関するもので

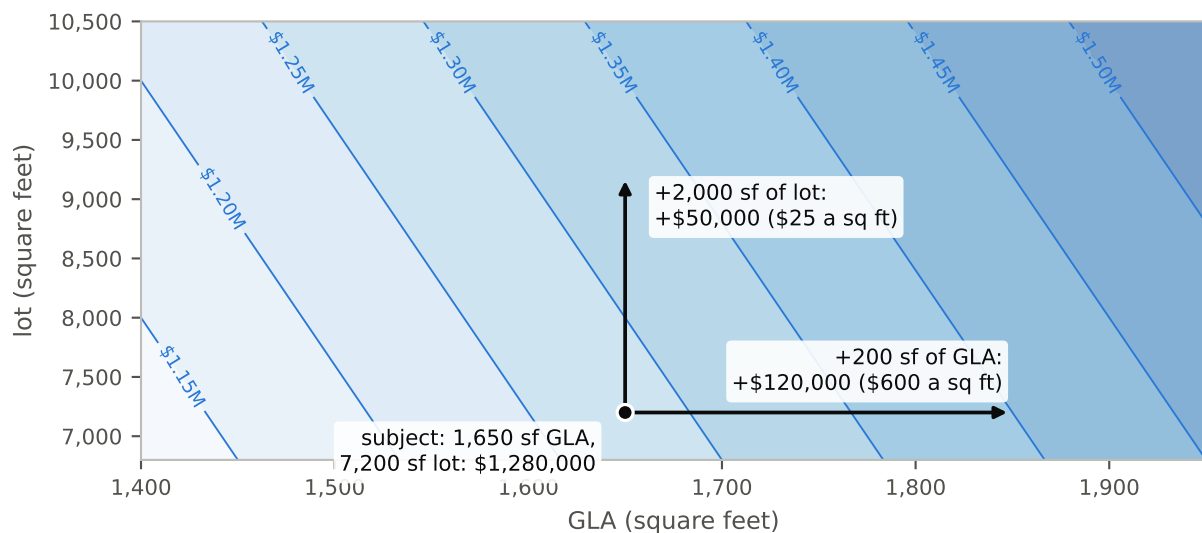


Figure 1: 構築されたパシフィカ市場の条件付き期待値関数（演習例）。眺望がなく、コンディションスコアが7.00、リフォーム歴のない住宅について、地形図のように描画したものです。各線は、平均価格が同じ GLA と敷地の組み合わせを結んでおり、平均価格が上がるにつれて陰影が濃くなっています。マークされた点は、本エントリーの対象となる住宅であり、GLA が 1,650 平方フィート、敷地面積が 7,200 平方フィートの土地に建ち、住所は \$1,280,000 である。2 本の矢印は 2 つの傾き（単価）を示している。GLA が 200 平方フィート増加すると平均価格は \$120,000 上昇し、これは 1 平方フィートあたり \$600 に相当する。敷地面積が 2,000 平方フィート増加すると、平均価格は \$50,000 上昇し、これは 1 平方フィートあたり \$25 に相当する。

あり、パラメータ同定（項目 010）は、販売データがそもそもその単価を示すことができるのかという、より根本的な問いを投げかけるものである。

評価実務で出会う場面。「この地域では、面積が 1,650 平方フィートで眺望のない住宅は、\$1.28 million 程度で売却される」と述べる鑑定評価者は、条件付き期待値を示している。ここで使われている言葉は「約」である。この発言は、そのような住宅がすべて \$1.28 million で売却されると主張しているわけではなく、むしろ \$1.28 million がそのタイプの住宅の売却価格の平均値であることを意味している。「約」という言葉を用いることで、鑑定評価者は個々の価格がその値を中心にばらつきがあることを認めているのである。

販売比較グリッドとは、評価対象の特性ベクトルにおける CEF の値を推定する手法である。各比較事例は、その比較事例の特性ベクトルにおける価格分布から得られた観測値であり、各補正は、その観測値を CEF 表面上で評価対象のベクトルの方向に沿ってシフトさせるものである。調整後の結果は、評価対象における CEF の高さの推定値となる。構造的に言えば、販売データから補正を抽出した場合、この比較表は条件付き平均の単純な非パラメトリック推定量（表面に関する式を仮定しないもの）と見なすことができる——これは、OLS の代わりに局所平均化と手動で指定された傾きを用いるものである；コストや判断に基づいて行われる補正は、構造化された意見であり、そのような種類の推定値ではない。

このように比較表を構成することで、鑑定評価者の暗黙の前提が明らかになります。比較事例が使用される場合、CEF は十分に平滑であり、近隣の地点の値が評価対象の地点の価値に関連すると仮定されます。線形補正は、その特定の特性に関する CEF の傾きが、対象範囲全体を通じてほぼ一定であることを前提としています。3 つの指標に等しい重みを与えることは、各指標が同じ条件付き平均を推定しており、体系的なずれがなく、かつすべてが同程度の精度を持っていることを前提としています。これらの仮定はいずれも珍しいものではなく、すべて曲面の形状に関するものであり、問題の曲面が特定されれば、すべてが議論の対象となる。

この枠組みは、予測区間に関するよくある誤解も解消する。不確実性には 2 種類ある。1 つは CEF がどこにあるか、つまり当てはめられた曲面が真の平均値からどれほど離れているかという点に関するものである。もう 1 つは、個々の価格がその平均値の周囲にどれほどばらついているかという点である。信頼区間は前者のみをカバーするが、予測区間は両方をカバーするため、常に 2 つのうちで幅が広い方となる。『Journal』の主要な回帰分析手法である MARS では、earth パッケージ（Milborrow 2026a）が、市場全体における価格のばらつきの変化を推定する第 2 のモデル、すなわち分散モデル（Milborrow 2026b）を用いて、評価対象に対する予測区間を算出する。前提条件として、十分な取引実績があり、誤差が大まかに正規分布に従い、互いに独立していることが求められる。このパッケージは、取引価格が予測区間内に収まる頻度を報告し、それが検証の根拠となる。RCA に基づく鑑定では、販売比較法に従い、評価対象の市場価値は、モデルによって推定された売買価格に、鑑定評価者が推定した残差を加えたものとなる。分散モデルは、これらの条件が満たされていることを前提として、評価対象の売買価格が 95% の確率で収まる範囲を示す。鑑定評価者に「どの程度確信がありますか？」と尋ねられた場合、通常は個々の価格のばらつき（予測区間）について問われているのですが、鑑定評価者はたいてい、平均値（信頼区間）に対する確信度を答えます。

推定対象——すなわち、推定対象となる母集団の特性——に名前を付けることは、正当性を示す行為である。なぜなら、それなしでは結論を判断できないからである。「補正額は \$100,000 である」という記述は、現時点では正しいとも間違っているとも判断できない。「総居住面積（GLA）、敷地、および状態を一定に保った場合、眺望のある不動産とない不動産との間の条件付き売買価格の差は、\$100,000 である」と述べたとき初めて、何を一定条件としたか、対象となる母集団、および対象となる物件数といった点に基づいて議論の余地のある主張となるのです。

妥当性の裏づけにのっての意義。これに伴い、3 つの正当性の帰結が生じる。

条件付け集合（一定に保たれる特徴量）の選択は、方法論上の問題である。これは推定対象を定義するものである。データによってその推定対象を推定できるかどうかは別の問題であり、

これはパラメータ同定の評価対象となる。式 $E[P | GLA, view]$ と $E[P | GLA, view, condition]$ は異なる関数を表しており、これらは異なる値や異なる傾きを持つ可能性があります。鑑定評価者が比較表に変数を追加する場合、単に推定値を改善しているのではなく、推定対象そのものを変更していることになります。妥当な姿勢としては、価値意見の根拠となる条件付け集合を明示することです。

「他の条件が同じである」という表現は、鑑定評価者が分析においてどのような推論を行うかを表すものであり、明確な意味を持っています。CEF の枠組みにおいて、これは、たった 1 つの特性のみが異なる 2 つの特性ベクトルにおける条件付き平均を比較することを意味します。これは母集団に関する記述であり、十分な市場データが存在する場合、それに対して検証を行うことができます。その背後にある分析は、通常、報告書には記載されず、ワークファイル (USPAP の「記録保持規則」に基づき、鑑定評価者が保持することが義務付けられている業務記録) に記載される。

残差のばらつきは情報であり、恥じるべきことではない。条件付き期待値関数を中心とした価格のばらつきは市場の特性であり、価格がその周囲に広くばらついている市場では、どのような手法を用いても、単一の数値による推定には大きな不確実性が伴うことになる。このばらつきに言及せずに数値を提示することは、市場が許容する範囲以上に結論を正確であるかのように見せかけることになる。

評価計算例。『』第 1 号に掲載されたパシフィカの 8 件の比較事例価格は、ノイズのないサンプルである。これらは、単一の線形方程式を正確に満たすように構築されたものである。これら 8 つの価格を GLA、区画、眺望、状態、およびリフォームの有無という変数で回帰分析すると、すべての価格が 1 ドル単位まで再現され——余剰項は一切残らない——、回帰分析の結果として、価格が構築された元の式が、各単価ごとに正確に導き出される：

$$E[P | Z] = -30,000 + 600 \cdot z_{GLA} + 25 \cdot z_{lot} + 100,000 \cdot z_{view} + 20,000 \cdot z_{cond} + 60,000 \cdot z_{remodel}.$$

ここでいう「状態」とは、本ジャーナルの状態スコアを指します。これは、定義された市場地域内の周辺住宅の状態と比較した、当該住宅の状態の推定値です。これは、鑑定評価者による当該市場地域の理解に基づいており、同じ市場地域で活動する他の鑑定評価者の理解と著しく矛盾してはなりません。スコアは 0.00 から 10.00 までの範囲で表され、市場地域内でこの物件よりも状態が悪い住宅の割合を 10 で割った値である。スコア 5.00 は、市場地域内の住宅の半数がこの物件よりも状態が悪いことを意味し、9.00 は 10 戸中 9 戸がそうであることを意味する。第 1 項では、状態を言葉で記述し、回帰分析のために 1~3 のコードを割り当てた。本ジャーナルのスコアでは、これらの言葉は「平均」5.00、「良好」7.00、「優秀」9.00 に対応するため、第 1 項の 1 段階は 2 ポイントに相当し、市場では 1 ポイントあたり \$20,000 が支払われる。最後の項は、構築されたデータがリフォームに付与する価格差である：\$60,000 は、同じ GLA、敷地、眺望、状態スコアを持つ住宅よりも高い価格である。これはリフォームという行為そのものの価格ではなく、他の特徴では記録されていない、リフォームがもたらした何らかの要素を表している (項目 008、省略変数バイアス)。

これは、まさに構築されたパシフィカ市場の CEF です。先ほどどのような対象が記述されたのか、そしてなぜそれほど正確に記述できるのかを明確にすることが重要です。このデータセットは構築されたものであり、構築されたデータセットは実際にはノイズを含まないことがありますが、実際のデータセットには常にノイズが含まれています。これを構築する利点は、価格からランダムなノイズを完全に除去することにある——すべての価格が式に正確に一致する——ため、項目 008 (省略変数バイアス) から 010 (パラメータ同定) において、推定値とその目標値との差がノイズに起因することはあり得ない。それぞれがバイアスであるか、あるいは「パラメータ同定」の場合、同定の失敗である。

この例では、従来型の評価や第 1 号と同様に、状態をドル単価を持つ特徴量として扱っており、項目 008 (省略変数バイアス)、009 (内生性)、および 010 (パラメータの同定) は、このア

プローチのどこに問題があるかを示しています。従来の評価手法は、本誌が採用している RCA (制約アップ) に基づく評価手法とは異なります。制約アップでは、回帰分析には測定可能な特徴量のみが用いられ、「状態」や「品質」は意図的に除外されています。回帰分析によって説明されなかった売買価格の部分が、残差となります。編集者が担当するカリフォルニア市場では、測定可能な特徴量によって売買価格の変動の約 80% が説明され、残りの約 5 分の 1 が残差となります。鑑定評価者は、ある制約の下で、自身の評価スコアを指針として、残差を「状態」「品質」、およびデータで測定されていないその他の特徴に配分する。その制約とは、残差構成要素の価値寄与の合計が、比較事例の残差と等しくなければならない、というものである。評価額の結論は、鑑定評価者がそれをどのように配分するかには依存しない——RCA 論文 (Craytor 2026, p. 36) の証明 II: すべての残差構成要素の価値寄与の合計が比較事例の残差と等しくなるという制約の下で、残差の内訳を変更しても、最終的な評価額の結論には影響を与えない。したがって、状態の評価率は決して推定されず、主張されることもない。鑑定評価者が「状態」に割り当てるのは、測定された特徴（その中には築年数も含まれる）がすでに説明している範囲を超える部分である。

第 1 号の回帰分析をこの CEF と比較すると、その回帰分析は——ただし、まず 1 点区別しておく必要があるが——次のように判断できる。一般的に言えば、OLS 回帰は対象とする条件付期待関数 (CEF) を再構築するものではなく、その最良の線形近似、すなわち線形射影を求めるものであり、これら二者が一致するのは CEF が線形である場合に限られる。MARS のようなノンパラメトリック手法は、CEF が線形である必要はない。CEF を区間ごとに近似し、曲面が曲がる「ヒンジ」(変化率が変化する点) や、ねじれる「相互作用項」(別の特徴量に依存する変化率) を設けることで、線形射影では平坦化されてしまう非線形の CEF を追跡する。それでもなお、これは同じ対象に対する近似であり、次の項目で取り上げる「省略変数」「内生性」「同定」に関する問題は、これに対してもまったく同じように適用される。前述の 5 変数の CEF は線形であるため、リモデリング指標を含む回帰分析によってそれを正確に再現できるのである。リモデリング指標を除く 4 つの特性で条件付けられた第 1 号の回帰分析は、 $E[P \mid \text{GLA, lot, view, cond}]$ に対する最良の線形近似を推定し、ジャーナルのスコアで条件を再定義すると (第 1 号では 1~3 のコーディングの各ステップごとに切片 80,777 および 49,824 を返した。これは同じ当てはめ結果である)、

$$\hat{p}(\mathbf{z}) = 6,041 + 575 \cdot z_{\text{GLA}} + 22.70 \cdot z_{\text{lot}} + 86,179 \cdot z_{\text{view}} + 24,912 \cdot z_{\text{cond}}.$$

各係数は、その直上の係数とは異なります。回帰分析を直線当てはめという観点から説明してしまえば、これ以上述べるべき点はありません。つまり、2 つのモデルと 2 組の数値があるだけです。しかし、CEF アプローチの下では、2 つの厳密な命題が導き出されます。第一に、2 つの回帰分析は異なる変数群を条件としており、したがって、異なる条件付き期待関数を近似しています:

$$E[P \mid \text{GLA, lot, view, cond}] \quad \text{and} \quad E[P \mid \text{GLA, lot, view, cond, remodel}].$$

一部のリフォーム済み不動産と未リフォーム不動産が同じ 4 つの特性を共有する市場において、これら 2 つの関数はそれらの点において異なる値をとる。この小さな人工市場では、4 つの特性が完全に一致する不動産は 2 つとないため、各不動産において 2 つの関数は一致するが、その形は異なる。すなわち、2 つ目の関数はその特性において直線式であるのに対し、1 つ目の関数はそうではない。いずれの場合も、回帰分析は対象関数に対する最良の線形近似を提供するに過ぎない。両方の対象関数は有効な対象である。第二に、5 つの変数を持つ関数のみが、潜在価格として読み取れる係数を持っている——リフォーム係数は、GLA、敷地面積、眺望、状態スコアが同じリフォーム済み住宅と未リフォーム住宅の価格差である——。その理由は、項目 008 (省略変数バイアス) で説明されている。

第 1 号の Comp H の残差は、この区別がいかに重要であることを如実に示している。第 1 号の 8 つの特性を市場全体を代表するものとして捉えるならば、各 4 変数の特性ベクトルは 8 つの特性のうちの 1 つと完全に一致することになり、したがって 4 変数の CEF は 8 つの価格すべて——Comp H もその一つ——を正確に通過することになる。Comp H は 4 変数 CEF からの偏差がゼロである。その第 1 号の残差 \$46,455 は、4 変数の CEF からゼロの偏差を示しており、これは 4 変数の CEF からゼロの偏差を示しており、その 8 つの不動産からなる市場の線形投影である 4 変数の線形回帰モデルとの差異を表している。そして、まさにその投影と CEF との間のギャップの中に、リモデルの効果が隠されているのである。5 変数の仕様では、予測値と CEF は同一となり、偏差は消失する。この残差は Comp H 自体に起因するものではなく、仕様に起因するものであった。

References.

- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Craytor, W. B. (2026). Residual Constraint Approach (RCA): Framework & Protocol. *The Valuation Engineer*, 1(1). <https://doi.org/10.67330/kgebx46>
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Julius Springer.
- Milborrow, S. (2026a). *earth: Multivariate Adaptive Regression Splines* (R package version 5.3.6; derived from mda:mars by T. Hastie and R. Tibshirani). <https://doi.org/10.32614/CRAN.package.earth>
- Milborrow, S. (2026b, August 7). *Variance Models in earth*. <https://www.milbo.org/doc/earth-varmod.pdf>

Cross-references: hedonic regression; implicit price; omitted variable bias; parameter identification.

Addendum: Prolog Formalization

The following machine-readable companion (007-conditional-expectation.pl) formalizes this entry in the journal's Prolog knowledge base. It is also published as a browsable galley alongside this article.

```
% SPDX-License-Identifier: AGPL-3.0-or-later
% Copyright (C) 2026 William Bert Craytor. Published by The Valuation Engineer
% Journal (Pacific Vista Net). This program is free software: you can
% redistribute it and/or modify it under the terms of the GNU Affero General
% Public License as published by the Free Software Foundation, either version 3
% of the License, or (at your option) any later version. It is distributed
% WITHOUT ANY WARRANTY, without even the implied warranty of MERCHANTABILITY or
% FITNESS FOR A PARTICULAR PURPOSE; see the COPYING file beside it or
% <https://www.gnu.org/licenses/agpl-3.0.html>.

%% 007-conditional-expectation.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 2, Entry 007
%% Title: Conditional Expectation
%% Author: Bert Craytor
%% Concept: conditional_expectation
%% Version: 1.2.0
%%
%% v1.2.0: entry text revised for readers who are not statisticians;
%% no Prolog clause changes.
```

```

%%
%% v1.1.0: condition on the Journal's 0.00-10.00 score (average 5.00, good
%% 7.00, excellent 9.00; 20,000 a point). The generating function lives in
%% kb/pacifica_comps_extended.pl, which is restated in those units with this
%% release; nothing here changes but the units it reads.
%%
%% This file is the Prolog galley for the entry: it encodes the entry's
%% main claims, its worked example, and the generating function of the
%% running dataset in a form other Prolog code in the journal can
%% consult, query, and verify. Read alongside entry.tex / entry.pdf.

% =====
% DICTIONARY IMPORT
% =====

% The dictionary (Issues/tools/dictionary) is not loaded at run time, so
% that the file runs unchanged in Trealla, SWI-Prolog and Scyer. The
% declarations below are plain facts, read by the dictionary linter.
dictionary_release_pinned('dictionary-2026.1').

% --- From the dictionary (quoted for reference) -----
% term(conditional_expectation, "Conditional Expectation",
%     econometric_method,
%     "The conditional expectation of an outcome given a set of
%     predictors is the mean of the outcome's distribution at each
%     fixed value of the predictors. Viewed as a function of the
%     predictors, it is called the conditional expectation function
%     (CEF), and it is the function that regression aims to
%     approximate. Ordinary least squares (OLS) regression gives the
%     best linear approximation to the CEF, and it reproduces the CEF
%     exactly when the CEF itself is linear.", published,
%     'dictionary-2026.1').
%
% depends_on(conditional_expectation, hedonic_regression).
% -----

uses_terms([ conditional_expectation, hedonic_regression,
             implicit_price, hedonic_price_function,
             characteristics_space, residual,
             ordinary_least_squares,
             sales_comparison_approach, adjustment,
             comp_selection, reconciliation ]).
uses_relations([depends_on/2, related/2]).

% =====
% ENTRY SUBJECT
% =====

article_concept(conditional_expectation).
article_issue(volume(1), number(2), year(2026)).
article_id('foundations.2026.007').

% =====
% ENTRY-LOCAL VOCABULARY
% =====
%
% Field names for the worked-example records and for the two properties
% of the CEF the entry relies on. Too narrow to promote to the
% dictionary.

```

```

local_term(cef).                % conditional expectation function
local_term(estimand).
local_term(conditioning_set).
local_term(prediction_interval).
local_term(confidence_interval).
local_term(linear_projection).   % best linear approximation to the CEF

% =====
% THE TWO PROPERTIES OF THE CEF THE ENTRY USES
% =====

%! cef_property(?Name, ?Statement) is nondet.

cef_property(best_predictor,
    "Among all functions of the conditioning variables, the CEF minimizes expected squared
    ↪ prediction error.").
cef_property(orthogonal_deviation,
    "P = E[P|Z] + eps with E[eps|Z] = 0 holds by construction, not by assumption.").

% The second property is the hinge for entries 008-010: each asks when
% the recovered object is the CEF of interest.
hinge_for(orthogonal_deviation, omitted_variable_bias).
hinge_for(orthogonal_deviation, endogeneity).
hinge_for(orthogonal_deviation, identification).

% =====
% CENTRAL CLAIM: CONDITIONING SET DETERMINES THE ESTIMAND
% =====
%
% Two regressions over the same data with different conditioning sets
% target DIFFERENT population objects, and each recovers only the best
% linear approximation to its target (the linear projection); OLS
% recovers the CEF itself exactly when the CEF is linear. Neither is
% wrong; they answer different questions.

conditioning_set(issue1_model,
    [gla_sqft, lot_sqft, view_numeric, cond_score]).
conditioning_set(issue2_model,
    [gla_sqft, lot_sqft, view_numeric, cond_score, remodel_numeric]).

%! different_estimand(?ModelA, ?ModelB) is nondet.
%
% True when two models condition on different variable sets and
% therefore approximate different conditional expectation functions.

different_estimand(A, B) :-
    conditioning_set(A, SA),
    conditioning_set(B, SB),
    A \== B,
    sort(SA, X), sort(SB, Y),
    X \== Y.

% What OLS recovers, relative to the CEF it conditions on.
ols_recovers(best_linear_approximation_to_cef).
ols_recovers_cef_exactly_when(cef_is_linear).

% =====
% WORKED EXAMPLE — THE GENERATING FUNCTION

```

```

% =====
%
% The running dataset is defined in ../../kb/pacifica_comps_extended.pl.
% Its generating function IS the conditional expectation function for
% the constructed market, exactly: the data carry no noise.

:- initialization(consult('../../kb/pacifica_comps_extended')).
% Scyer Prolog loads forall/2 and the list predicates on request;
% SWI-Prolog and Trealla have them built in. The guarded loads below
% let the file run unchanged in all three.
:- initialization(catch(use_module(library(lists)), _, true)).
:- initialization(catch(use_module(library(iso_ext)), _, true)).

%! cef_exact(+Property, -Price) is det.
%
% The CEF evaluated at a property's characteristic vector. Identical to
% true_price/2 in the dataset file; named here to make the entry's claim
% ("the generating function IS the CEF") explicit in the galley.

cef_exact(P, Price) :- true_price(P, Price).

%! noiseless_dataset is semidet.
%
% Verifies the entry's central factual claim about the data: every
% observed sale price equals the CEF at that property, so the residual
% is zero everywhere.

noiseless_dataset :-
    forall(property(P),
        ( cef_exact(P, Price), sale_price(P, Price) )).

% =====
% THE COMP H POINT
% =====
%
% Issue 1 reported a $46,455 residual for Comp H under the four-variable
% model. Taking issue 1's eight properties as the market, every
% four-variable characteristic vector belongs to exactly one property, so
% the four-variable CEF interpolates all eight prices and Comp H deviates
% from it by ZERO. The 46,455 is Comp H's deviation from the fitted
% four-variable LINEAR regression --- the linear projection --- and the
% projection-CEF gap is where the remodel effect hides. Under the
% five-variable specification, projection and CEF coincide and the
% deviation vanishes.

issue1_residual(comp_h, 46455).

% deviation_from(Property, Object, Dollars).
deviation_from(comp_h, four_variable_projection, 46455).
deviation_from(comp_h, four_variable_cef, 0).
deviation_from(comp_h, five_variable_cef, 0).

residual_under(issue2_model, comp_h, 0) :-
    cef_exact(comp_h, Price),
    sale_price(comp_h, Price).

%! residual_is_artifact_of_specification(?Property) is nondet.
%
% The issue-1 residual is a property of the specification (a linear

```

```

% projection under a smaller conditioning set), not of the property.

residual_is_artifact_of_specification(P) :-
    issue1_residual(P, R), R \= 0,
    deviation_from(P, four_variable_cef, 0),
    residual_under(issue2_model, P, 0).

% =====
% APPRAISER ENCOUNTERS
% =====

grid_assumption(smoothness,
    "Using comparables at all assumes the CEF is smooth enough that nearby points inform the
    ↪ subject's point.").
grid_assumption(local_linearity,
    "Adjusting linearly assumes the CEF's slope in that characteristic is roughly constant over the
    ↪ range crossed.").
grid_assumption(unbiased_indications,
    "Averaging indications assumes each is an unbiased draw around the same conditional mean.").

% The grid is structurally a nonparametric estimator of a conditional
% mean, using local averaging and hand-specified slopes.
estimator_role(sales_comparison_approach, conditional_mean_estimator).

% =====
% DEFENSIBILITY
% =====

defensibility_consequence(conditioning_set,
    "Adding a variable to a grid changes the estimand, not merely the precision of an estimate.").
defensibility_consequence(ceteris_paribus,
    "'Holding other things equal' means comparing conditional means at vectors differing in one
    ↪ coordinate.").
defensibility_consequence(residual_spread,
    "Conditional dispersion is a market property; a point estimate reported without it overstates
    ↪ precision.").

% The confidence/prediction interval confusion the entry names.
interval_answers(confidence_interval, "uncertainty about where the CEF is").
interval_answers(prediction_interval, "spread of individual prices around the CEF").
commonly_confused(confidence_interval, prediction_interval).

```