東京区部新築戸建住宅における一般化加法モデル(GAM)による ノンパラメトリック・ヘドニック価格分析

A Study of Nonparametric Hedonic Regression Analysis for Tokyo Newly Built Detached House Using a Generalized Additive Model

MEIKAI Univ.　KIM Donghwan　　明海大学　不動産学部　金東煥

本研究は、東京都区部の新築戸建住宅約5万件の成約データを用い、線形OLSモデルと一般化加法モデル（GAM）によるヘドニック価格分析を比較した。一般化加法モデル（GAM）では土地面積、建物面積、最寄駅距離などの主要説明変数を平滑化関数として推定し、住宅価格形成における非線形関係を検証した。その結果、一般化加法モデル（GAM）は線形 OLS モデルよりわずかに高い予測精度を示した。また、建物面積や駅距離では顕著な非線形性が確認され、戸建住宅の価格分析ではノンパラメトリック手法を導入することが予測精度の向上と戸建住宅価格形成メカニズムの理解に有効であることが明らかとなった。

1. はじめに

東京都区部の新築戸建住宅市場では、住宅価格が土地面積、建物面積、最寄駅までの距離、用途地域など多様な住宅属性や立地条件によって形成されている。このような住宅価格形成要因を分析する代表的な手法として、ヘドニック価格モデルが広く用いられている。従来研究では、対数線形の最小二乗法（線形 OLS モデル）によるパラメトリック回帰が主流であり、各説明変数の価格弾力性を容易に解釈できるという利点を有している。一方で、住宅価格と住宅属性との関係は必ずしも線形ではなく、土地面積や建物面積、最寄駅までの徒歩距離などでは住宅価格との非線形関係が存在することが指摘されている（Yatchew, 1998）。このような場合、関数形を事前に仮定する線形モデルでは、価格形成の実態を十分に表現できず、推定精度や予測性能が低下する可能性がある。

近年では、この問題を乗り越えるために、ノンパラメトリック回帰や半パラメトリック回帰が住宅価格分析へ応用されており、一般化加法モデル（Generalized Additive Model：GAM）は、主要説明変数の非線形関係を柔軟に推定できる手法として注目されている。一般化加法モデル（GAM）は、説明変数ごとの平滑化関数を用いることで、線形モデルでは表現できない複雑な価格形成を捉えることが可能であり、近年の住宅価格分析や不動産評価への適用が進んでいる。

本研究の目的は、東京都区部の新築戸建住宅を対象として、従来の線形 OLS モデルによるパラメトリックのヘドニック価格分析と、ノンパラメトリック手法を導入した半パラメトリック一般化加法モデル（GAM）によるヘドニック価格分析を比較し、戸建住宅の価格形成における非線形関係を明らかにするとともに、より高い予測精度を有するヘドニック価格モデルを構築することである。また、各住宅属性が戸建住宅価格に及ぼす影響を、線形モデルと非線形モデルの双方から比較・検討し、東京都区部における戸建住宅の価格形成の特徴を明らかにする。本研究では、以上の研究目的に対して、R言語のmgcv::gamパッケージ（Wood, 2017）を用いた REML 法による半パラメトリック回帰推定を行い、実測値に基づいてモデルの性能を評価する。

本論文の構成は以下のとおりである。第 2 章ではヘドニック価格モデルおよびノンパラメトリック回帰に関する先行研究を整理する。第 3 章では分析モデルおよび使用データについて説明する。第 4 章では一般化加法モデル（GAM）と線形OLSモデルの推定結果を比較し、モデル適合度、平滑化項の非線形関係および主要説明変数の影響について考察する。最後に、第 5 章で本研究の結論と今後の課題を述べる。

2. 先行研究

ヘドニック価格モデル

ヘドニック価格法（HPM）の起源は Court（1939）による自動車価格指数の研究に遡る。

Lancaster（1966）は財を属性の束として捉える需要理論を提唱し、財から得られる効用が財そのものではなく、それが持つ属性から生じる「ヘドニック分析の需要側理論的基礎」として定式化した。Rosen（1974）が競争均衡モデルとして理論を体系化した。Rosen（1974）において住宅価格は属性の暗黙価格関数として表現される。

$$P(z) = P(z_1, z_2, \ldots, z_k) \quad (1)$$

この暗黙価格（インプリシット・プライス）$\partial P/\partial z_k$は、属性z_kの 1 単位増加に対して消費者が支払う意思のある価格の増分（Willingness to Pay）と生産者の供給余剰が均衡した市場価格として解釈される。
住宅価格分析では（1）式の対数線形モデルが一般的に採用されている（Sirmans et al., 2005）。

$$\ln P = \beta_0 + \Sigma_j \beta_j \ln z_j + \Sigma_k \gamma_k d_k + \varepsilon \quad (2)$$

（2）式において、lnP は住宅価格の対数、z_jは連続属性変数の対数、d_kはダミー変数（立地・時点等）、β_j は弾力性を表す。この定式化では、係数が弾力性として解釈できるとともに、価格分布の歪度を低減できる利点がある。日本では清水（2008）が GIS データを用いた精密なヘドニック分析を行い、前面道路幅員・最寄駅距離・容積率を精密に再測定し、推定精度を向上させた。

金（2026）は、戸建住宅データを用いて線形ヘドニック回帰を実施し、区ダミーと年度ダミーを含む拡張モデルを構築し、高い水準の自由度修正決定係数（R^2=0.889）を達成した。金（2026）は、本研究のベースラインモデルとなる。

ノンパラメトリック手法の発展

パラメトリックモデルは、説明変数と目的変数の関係を線形、対数線形などの仮定の基で推定するため、解釈が容易で推定効率も高いという利点がある。一方、パラメトリックモデルは、説明変数と目的変数の関係を線形などで事前に仮定する必要があり、非線形関係の価格形成を十分に表現できない。また、複雑な変数間の関係を捉えにくく、関数形の誤指定により推定精度が低下する可能性があるという限界も有する。このようなパラメトリックモデルの限界を克服するため、ノンパラメトリック・半パラメトリック手法が発展してきた。Mason & Quigley（1996）はノンパラメトリック回帰を住宅データに適用し、住宅価格形成には線形 OLS モデルでは十分に表現できない非線形性が存在することを示した。Anglin & Gençay（1996）は半パラメトリックモデルがパラメトリックモデルを予測精度で上回ることを示した。
一般化加法モデル（GAM）への応用では Pace（1998）が不動産評価への先駆的適用を行い、Bin（2004）が半パラメトリック対パラメトリックの予測比較を実施した。Wood（2017）は REML 法による平滑化パラメータ自動選択など、一般化加法モデル（GAM）の現代的実装を確立した。Bailey etal.（2022）は P-スプライン一般化加法モデル（GAM）が線形モデルより高い予測精度を持ち、属性の非線形関係を捕捉できることを示した。

ヘドニック価格モデルは、不動産市場を分析対象にするとき、各変数の線形を仮定して、線形のヘドニック価格モデルを用いて構築することが一般的である。本研究は、戸建住宅データを用いて、各属性データの分布を線形と決めずに、データそのものの関係性を推定するノンパラメトリック推定（Nonparametric Estimation）を行うことで、戸建住宅市場の分析に多様化を図ることが有意義である。

3. モデルとデータ

3.1 ヘドニック価格モデル

ヘドニック価格モデルとしては、パラメトリック回帰、ノンパラメトリック回帰、半パラメトリック回帰が使われている。

パラメトリック回帰（線形 OLS モデル）は、説明変数と目的変数の関係を、係数を持つ明示的な関数形で表す手法である。パラメトリック回帰の典型的な例として線形回帰があり、次式で表される。

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

線形 OLS モデル（最小二乗法）は、$\Sigma(y_i - \hat{y}_i)^2$ を最小にするように係数を推定する方法で、解釈しやすく基準的な回帰モデルとして広く用いられる。ただし、β_k は各変数の推定係数、x_k は各変数、y_iは実測値、$\hat{y}_i$は予測値を示す。

ノンパラメトリック回帰は、事前に厳密な関数形を定めず、データから柔軟に回帰曲線を推定する手法である。一般形は $y = m(x) + \varepsilon$ で、未知の関数 m(x)をスプライン法などで推定する。ノンパラメトリック回帰は非線形関

係を捉えやすい一方で、結果の解釈がやや難しく、十分なデータ量が必要になることがある。

半パラメトリック回帰は、一部を線形部分のパラメトリック、残りを非線形部分のノンパラメトリックに扱う複合型の手法である。代表的な形は $y = x\beta + g(z) + \varepsilon$ で、$x\beta$ は線形部分、$g(z)$は未知の非線形関数である。解釈性と柔軟性の両立を図れるため、実務分析でも有用である。

表 1 は、パラメトリック回帰（線形 OLS モデル）、ノンパラメトリック回帰、半パラメトリック回帰についての特徴を示す。本研究では、戸建住宅のデータに基づいて、ヘドニック価格モデルのパラメトリック回帰とノンパラメトリック手法を導入した回帰モデルを推定する。なお、本研究で推定する一般化加法モデル（GAM）は、平滑化項と線形項を組み合わせた半パラメトリックモデルである。

3.2　データの概要と基本統計量

分析データは、業プロ株式会社が提供する東京都区部新築戸建住宅の成約データである（表 2）。対象期間は 2009 年から 2026 年 1 月までとし、原データ 54,007 件のうち、目的変数および説明変数に欠損値を含む観測値を除外した結果、49,254 件を分析対象とした。

推定モデルの構築および予測性能の検証を行うため、分析対象データを訓練データ（Training Data）とテストデータ（Test Data）に分割した。具体的には、データ全体の 80%（39,403 件）を訓練データとしてモデルの推定に使用し、残り 20%（9,851 件）をテストデータとしてモデルの予測性能を評価した。データの分割はランダムサンプリングにより行い、推定に使用していないテストデータに対する予測結果を用いて、ヘドニック価格モデルの汎化性能を検証した。このようにデータを分割して評価することで、同一データによる推定・評価に起因する過学習（Overfitting）の影響を抑え、未知データに対する予測精度を客観的に評価することが可能となる。

表 3 に分析対象データの基礎統計量を示す。成約価格の平均は約 5,206 万円（中央値 4,880 万円）であり、建物面積の平均は 94.9 ㎡、最寄駅までの徒歩距離は平均 13.4 分（中央値 12 分）であった。また、価格、土地面積、建物面積、建蔽率、容積率など右裾の長い分布を示す変数については、分布の歪みを緩和し、線形関係を改善するため自然対数変換を与えた。建蔽率および容積率についても対数変換後の値を説明変数としてモデルに投入した。

4. モデル推定と結果

4.1　一般化加法モデル（GAM）モデルと線形 OLS モデル

本研究では、一般化加法モデル（GAM）の平滑化関数として薄板回帰スプライン（Thin Plate Regression Spline）を採用した。薄板回帰スプライン [1)、2)] は、説明変数と目的変数の非線形関係を滑らかな関数として推定する手法であり、本研究で用いた R 言語の mgcv パッケージでは、節点（ノット）の位置を研究者が事前に指定する必要がなく、基底関数、

表 1　ヘドニック価格モデルの比較

アプローチ	関数形の仮定	解釈可能性
パラメトリック回帰（線形 OLS モデル）	強い（対数線形）	高い
半パラメトリック回帰（一般化加法モデル：GAM）	主要説明変数の仮定なし	高い
完全ノンパラメトリック回帰（カーネル回帰）	なし	低い

表 2　データ概要

項目	内容
データ提供	業プロ株式会社
対象	東京都区部新築戸建住宅 成約データ
対象期間	2009 年～2026 年 1 月
原データサンプル	54,007 件
分析サンプル	49,254 件（欠損除去後）
訓練データ	39,403 件（80%）
テストデータ	9,851 件（20%）
目的変数	ln_price = ln（成約価格）（注）ln は変数の対数変換を示す。

表 3　変数の基礎統計量（原データ、2009～2026 年）

変数	サンプル数	平均	標準偏差	最小値	中央値	最大値
成約価格（円）	54,005	52,063,300	18,808,700	10,000	48,800,000	450,000,000
単価（円/㎡）	54,002	675,111	273,081	1,000	621,436	4,295,447
土地面積（㎡）	54,003	82.07	49.23	10	81.70	8,579
建物面積（㎡）	54,000	94.87	14.50	10	94.72	322.28
建物階数（階）	53,999	2.43	0.50	1	2	4
最寄駅までの徒歩距離（分）	51,560	13.36	6.56	2	12	63
建蔽率	53,744	0.591	0.097	0.25	0.60	1.00
容積率	53,704	1.696	0.602	0.30	1.60	8.00
前面道路幅員（m）	53,280	5.28	2.37	1.58	4.50	52.00
部屋数（室）	53,582	3.01	0.90	0	3	6

平滑化パラメータ、曲線の複雑さを「制限付き最尤法（Restricted Maximum Likelihood）」などを用いて最適化する。また、平滑化ペナルティにより過度な曲線の複雑化を抑制するとともに、データの構造に応じた柔軟な推定を可能とするため、ヘドニック価格モデルを用いる不動産価格分析において広く利用されている。

本研究の目的変数は成約価格の対数（y=ln_price）とする。説明変数のうち、土地面積、建物面積、最寄駅までの徒歩距離、建蔽率、容積率、前面道路幅員の 6 変数については薄板回帰スプラインの非線形関係を設定し、その他の変数については線形項として扱った（本研究では半パラメトリック一般化加法モデル（GAM）を R の mgcv::gam（REML 法）で推定）。（3）式は半パラメトリック一般化加法モデル（GAM）を示す。

$$y=f_1(\ln_land)+f_2(\ln_building)+f_3(\ln_station)+f_4(\ln_bcr)+f_5(\ln_far)+f_6(\ln_road_width)+\gamma_1 \ln_floors+\gamma_2\, rooms+\Sigma_k\, \delta_k\, ward_k+\Sigma_t\, \varphi_t\, year_t+\Sigma_m\, \psi_m\, dummy_m+\varepsilon \quad (3)$$

ただし、ln_land は土地面積、ln_building は建物面積、ln_station は最寄駅までの徒歩距離、ln_bcr は建蔽率、ln_far は容積率、ln_road_width は前面道路幅員、ln_floors は建物階数、rooms は部屋数、$ward_k$は区ダミー（中央区基準）、$year_t$は成約年ダミー（2009 年基準）、$dummy_m$は、用途地域、接道条件、構造等の属性を表すダミーである。ln は変数の対数変換を示す。f_j は薄板スプライン（非線形関係）による平滑化関数である。

（4）式は、一般化加法モデル（GAM）が「どのような曲線を選ぶか」を表す薄板スプライン（非線形関係）を選択する際に、使われる式であり、平滑化パラメータ λ_j が「制限付き最尤法（Restricted Maximum Likelihood）」で R の mgcv::gam（REML 法）では自動選択される[3)]。

$$\min_{\{\beta,f\}} \Sigma_i(y_i-\hat{y}_i)^2+\Sigma_j\, \lambda_j \int[f_j''(x)]^2 dx \quad (4)$$

（5）式は、全変数を線形項として投入した線形ヘドニック価格モデルであり、通常の線形 OLS モデルである。本研究では（4）式の一般化加法モデル（GAM）と比較する。

$$y=\beta_0+\beta_1 \ln_land+\beta_2 \ln_building+\beta_3 \ln_floors+\beta_4 \ln_station+\beta_5 \ln_bcr+\beta_6 \ln_far+\beta_7\, rooms+\beta_8 \ln_road_width+\Sigma_k\, \gamma_k\, dummy_k+\varepsilon \quad (5)$$

4.2　モデル推定結果とモデル適合度の比較

表 4 は、一般化加法モデル（GAM）と線形 OLS モデルのテストデータでの予測精度を示す。一般化加法モデル（GAM）はテスト R^2が 0.8558、線形 OLS モデルのテスト R^2は 0.8416 であり、両モデルの決定係数がほぼ同じ程度を示すが、わずかに一般化加法モデル（GAM）が高い（推定データでの一般化加法モデル（GAM）の自由度修正 R^2は 0.865、線形 OLS モデルの自由度修正 R^2は 0.851）。これは、一般化加法モデル（GAM）と線形 OLS モデルが戸建住宅価格の変動をほぼ同じ程度で説明するが、一般化加法モデル（GAM）がわずかに説明力が高く、未知データに対しても優れた予測性能を有すると考えられる。また、平均

絶対パーセント誤差（MAPE）は、線形 OLS モデルの 9.408％に対して一般化加法モデル（GAM）では 8.988％となり、一般化加法モデル（GAM）がわずかに優れる。MAPE は予測値と実測値との平均的な誤差率を示す指標であり、値が小さいほど予測精度が高いことを意味する。したがって、一般化加法モデル（GAM）は線形 OLS モデルよりも実際の成約価格に近い予測を行っていることが確認された。さらに、本研究では、一般化加法モデル（GAM）の逸脱度説明率（Deviance Explained）[4] は 86.6％であり、戸建住宅価格形成に関する情報の大部分をモデルが捉えていることが確認された。一般化加法モデル（GAM）の逸脱度説明率は、自由度修正 R^2（0.865）ともほぼ一致しており、本モデルが高い適合度を有していると評価できる。

以上の結果から、一般化加法モデル（GAM）は主要説明変数の非線形関係の価格形成を柔軟に表現できるため、線形 OLS モデルと比較して予測性能およびモデル適合度の双方で優れていることが明らかとなった。特に、建物面積、土地面積、駅距離などの説明変数について、一定の線形効果を仮定する線形 OLS モデルでは捉えにくい非線形性を一般化加法モデル（GAM）が適切に表現したことが、予測精度の向上につながったものと考えられる。

MAPE も一般化加法モデル（GAM）（8.988%）の方が線形 OLS モデル（9.408%）より低く、予測精度の改善が確認された。

4.3　平滑化項各変数の有効自由度と有意性

平滑化項の有効自由度（Effective Degrees of Freedom：EDoF）は、一般化加法モデル（GAM）における平滑化関数の複雑さを表す指標である。EDoF が 1 に近い場合は線形関係に近く、値が大きくなるほど非線形性が強く、より複雑な曲線が推定されていることを示す。したがって、EDoF を確認することで、各説明変数に対して線形モデルで十分か、それとも一般化加法モデル（GAM）による非線形モデルが必要かを判断することができる。表 5 は、一般化加法モデル（GAM）の平滑化項の有効自由度の F 検定結果を示す。6 つの変数の平滑化項が $p<0.001$ で有意であり、すべての平滑化対象変数において有意な非線形関係が確認された。建物面積（EDoF=6.897）・容積率（6.807）・最寄駅までの徒歩距離（6.222）で EDoF が高く、強い非線形性を示している。

4.4　一般化加法モデル（GAM）の平滑化項各変数の解釈

表 6 は、一般化加法モデル（GAM）の平滑化項の平均傾きを示す。平均傾き（Average Slope）は、一般化加法モデル（GAM）で推定した平滑化関数の観測範囲全体における平均的な限界効果を示し、値が正（負）の場合は、説明変数の増加に伴い住宅価格が平均的に上昇（低下）することを意味する[5]。

表 4　一般化加法モデル（GAM）と線形 OLS モデルの予測精度比較（テストデータ）

指標	一般化加法モデル（GAM）	線形 OLS モデル
テスト R^2	0.8558	0.8416
テスト RMSE	0.1160	0.1216
テスト MAE	0.0895	0.0936
テスト MAPE（%）	8.988	9.408
逸脱度説明率（%）	86.6	—

表 5　一般化加法モデル（GAM）平滑化項の有効自由度の検定

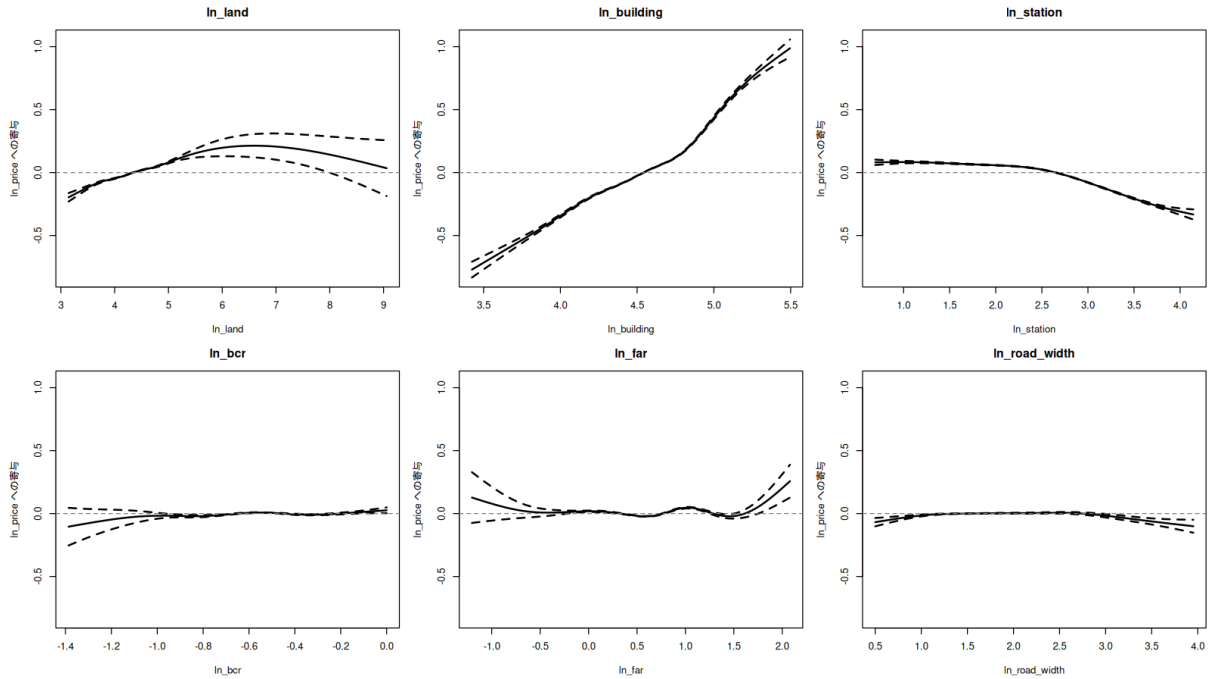
モデル変数	変数	EDoF	F 値	p 値
s(ln_land)	土地面積	6.285	126.09	<0.001***
s(ln_building)	建物面積	6.897	1827.77	<0.001***
s(ln_station)	最寄駅までの徒歩距離	6.222	1631.77	<0.001***
s(ln_bcr)	建蔽率	5.374	13.33	<0.001***
s(ln_far)	容積率	6.807	75.98	<0.001***
s(ln_road_width)	前面道路幅員	5.578	8.58	<0.001***

（注）ln は変数の対数変換を示す。

表 6　一般化加法モデル（GAM）平滑化項の平均傾き（Average Slope）の推定結果

平滑化項	平均傾き	影響方向
s(ln_land)	0.039	正（＋）
s(ln_building)	0.849	正（＋）
s(ln_station)	-0.12	負（－）
s(ln_bcr)	0.094	正（＋）
s(ln_far)	0.04	正（＋）
s(ln_road_width)	-0.009	負（－）

（注）ln は変数の対数変換を示す。



（注）点線は 95%信頼区間を示し、ln は変数の対数変換を示す。

図 1　一般化加法モデル（GAM）の平滑化項の非線形関係

各変数の平滑化項の平均傾きは、その解釈に一定の留意が必要であるが、以下の解釈ができる。土地面積の増加は住宅価格を上昇させるが、その平均的な効果は比較的小さい。建物面積の増加は住宅価格を最も大きく押し上げる。駅から遠くなるほど住宅価格は低下する。平均的には建蔽率の上昇は住宅価格を押し上げる。容積率の上昇は住宅価格に正の影響を与える。前面道路幅員の平均的な影響は小さく、住宅価格への寄与は限定的である。

図 1 は、一般化加法モデル（GAM）の平滑化項のグラフを示す。建物面積は戸建住宅価格に対して最も大きな正の非線形関係を示し、建物規模が大きくなるほど価格上昇効果が強まる傾向が確認される。一方、土地面積は一定規模までは戸建住宅価格を押し上げるが、その後は限界効果が逓減する傾向が確認される。最寄駅までの徒歩距離は、駅から遠くなるほど戸建住宅価格が低下する明確な非線形関係を示し、交通利便性が住宅価格形成に大きく影響していることが確認される。また、建蔽率および容積率では緩やかな非線形性が認められるが、その影響は建物面積や駅距離と比較して小さかった。前面道路幅員についても住宅価格への影響は限定的である。

以上より、戸建住宅価格形成においては建物面積と最寄駅までの徒歩距離が主要な非線形関係を示す要因であり、土地面積にも限界効果の逓減が確認される。一方、建蔽率、容積率および前面道路幅員については影響は比較的小さいが、線形 OLS モデルでは表現できない非線形性が存在することが明らかである。

表 7　一般化加法モデル（GAM）線形項各変数の主要係数

モデル変数	変数	係数	標準誤差	t 値	p 値
ln_floors	建物階数	-0.107	0.0055	-19.53	<0.001***
rooms	部屋数	0.005	0.0008	7.03	<0.001***
minato_d	港区	0.332	0.1207	2.76	0.006**
shinagawa_d	品川区	-0.256	0.1154	-2.22	0.026*
setagaya_d	世田谷区	-0.342	0.1151	-2.97	0.003**
suginami_d	杉並区	-0.399	0.1151	-3.47	<0.001***
nakano_d	中野区	-0.478	0.1152	-4.15	<0.001***
kita_d	北区	-0.606	0.1152	-5.27	<0.001***
edogawa_d	江戸川区	-0.719	0.1151	-6.25	<0.001***
adachi_d	足立区	-0.906	0.1151	-7.87	<0.001***
wooden_d	木造	-0.120	0.0167	-7.18	<0.001***
ownership_d	所有権	0.259	0.0034	75.27	<0.001***

（注）ln は変数の対数変換を示す。

表 8　線形 OLS モデルの説明変数・主要ダミーの係数

モデル変数	変数	係数	標準誤差	t 値	p 値
ln_building	建物面積	0.648	0.0058	111.72	<0.001***
ln_land	土地面積	0.111	0.0043	25.66	<0.001***
ln_floors	建物階数	-0.127	0.0055	-23.17	<0.001***
ln_station	最寄駅までの徒歩距離	-0.124	0.0014	-89.26	<0.001***
ln_bcr	建蔽率	-0.018	0.0074	-2.41	0.0157*
ln_far	容積率	0.019	0.004	4.85	<0.001***
ln_road_width	前面道路幅員	0.004	0.0021	2.05	0.0395*
rooms	部屋数	0.003	0.0008	4.187	<0.001***
wooden_d	木造	-0.1481	0.0174	-8.5	<0.001***
ownership_d	所有権	0.26	0.0036	72.179	<0.001***

（注）ln は変数の対数変換を示す。

4.5　一般化加法モデル（GAM）と線形 OLS モデルの線形項各変数の比較

主要説明変数の比較

表 7 は、一般化加法モデル（GAM）の線形項各変数の推定結果であり、表 8 は線形 OLS モデルによる主要説明変数の推定結果を示す。一般化加法モデル（GAM）では、土地面積、建物面積、最寄駅までの徒歩距離、建蔽率、容積率、前面道路幅員の各変数を平滑化関数として推定し、前節で説明したため、表 7 には線形項のみを掲載する。一方、線形 OLS モデルでは、すべての説明変数を線形項として推定したため、各係数は住宅価格に対する平均的な影響を表す。

まず、線形 OLS モデルの結果から、建物面積（対数）の係数は 0.648 となり、建物面積が 1%増加すると住宅価格は約 0.65%上昇することが示された。これは本研究で最も大きな正の影響を有する変数であり、東京都区部の新築戸建住宅の建物規模が戸建住宅の価格に最も大きく影響を与えることを示す。土地面積（対数）の係数は 0.111 であり、土地面積の増加は戸建住宅価格を押し上げるが、その影響は建物面積より小さい。東京都区部では土地価格が高く、限られた敷地を有効利用した戸建住宅の供給が一般的であるため、土地規模よりも建物規模の方が価格への影響が大きいことを表すと考えられる。最寄駅までの徒歩距離（対数）の係数 [6] は−0.124 となり、駅距離が 1%増加すると住宅価格は約 0.12%低下した。最寄駅への近接性が住宅価格を左右する重要な要因であることが確認され、交通利便性の高さが戸建住宅価格に正の影響を与えていることが示される。ただし、一般化加法モデル（GAM）では、駅距離を平滑化関数として推定しており、駅から遠くなるほど戸建住

宅価格が低下するが、一定距離を超えるとその影響が小さくなるなど、線形 OLS モデルでは表現できない住宅価格との非線形関係が確認される。建蔽率と容積率の係数は、それぞれ−0.018 および 0.019 と小さく、住宅価格への影響は限定的である。また、前面道路幅員の係数は 0.004 と極めて小さく、道路幅員単独の影響は、他の立地条件や住宅属性と比較すると小さいことが示される。

ダミー変数の比較

建物階数は、一般化加法モデル（GAM）で −0.107、線形 OLS モデルで、−0.127 となり、いずれも有意な負の係数を示す。東京都区部では 3 階建て住宅は、狭小敷地に建設されるケースが多く、住環境や居住性の観点から価格単価が相対的に低く評価される傾向があることに起因すると考えられる。部屋数は、一般化加法モデル（GAM）で 0.0053、線形 OLS モデルで 0.003 といずれも正の係数を示したが、その効果は小さい。これは部屋数そのものよりも、建物面積や住宅全体の規模が戸建住宅価格形成に与える影響が大きいことを示唆すると考えられる。所有権ダミーは、一般化加法モデル（GAM）で 0.259、線形 OLS モデルで 0.260 とほぼ同一の係数を示す。所有権住宅は、借地権住宅より高く評価される傾向が示された。これは土地所有権が住宅資産価値に与える影響を反映していると考えられる。木造ダミーは一般化加法モデル（GAM）で −0.120 となり、非木造住宅と比較して木造住宅の価格は約 12%低いことを示す。東京都区部では、耐火性や耐久性に優れる鉄骨造や鉄筋コンクリート造の住宅が相対的に高く評価されるため、このような結果が得られたと考えられる。

区ダミー[7] については、中央区を基準として推定した結果、港区（+0.3327）は正の係数を示す一方、品川区（−0.257）、世田谷区（−0.342）、杉並区（−0.400）、中野区（−0.478）、北区（−0.607）、江戸川区（−0.720）、足立区（−0.906）の順に負の係数の絶対値が大きくなる。これは、東京都区部の価格水準を示すと考えられる。特に本稿で示した区の中では、港区と足立区の推定係数の差が最も大きく、都心へのアクセス性や地域ブランド、周辺環境などの立地要因が住宅価格に大きく反映されている可能性がある。

以上より、線形 OLS モデルは、各説明変数が戸建住宅価格に与える影響を把握する上で有効である。加えて、一般化加法モデル（GAM）は、土地面積、建物面積、最寄駅までの徒歩距離などの主要説明変数と住宅価格との非線形関係を表現できる点に特徴がある。例えば、最寄駅までの徒歩距離は、駅から遠くなれば、住宅価格が低下するが、一定距離を超えるとその影響が小さくなるなど、線形 OLS モデルでは表現できない非線形関係の価格形成を一般化加法モデル（GAM）が考慮できると考えられる。

本研究では、一般化加法モデル（GAM）と線形 OLS モデルの予測精度がほぼ類似であるが、一般化加法モデル（GAM）がわずかに高いことから、東京都区部の新築戸建住宅市場では、主要説明変数の非線形性を考慮することが戸建住宅価格の推定精度向上に有効であることが考えられる。

5. 結論

本研究は、東京都区部新築戸建住宅の成約価格を目的変数として、一般化加法モデル（GAM）と線形 OLS モデルの 2 モデルを推定し、比較した。その結果、以下の 3 点が確認された。

第一に、一般化加法モデル（GAM）（テスト R^2=0.8558）は線形 OLS モデル（テスト R^2=0.8416）とほぼ類似であるが、わずかに上回り、ノンパラメトリック手法を導入した半パラメトリック GAM が、東京都区部の戸建住宅の価格推定モデルとして有効であることが確認された。

第二に、一般化加法モデル（GAM）の主要説明変数の平滑化項（非線形関係）が統計的に有意であり、EDoF も 5.4〜6.9 を示したことから、主要説明変数と戸建住宅価格との間に非線形な関係が存在することが示された。例えば、最寄駅までの徒歩距離と戸建住宅価格は、負の関係を有するが、一定距離を超えるとその影響が小さくなるなどの非線形関係を有する。

第三に、一般化加法モデル（GAM）と線形 OLS モデルは、区ダミーについて統計的に有意な結果が確認され、戸建住宅の立地が価格に与える影響が明確であることを示した。戸建住宅価格には、立地を表す区別の価格差が明確に確認され、東京都区部における住宅価

格の地域差が示された。

一方で、戸建住宅価格推定のヘドニック価格モデルにおいては、一般化加法モデル（GAM）と線形OLSモデルがほぼ同じ予測精度を示し、主要説明変数の価格への影響をおおむね同じ傾向で捉えていることが確認された。つまり、解釈可能性（各係数の直接的解釈）においては、線形 OLS モデルの方が優れるため、実務的な価格予測などでは、線形OLSモデルも引き続き有用である。

したがって、今後の課題として、一般化加法モデル（GAM）と線形OLSモデルをさらに拡張して、予測精度を高める必要がある。その方法としては、一般化加法モデル（GAM）に空間自己相関を考慮した空間 GAM（Geographically Weighted GAM）の導入又は、変数間の交互作用項のモデル化などを検討してより高い予測精度を有するモデルの構築を目指す。

【注】

1）スプライン関数（Spline Function）とは、複数の多項式を滑らかにつなぎ合わせて作る関数であり、複雑な非線形関係を柔軟に表現するための数学的手法である。一般化加法モデル（GAM）の中核となる技術であり、不動産価格分析でも広く利用される。

2）薄板回帰スプライン回帰の詳細については、Simon N. Wood (2003)を参照されたい。

3）R の mgcv::gam（REML 法）の詳細については、Simon Wood（2026）を参照されたい。Simon Wood（2026）、“Package ‘mgcv’”（オンライン：https://cran.uib.no/web/packages/mgcv/mgcv.pdf）

4）逸脱度説明率（Deviance Explained）は、平均値のみを用いたモデル（Null Model）と比較して、推定モデルが目的変数の変動をどの程度説明できたかを示すGAMで報告される指標である。

5）一般化加法モデル（GAM）では各説明変数が非線形関係として推定されるため、平均傾きは全区間における平均的な傾向を表す指標であり、各点における限界効果は平滑化関数の形状に応じて変化する。

6）本研究では、モデル全体の説明力を優先するため、最寄駅までの徒歩距離を含む主要説明変数に対して、対数変換を与えて分析した。

7）区ダミーの推定結果は、紙面制約のため、東京都区部の特徴を示す一般加法モデル（GAM）による一部の区の推定結果のみを結果として示す。また、線形 OLS モデルの区ダミー推定結果は、概ね一般加法モデル（GAM）の推計結果と類似な値を示した。

【参考文献】

Court, A. T. (1939), " Hedonic Price Indexes. The Dynamics of Automobile Demand", New York: General Motors, pp.98–119.

Griliches, Z. (1961), "Hedonic Price Indexes for Automobiles: An Econometric of Quality Change", NBER Chapters, in: The Price Statistics of the Federal Goverment, National Bureau of Economic Research, Inc, pp.173-196.

Lancaster, K. J. (1966), "A New Approach to Consumer Theory", Journal of Political Economy, 74(2), pp.132-157.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", Journal of Political Economy, 82(1), pp.34-55.

Sirmans, G. S., Macpherson, D. A. & Zietz, E. N. (2005), "The Composition of Hedonic Pricing Models", Journal of Real Estate Literature, 13(1), pp.1-44.

清水千弘（2008）,「不動産市場の情報不完全性と価格形成要因に関する研究」東京大学学位（博士）請求論文.

金東煥（2026）,「時点ダミー型ヘドニック法による東京 23 区戸建住宅価格指数の構築と検証」,明海大学不動産学部論集 Vol.38, pp.15–29。

Mason, C. & Quigley, J. M. (1996), "Non-parametric Hedonic Housing Prices" , Housing Studies, 11(3), pp.373-385.

Anglin, P. & Gençay, R. (1996), "Semiparametric Estimation of a Hedonic Price Function", Journal of Applied Econometrics, 11(6), pp.633-648.

Yatchew, A. (1998), "Nonparametric Regression Techniques in Economics", Journal of Economic Literature, 36(2), pp.669-721.

Pace, R. K. (1998), "Appraisal Using Generalized Additive Models", Journal of Real Estate Research, 15(1), pp.77-99.

Bin, Okmyung, (2004), "A Prediction Comparison of Housing Sales Prices", Journal of Housing Economics, 13(1), pp.68-84.

Wood, S. N. (2003), "Thin Plate Regression Splines " Journal of the Royal Statistical Society:

Series B (Statistical Methodology), 65(1), pp.95–114.
Wood, S. N. (2017), "Generalized Additive Models: An Introduction with R(2nd ed)", New York, CRC Press.
Bailey, Lauria, Lindquist, Mittnik & Rachev (2022), "Hedonic Models of Real Estate Prices: GAM Models; Environmental and Sex-Offender-Proximity Factors", Journal of Risk and Financial Management, 15(12), 601.