ベイズ推定による小地域単位の賃料形成要因に対する評価の違いに着目した賃貸集合住宅市場の地理的分割分析：Generalized Group Lasso の応用

辺見陽一朗*・井上亮**

Geographic Segmentation Analysis of the Rental Housing Market Focusing on Sub-regional Differences in Rent-determining Factors Using Bayesian Estimation: An Application of the Bayesian Generalized Group Lasso

Yoichiro Hemmi*, Ryo Inoue**

Abstract: The real estate market is often segmented into sub-regions with distinct rent formation mechanisms. This study extends research on geographic market segmentation by incorporating Bayesian estimation into the Generalized Group Lasso (GGL) framework. Specifically, for each explanatory variable we set region-specific coefficients that measure deviations from the study-wide mean effect, and place sparse multivariate-Laplace priors on them. The model defines sub-regions at the *Chome* level and explicitly accounts for adjacent relationships between them. This approach enables both the detection of spatial heterogeneity in rent determinants and the quantification of parameter uncertainty. The proposed Bayesian GGL model is applied to rental apartment data in Suginami Ward, Tokyo. The results demonstrate differences in the evaluation of rent determinants across neighborhoods, providing new insights into the geographic segmentation of the rental housing market.

Keywords: 賃料形成要因，賃貸集合住宅市場，地理的市場分割，Generalized group lasso，スパースベイズ推定

１．はじめに

不動産市場は, 物件属性や需要者層, 地区特性など様々な要因で複雑に細分化され，各市場で異なる賃料形成が成されている．この分割実態の把握は，不動産市場の構造を理解する上で重要である．

その 1 種である地理的市場分割では，高級住宅地を表すブランド地名を持つ町丁目や，教育水準の高い学校区において，周辺地域と異なる賃料形成が行われている現象が見られる．この地理的市場分割の分析に関して，これまで多くの研究がなされてきた．

Goodman and Thibodeau (2003) や Islam and Asami (2011) は，事前設定した複数の小地域区分を利用し，市場分割構造を明らかにした．市場分割単位をより柔軟に把握するため, 黒田ら (2014) は分析上の最小地域区分を設定した上で，その隣接地域の組み合わせで分割単位を表現することとし，焼きなまし法での分割単位抽出法を提案した. 井上ら (2020) は様々な空間単位での市場分割検出法として，Generalized lasso (Tibshirani and Taylor, 2011) の分析を提案したが，各小地域の賃料形成の違いを地域のダミー変数を用いた賃料水準の違いで表すに留まっている．

そこで, 辺見ら (2024) は各説明変数のパラメータも小地域毎に設定し，地域ごとの変数をまとまりとして推定するために Group lasso (Yuan and Lin, 2006) に基づく正則化を導入した．この手法により，地域単位での説明変数効果の有無をまとめて判別できるとともに，Tree-guided group lasso (Kim and Xing, 2010) の木構造を地域の階層構造に応用して分析を行った．小地域固有の説明変数のパラメータは分析対象領域全体の平均的評価からの乖離を表している．しかし，この研究では分析の空間スケールが固定されており，どのような大きさで賃料形成が成されているかを分析できていなかった．また，推定結果として各パラメータの点推定値のみが出力されるため，地域ごとの推定結果の信頼性を評価する上で有用な，パラメータの不確実性に関する情報は得られないことも課題であった．

パラメータの不確実性を定量的に評価できる推定法として，ベイズ推定が挙げられる．ベイズ推定はパラメータ自身を確率変数として取り扱い，パラメータの従う確率分布を推定する．また，パラメータのとりうる値に関して何らかの事前知識を有している場合，その知識を表現した確率分布を事前分布として与えることができる．特に，推定パラメータのうち一部のパラメータの真値が 0 であるという事前知識がある場合では，ラプラス事前分布を使用すると事後分布の MAP 推定値が Lasso 推定

* 学生会員 東北大学情報科学研究科（Tohoku University）
〒980-0845 宮城県仙台市青葉区荒巻青葉 6-6-06 E-mail：hemmi.yoichiro.s1@dc.tohoku.ac.jp
** 正会員 東北大学情報科学研究科（Tohoku University）

値と一致する (Tibshirani, 1996) というスパース事前分布が有用である．Masuda and Inoue (2022) ではこの特性を生かし，Generalized lasso のベイズ推定によって点事象集積領域の検出における有効性を示した．

以上を踏まえ，本研究では辺見ら (2024) の設定に井上ら (2020) で提案された Generalized lasso を加えた，Generalized group lasso (GGL) をモデルとして採用し，そこにスパース事前分布として知られるラプラス事前分布を応用することで，ベイズの枠組みへの拡張を目指す．

本研究では，ベイズ推定による小地域固有のパラメータ値の推定を通して，賃料形成要因に対する小地域単位の評価の違いを定量的な信頼性と併せて把握し，地理的市場分割のより詳細な実態を解明することを目的とする．

２．GGL とスパースベイズ推定

２．１　Generalized group lasso

Alaiz and Dorronsoro (2015) によって提案された Generalized group lasso は，Group Lasso と同じく，複数のパラメータがグループを構成している場合に，各グループのパラメータの L2 ノルムの総和に関する正則化を行う．それに加えて Generalized Lasso のように指定された組み合わせの中の隣接関係にあるグループ同士の L2 ノルムの差の絶対値に関しても正則化を行う．Generalized group lasso の推定は，式 (1) で表される．

$$\min_{\boldsymbol{\beta}} \ \frac{1}{2}\sum_{i=1}^{N}\|y_i-\mathbf{x}_i{}^{\mathrm{T}}\boldsymbol{\beta}\|_2^2+\lambda\sum_{i,j\in A}\left\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_j\right\|_2+\delta\lambda\sum_{i=1}^{N}\|\boldsymbol{\beta}_i\|_2 \quad (1)$$

このモデルは，被説明変数への寄与が小さい説明変数パラメータをグループ単位で 0 に推定できることに加え，隣接関係を設定されたグループ間で影響が類似している場合は共通のパラメータとして推定することができる．

２．２　ベイズ推定とスパース事前分布

ラプラス事前分布は，スパースモデリングの代表的な手法である L1 正則化をベイズ推定の枠組みに拡張するスパース事前分布である．Park and Casella (2008) ではラプラス事前分布を用いた線形回帰モデルの Bayesian lasso を提案している．また，L2 正則化を用いた Group lasso のベイズ拡張は Bayesian group lasso (Raman et al., 2009) として知られている．ここでは多変量ラプラス分布を事前分布として設定することで，事後分布の MAP 推定値は Group lasso 推定値と一致することが示されている．

２．３　既往研究の課題

既往研究における Bayesian group lasso では，地域間の隣接関係を同時に考慮した推定は行われていない．また，Generalized lasso に基づくアプローチでは，地域固有の複数パラメータをグループとして扱うことができない．したがって，隣接関係を考慮しつつ，地域単位のパラメータをグループとして同時にスパース化するベイズ推定が必要である．本研究の新規性は，この枠組みを Generalized group lasso にラプラス事前分布を組み合わせることで実現し，ベイズの枠組みに拡張した点にある．

３．提案手法

３．１　モデル

分析対象領域全体の平均的な影響からの乖離を表す各説明変数のパラメータを小領域ごとに設定する．領域数をGとして，領域g内の物件インデックスをi，物件数をn_gとし，説明変数ベクトルは$\mathbf{x}_{g,i}$，物件賃料は$y_{g,i}$と表す．全体共通のパラメータベクトルを$\boldsymbol{\beta}$，領域g固有のパラメータベクトルを$\boldsymbol{\gamma}_g$とし，β_kと$\gamma_{g,k}$は説明変数の種類数Kの要素を持つ．誤差の標準偏差をσとして，正規分布に従うとすると，線形回帰モデルは式 (2) で表される．

$$y_{g,i}\sim N\left(\mathbf{x}_{g,i}{}^{\mathrm{T}}\left(\boldsymbol{\beta}+\boldsymbol{\gamma}_g\right),\sigma^2\right) \quad (2)$$

次に，ベイズの定理で使用する尤度関数は，式 (2) の回帰モデルから式 (3) の形に自動的に導かれる．

$$p\left(y_{g,i}\mid\boldsymbol{\beta},\boldsymbol{\gamma}_g,\sigma^2\right)=\prod_{g=1}^{G}\prod_{i=1}^{n_g}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(y_{g,i}-\mathbf{x}_{g,i}{}^{\mathrm{T}}\left(\boldsymbol{\beta}+\boldsymbol{\gamma}_g\right)\right)^2}{2\sigma^2}\right) \quad (3)$$

$\boldsymbol{\beta}$と$\boldsymbol{\gamma}$の事前分布には，隣接地域の集合をAと表し，λ_1, λ_2をハイパーパラメータとした式 (4) の同時分布を設定する．βは非正則化パラメータであるため，弱情報事前分布を設定し，$\boldsymbol{\gamma}$には隣接関係の差の正則化と，グループ単位の正則化の部分に多変量ラプラス分布を設定することで，GGL をベイズの枠組みへ拡張する．

$$\begin{aligned}p(\boldsymbol{\beta},\boldsymbol{\gamma}\mid\lambda_1,\lambda_2,\sigma^2)\ =\ &\left[\prod_{k=1}^{K}N(\beta_k|0,5^2)\right]\\ \times&\left[\prod_{(g,g')\in A}Multi\text{-}Laplace\left(\gamma_g-\gamma_{g'}\mid 0,\lambda_1\right)\right]\\ \times&\left[\prod_{g=1}^{G}Multi\text{-}Laplace\left(\gamma_g\mid 0,\lambda_2\right)\right]\\ \times\ &Exp(\sigma\mid 1)\end{aligned} \quad (4)$$

３．２　WAIC によるハイパーパラメータの決定

提案手法の事前分布 (式(4)) に含まれるλ_1, λ_2はハイパーパラメータで，推定前に値を設定する必要がある．本研究では，ハイパーパラメータの決定にあたり，事前に各ハイパーパラメータに複数の候補値を設定し，すべての候補値の組み合わせでパラメータ事後分布の推定・情報量規準 WAIC (Watanabe, 2010) の算出を行う．その上で，マルコフ連鎖列が収束判定を満たし，かつ WAIC が最小となったときの候補値の組み合わせを採用し，ハイパーパラメータを決定する．

４．杉並区の不動産賃料を対象とした分析

４.１　使用データ

本分析では，2022 年にアットホーム株式会社が収集した，東京杉並区を対象とした賃貸集合住宅の募集賃料データを利用する．目的変数は 1 ㎡あたりの賃料，説明変数には専有面積 (㎡)，所在階 (階)，築年数 (年)，最寄り駅までの徒歩時間 (分) を採用する．その中から，新耐震基準を満たす築年が 1982 年以降の物件のみを採用し，所在階が地下の物件は地上階の物件と性質が異なると考えられるため除外した．他にも，敷地面積が 10 ㎡未満，600 ㎡以上の物件，1 ㎡あたりの賃料 (賃料単価) が 1,000 円/㎡未満，10,000 円/㎡以上の物件は入力ミスや外れ値と見なし，除外した．ここから欠損値と重複データを取り除いた全 10,011 件の物件を分析の対象とする．

小地域固有のパラメータは丁目単位で設定し，丁目ごとに K 個の$\gamma_{g,k}$をグループとしている．また，$\boldsymbol{\gamma}_g$の基準地域は「高井戸東 4 丁目」と定めた．この基準は，杉並区の全 138 個の丁目の中から，小地域固有のパラメータを使用せずに賃料モデルを推定し，残差の絶対値を丁目ごと平均した値が最小の丁目である．

４.２　MCMC の設定

提案手法のパラメータ推定は，確率的プログラミング言語である Stan を用いたハミルトニアンモンテカルロ法により MCMC サンプリングを行う．1 回の推定あたり 4 本のマルコフ連鎖列を独立に発生させた後，各チェーンにおいてパラメータの事後分布を 3,000 回サンプリングし，このうち最初の 1,000 回は burn-in period として破棄した．サンプリング終了後，Gelman-Rubin 統計量 $\hat{R}$を確認し，全推定パラメータの $\hat{R}$が 1.1 以下となった場合，マルコフ連鎖列が定常分布に収束したと判定した．

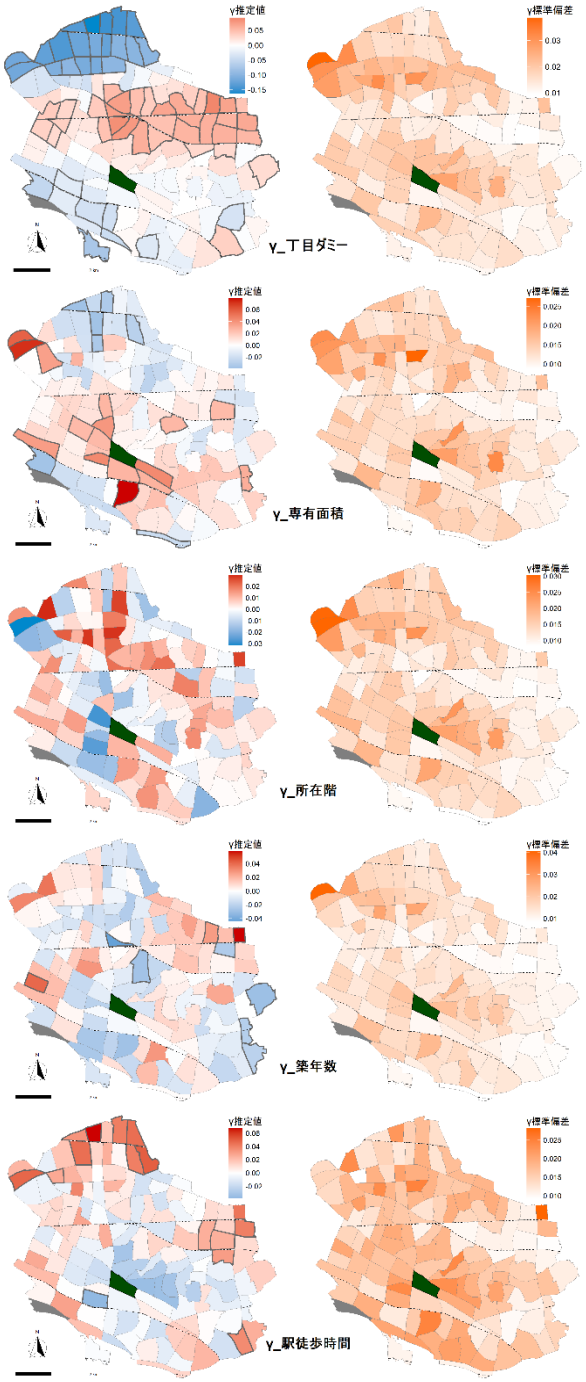
４.３　推定結果

本推定では，ハイパーパラメータ λ_1 は$10^{0.0}$~$10^{2.0}$の範囲で$10^{0.1}$刻み，λ_2は$10^{0.0}$~$10^{4.0}$の範囲で$10^{0.5}$刻みで変化させ，WAIC が最小となる組み合わせを探索した．

物件データが存在しない領域はグレー，基準地域は緑，鉄道路線を白黒の点線で図示している．点推定値は各パラメータの事後分布から MAP 推定値を採用し，図 1 では事後分布の 95%信用区間が 0 を跨いでいない地域を非ゼロに推定されたと見なし，その領域は太線で囲っている．

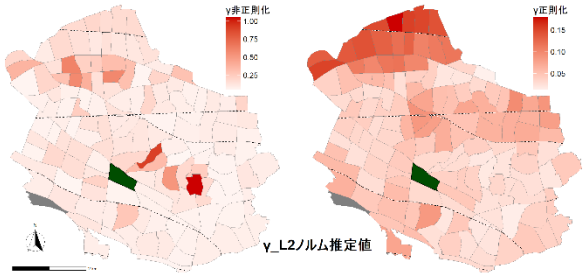
図 2 では，式 (4) の$\boldsymbol{\gamma}$について，$\boldsymbol{\beta}$と同様，平均 0，分散 5 の正規分布に従うとし，他の正則化条件を排除した$\boldsymbol{\gamma}$非正則化の分析を行い，$\boldsymbol{\gamma}$推定値について非正則化と正則化の結果の L2 ノルム推定結果を比較した．

非正則化の最小 WAIC は-16081，正則化をした結果の最小 WAIC は-16276 であった．



(a) 事後分布 MAP 推定値　　(b) 事後分布標準偏差

図 1　杉並区の丁目固有パラメータ推定結果



(a) 非正則化 L2 ノルム　　(b) 正則化 L2 ノルム

図 2　正則化の有無による$\boldsymbol{\gamma}$の L2 ノルム推定結果比較

表 1　全体共通パラメータ**β**の推定値

	推定値	標準偏差
切片	8.0553	0.0074
専有面積 (㎡)	-0.0966	0.0058
所在階 (階)	0.02052	0.0059
築年数 (年)	-0.0867	0.0065
駅徒歩時間 (分)	-0.0222	0.0068

γの推定結果を表す図 1 では，丁目ダミーは西武新宿線や井の頭線沿線で負，中央線や丸の内線沿線で正に推定される結果となった．最寄り駅までの徒歩時間では，鉄道沿線では正の傾向があり，逆に沿線から離れた中央線と井の頭線の中間地域などでは，負に推定される領域が多いことが分かる．一方，所在階と築年数の**γ**推定値は正負が混合する解釈が難しい結果となった．

専有面積に注目すると，井の頭線中央部付近でまとまって正に推定されており，周辺地域の地区特性を確認すると，河川沿いの緑地や大規模な公園が集積している地域である．これは自然環境を重視する居住者は家族連れ世帯が多いと考えられ，これらの地域にはその層をターゲットとした大規模集合住宅が点在していることから，局所的な需要者層の違いによって，物件の専有面積に対する評価が全体平均よりも高くなっていると考察できる．

また，全体的な傾向として，非ゼロに推定された地域の標準偏差は小さく推定されたことから，本分析の結果が不確実性の少ない有意な結果であることも示された．

図 2 の比較結果では，提案手法による正則化によって，賃料形成において特徴的なパターンを示す地域を明確に抽出できていることが分かる．さらに，空間的な平滑性を組み込むことで，鉄道沿線に沿って特徴的な領域がまとまって現れるなど，地理的に意味のある地域構造を捉えられている．このことから，隣接する地域間では賃料形成の構造が似通っていることが考察できる．

５．おわりに

本研究では，賃貸集合住宅市場における地理的市場分割の実態を明らかにすることを目的として，小地域単位での賃料形成要因の評価の違いに着目した分析を行った．具体的には，Generalized group lasso をベイズ推定の枠組みに拡張し，ラプラス事前分布を導入することで，小地域固有のパラメータの不確実性を考慮しつつ，地域の特徴的な賃料形成パターンを抽出できるモデルを構築した．

杉並区を対象とした実証分析の結果，鉄道沿線や公園周辺など，地区特性に関連して一部の説明変数の評価が全体平均から乖離する地域が捉えられた．特に専有面積や最寄り駅までの徒歩時間において，局所的な需要者層の違いが賃料形成に反映されていることが確認された．また，隣接関係を考慮した正則化によって鉄道沿線に沿ったまとまりなど空間的に意味のある地域構造が抽出され，地域の賃料形成構造を解釈可能な形で明らかにすることができた．さらに，非ゼロ推定された地域では標準偏差が小さく，不確実性の少ない安定した推定が得られたことから，GGL をベイズの枠組みに拡張した本手法の有効性が裏付けられた．

以上より，本研究は小地域単位における賃料形成要因の評価を定量的かつ信頼性を伴って把握できる枠組みを提示した点で意義があり，今後の不動産市場分析においても有用な手法となりうると考えられる．

謝辞

本研究は，国立情報学研究所の IDR データセット提供サービスによりアットホーム株式会社から提供を受けた「アットホームデータセット」を利用した．また本研究は，JSPS 科研費 21H01447・24K00997 の助成を受けた．

参考文献

井上亮，石山里穂子，杉浦綾子 (2020) 東京都区部の賃貸マンション市場の地理的分割の実態把握－スパースモデリングによるアプローチー. 土木学会論文集 D3 (土木計画学), 76(3), 251–263.

黒田翔，堤盛人，奈良岡浩二 (2014) 価格構造の空間的異質性に基づく不動産市場の地理的細分化．ジャレフ・ジャーナル (不動産ファイナンス・不動産経済学研究), 7, 1–18.

辺見陽一朗，井上亮 (2024) 小地域毎の賃料形成要因に対する評価の違いに着目した賃貸集合住宅市場の地理的分割分析．GIS 学会 学術研究発表大会講演集, 32, 96(60).

Goodman, A. C., and Thibodeau, T. G. (2003) Housing market segmentation and hedonic prediction accuracy. *J. Housing Econ.*, 12(3), 181–201.

Islam, K. S., and Asami, Y. (2011) Addressing structural instability in housing market segmentation of the used houses of Tokyo, Japan. *Procedia Soc. Behav. Sci.*, 21, 33–42.

Kim, S., and Xing, E. P. (2010) Tree-guided group lasso for multi-task regression with structured sparsity. *ICML 2010*, 1–14.

Masuda, R.., Inoue, R. (2022) Point-event cluster detection via the Bayesian generalized fused lasso. *ISPRS International Journal of Geo-Information*, 11 (3): 187.

Park, T., and Casella, G. (2008) The Bayesian lasso. *J. Am. Stat. Assoc.*, 103(482), 681–686.

Raman, S., Fuchs, T. J., Wild, P. J., Dahl, E., and Roth, V. (2009) The Bayesian group-lasso for analyzing contingency tables. *ICML 2009*, 881–888.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B*, 58(1), 267–288.

Tibshirani, R., and Taylor, J. (2011) The solution path of the generalized lasso. *Ann. Stat.*, 39(3), 1335–1371.

Watanabe, S. (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.*, 11, 3571–3594.

Yuan, M., and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. *J.R. Stat. Soc. B*, 68(1), 49–67.