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

金東煥（キムドンファン）
明海大学 不動産学部
2026年12月5日

2026年度日本不動産学会秋季全国大会（於：名城大学）

※本発表の内容は個人の見解であり、組織を代表するものではありません。

目次

1. 研究の目的と背景
2. 先行研究
3. 分析データ・分析方法
4. 分析結果
5. 結論

研究の目的

東京都区部の新築戸建住宅の約5万件の成約データを用い、線形OLSと一般化加法モデル（GAM）によるヘドニック分析を比較

① 住宅価格形成における非線形関係を明らかにする。

⇒　土地面積・建物面積・駅距離等を平滑化関数として推定

② 高い予測精度のヘドニック価格モデルを構築する。

⇒　訓練80%／テスト20%分割による予測パフォーマンス検証

③ 各住宅属性の価格への影響を両モデルから比較・検討する。

平均傾き・係数の対比により解釈可能性を維持

手法：R言語mgcvパッケージのgam関数（Wood, 2017）、REML法による半パラメトリック回帰

【参考】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。⇒線形ヘドニック拡張モデル R^2=0.889

研究の背景

- ヘドニック価格分析：住宅を属性の束とみなし、各属性の暗黙価格を回帰的に推定

✓ 従来は対数線形OLSによるパラメトリック回帰が主流

✓ 価格弾力性としての解釈が容易

✓ しかし、関数形（対数線形）を事前に仮定

- **住宅価格と属性の関係は必ずしも線形ではない可能性が高い。**

✓ 土地面積・建物面積・駅距離等で非線形関係の指摘（Yatchew, 1998）

✓ 誤った関数形の仮定は推定精度・予測性能を低下させる。

➢ **関数形を柔軟に推定する平滑化関数で主要変数を推定するGAMが有力な代替になる可能性がある。**

先行研究（ヘドニック理論）

- Court（1939）：自動車価格指数研究が起源
- Lancaster（1966）：財を属性の束として捉える消費者理論
- Rosen（1974）：競争均衡モデルとして体系化

  $P(z) = P(z_1, z_2, \cdots, z_k)$、暗黙価格 $\partial P / \partial z_k$ は需給の均衡価格

- Sirmans et al.（2005）：対数線形の定式化が標準

  $\ln P = \beta_0 + \Sigma \beta_j \ln z_j + \Sigma \gamma_k d_k + \varepsilon$

- 清水（2008）：GISによる道路幅員・駅距離・容積率の精密測定
- 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。※線形ヘドニック拡張モデル R^2=0.889

先行研究（ノンパラメトリック手法）

- Mason & Quigley（1996）：線形OLSで表現できない非線形性の存在を提示
- Anglin & Gençay（1996）：半パラメトリックモデルが予測精度で優位
- Pace（1998）：GAMの不動産評価への先駆的適用
- Bin（2004）：半パラメトリック vs パラメトリックの予測比較
- Wood（2017）：REML法による平滑化パラメータ自動選択（mgcv）を確立
- Bailey et al.（2022）：P-スプラインGAMが線形モデルより高精度

→ 海外では蓄積が進む一方、日本の戸建市場での大規模検証は限られる。

ヘドニック価格モデルの比較

表1　ヘドニック価格モデルの種類と特徴

アプローチ	関数形の仮定	解釈性
線形OLS（パラメトリック）	強い（対数線形）	高い
GAM（半パラメトリック）	主要変数の関数形を柔軟に推定	高い
カーネル回帰（ノンパラメトリック）	なし	低い

● **本研究は半パラメトリックGAMを採用**

・連続6変数（土地・建物・駅・建蔽率・容積率・幅員）→薄板回帰スプライン

・階数・部屋数・各種ダミー→線形項、λはREML法で自動選択

・関数形の柔軟性と解釈可能性を両立

データ概要

表2　使用データ状況

項目	内容
データ提供	業プロ株式会社（2009～2026年1月）
対象	東京都区部の新築戸建住宅
原データ	54,007件
分析サンプル （欠損除去後）	49,254件
訓練／テスト	39,403件／9,851件（80/20）
目的変数	ln_price＝ln（成約価格）

- ランダム分割により過学習を抑制し、テストデータで予測パフォーマンスを検証

基礎統計量（主要変数）

表3　使用データの基礎統計量

変数	平均	標準偏差	中央値	最大値
成約価格（万円）	5,206	1,881	4,880	45,000
土地面積（㎡）	82.07	49.23	81.70	8,579
建物面積（㎡）	94.87	14.50	94.72	322.28
最寄駅徒歩距離（分）	13.36	6.56	12	63
建蔽率	0.591	0.097	0.60	1.00
容積率	1.696	0.602	1.60	8.00

- 多くのデータに対しては、自然対数に変換して使用

※階数（平均2.43）・部屋数（3.01）・幅員（5.28m）は表示を省略

モデル設定

GAM（半パラメトリック）

$$y = f_1(\text{ln_land}) + f_2(\text{ln_building}) + f_3(\text{ln_station})$$
$$+ f_4(\text{ln_bcr}) + f_5(\text{ln_far}) + f_6(\text{ln_road_width})$$
$$+ \gamma_1 \text{ln_floors} + \gamma_2 \text{rooms} + \Sigma \delta_k \text{ward}_k + \Sigma \phi_t \text{year}_t + \Sigma \psi_m \text{dummy}_m + \varepsilon$$

※$f_j(\cdot)$＝薄板回帰スプライン、λはREML法で自動選択

線形OLS（一般的なOLS）

$$y = \beta_0 + \beta_1 \text{ln_land} + \beta_2 \text{ln_building} + \cdots + \beta_8 \text{ln_road_width}$$
$$+ \Sigma \gamma_k \text{dummy}_k + \varepsilon$$

共通仕様

- 区ダミー（中央区基準）・成約年ダミー（2009年基準）・用途地域・接道・構造等
- 評価指標：テストR^2・RMSE・MAE・MAPE

分析結果（予測精度比較）

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	GAM	線形OLS
テストR²	0.8558	0.8416
RMSE※	0.1160	0.1216
MAE※	0.0895	0.0936
MAPE（%）	8.988	9.408

※逸脱度説明率(%)　　86.6

- **GAMが全指標で線形OLSを上回る**

✓ R²：（OLS＝0.8416）→（GAM＝0.8558（+0.014））、MAPE：（OLS＝ 9.41%）
→（GAM＝8.99%（0.4pt改善））

※対数価格上の値

分析結果（予測パフォーマンスの比較）

- テストデータでの実測値と予測値の対応：GAMがOLSよりばらつきが小さい。

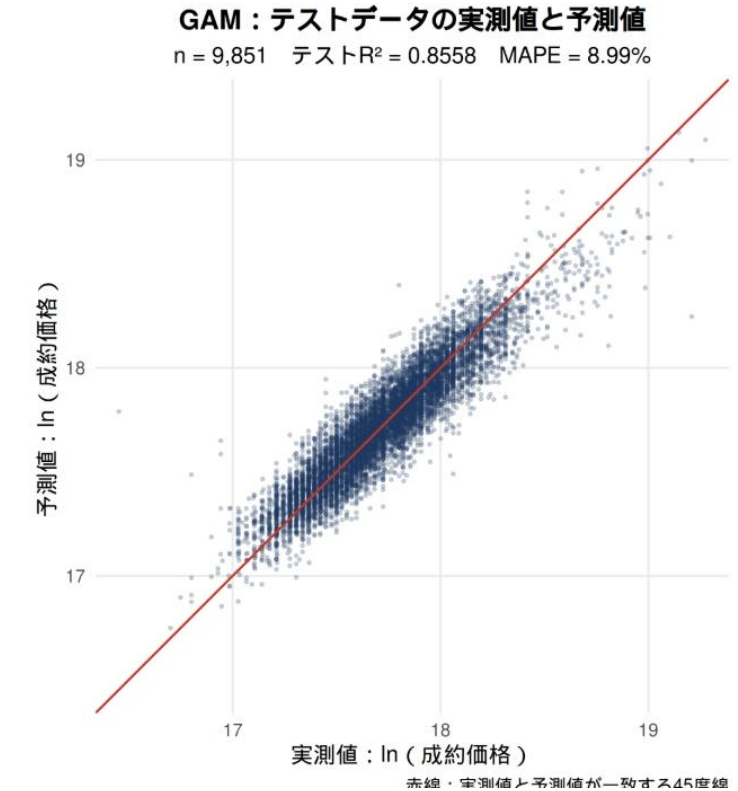
✓ GAMのテスト$R^2$0.8558　OLSのテスト$R^2$0.8416

- 特に高価格帯・駅至近など「分布の端」で両モデルの差が現れる可能性がある。

※テスト(R^2) は、モデルの推定に使用していないデータに対して、どの程度正確に予測できるかを評価

$$R^2_{\text{test}} = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y}_{\text{test}})^2}$$

y_i は実測値、$\hat{y}_i$ は学習済みモデルによる予測値、$\bar{y}_{\text{test}}$ はテストデータの平均値



分析結果（平滑化項の非線形性検定）

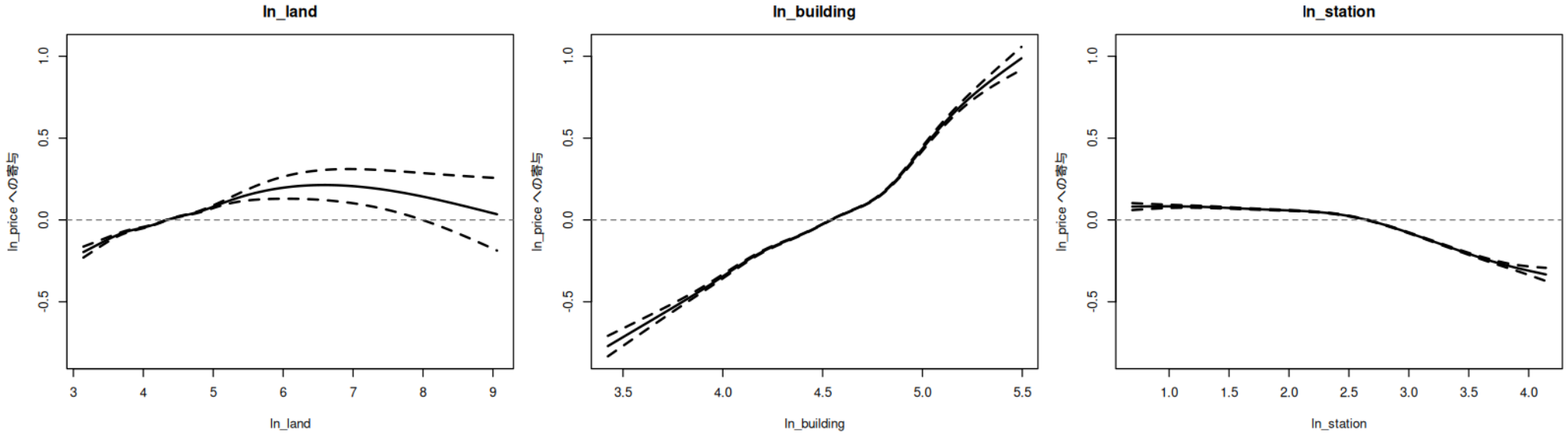
表 5 一般化加法モデル（GAM）平滑化項の有効自由度の検定

平滑化項	EDoF	F値	p値
土地面積	6.285	126.09	***
建物面積	6.897	1827.77	***
駅徒歩距離（分）	6.222	1631.77	***
建蔽率	5.374	13.33	***
容積率	6.807	75.98	***
前面道路幅員	5.578	8.58	***

- 上記6変数が $p<0.001$（***）で有意。EDoF 5.4～6.9 ＞ 直線の1

分析結果（平滑化関数の形状）

図 1 一般化加法モデル（GAM）の平滑化項の非線形関係

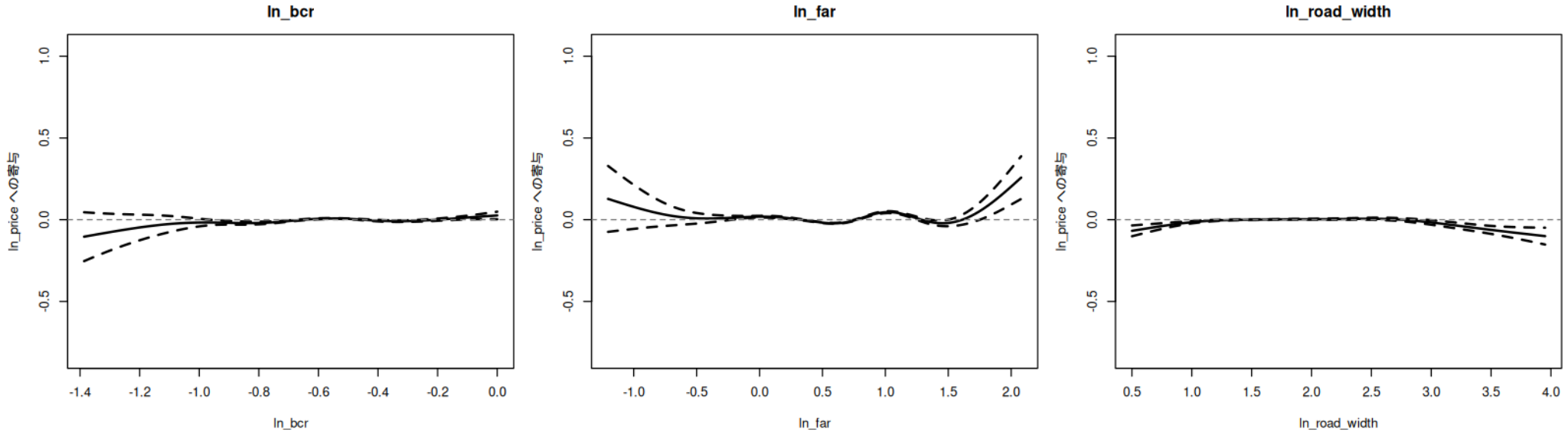


- 建物面積：最大の正の非線形効果（規模拡大で効果が強まる）
- 駅徒歩距離：近距離側では比較的緩やかに価格が低下して、その後価格低下が大きくなり、遠距離側でも低下が続く。
- 土地面積：一定規模までは押し上げ、その後は限界効果が逓減

分析結果（平滑化関数の形状）

- 建蔽率：緩やかに住宅価格を押し上げる。
- 容積率：住宅価格に正の影響を与える。
- 前面道路の幅員：住宅価格への影響は限定的

図 1-2 一般化加法モデル（GAM）の平滑化項の非線形関係



分析結果（平滑化項の平均傾き）

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	方向
土地面積	0.039	正（＋）
建物面積	0.849	正（＋）
駅徒歩距離（分）	−0.12	負（−）
建蔽率	0.094	正（＋）
容積率	0.04	正（＋）
前面道路幅員	−0.009	負（−）

- **建物面積（+0.849）が突出。駅距離はOLS係数（−0.124）と整合**

※平均傾き（AverageSlope）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す（＝グリッド上の端点差分（簡便法））。

分析結果（GAM線形項：区ダミー等）

表 7 一般化加法モデル（GAM）線形項各変数の主要係数

変数	係数	p値
建物階数	−0.107	<0.001***
部屋数	0.005	<0.001***
木造	−0.120	<0.001***
所有権	0.259	<0.001***

区（中央区基準）	係数
港区	+0.332**
世田谷区	−0.342**
杉並区	−0.399***
北区	−0.606***
江戸川区	−0.719***
足立区	−0.906***

- 所有権は成約価格に約+30%・木造は成約価格に約−11%影響する。
- 港区のみ正で足立区と価格差が大きい。

分析結果（線形OLS係数との比較）

表 8 線形 OLS モデルの説明変数・主要ダミーの係数

変数	OLS係数	解釈（弾力性）
建物面積（ln）	0.648***	1%増→+0.65%（最大）
土地面積（ln）	0.111***	1%増→+0.11%
駅徒歩距離（ln）	−0.124***	1%増→−0.12%
建物階数（ln）	−0.127***	他の属性を一定として、階数と成約価格に負の関係
所有権	0.260***	借地権比で高評価

- **OLSとGAMにおける主要な規模・利便性の変数では、方向が一致する。**
  - GAMの付加価値は「形状」（非線形性）の把握。建蔽率等3変数は影響小で省略

【参考】GAMとOLSの予測条件による計算シミュレーション

表　GAMとOLSの予測条件による成約価格の計算シミュレーション結果

物件予測条件：立地	予測条件：土地・建物・駅距離	GAMの予測値	OLSの予測値	差
① 世田谷・標準	80/95㎡・10分	10,429	10,572	−1.4%
② 港区・高級	100/110㎡・5分	26,053	28,779	−9.5%
③ 足立・郊外	70/90㎡・15分	5,377	5,402	−0.5%
④ 杉並・駅近	60/80㎡・3分	8,567	9,634	−11.1%
⑤ 練馬・広め	120/105㎡・12分	9,064	8,945	+1.3%
⑥ 江戸川・角地	85/98㎡・10分	7,652	7,739	−1.1%

- 価格は万円。標準域は差±1.5%以内。②と④は差が拡大（モデルの関数形によって予測値に差が生じる例の可能性がある）。

※その他の予測条件：成約年（2026年成約想定）、構造:木造（②のみ非木造）、接道:接道一方(⑥のみ角地)、用途地域: 第一種低層(②のみ第一種住居)、権利:所有権、地目: 宅地

結論

① GAM（テストR²=0.8558）は線形OLS（テストR²=0.8416）を上回り、半パラメトリックGAMの有効性を確認

② 主要6変数の平滑化項は全て有意（EDoF 5.4～6.9）

✓ 価格との非線形関係の存在を統計的に確認（例：駅距離は近距離側では比較的緩やかに価格が低下して、その後価格低下が大きくなり、遠距離側でも低下が続く。）

③ 区ダミーが両モデルで有意：立地による価格差（港区+0.33～足立区−0.91）

- 留意点：予測精度の差は限定的であり、解釈容易な線形OLSも実務上有用
- 今後の課題：空間GAM（Geographically Weighted GAM）、交互作用項の導入

➢ 本研究は、GAMがOLSより予測精度がやや高く、さらに属性と価格の関係が一定ではないことを可視化した（研究の貢献）。

ご清聴ありがとうございました。