

東京都区部の賃貸住宅市場における 地理的区分の把握 — スパースモデリングによるアプローチ —

井上 亮¹・石山 里穂子²・杉浦 綾子³

¹正会員 東北大学准教授 大学院情報科学研究科人間社会情報科学専攻
(〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06)

E-mail: rinoue@tohoku.ac.jp

²学生会員 東北大学博士前期課程 大学院情報科学研究科人間社会情報科学専攻
(〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06)

E-mail: rihoko.ishiyama.r5@dc.tohoku.ac.jp

³非会員 株式会社フロネシス
(〒105-0003 東京都港区西新橋1丁目14-9)

E-mail: a.sugiura@phronesis.link

不動産価格の形成要因やその評価は地域により異なるとされ、不動産市場は地理的に分割されていると考えられている。本研究は、賃料データに基づき賃貸マンション市場の地理的区分を把握するため、地理的に分割された賃料形成を表現できるような地域別のパラメータを設定した賃料モデルに対し、スパースモデリング(Fused lasso)による変数選択を通じた分析を提案した。東京都区部を対象とした分析から、賃貸マンション市場は、区、および、鉄道沿線や駅勢圏、町丁目など、様々な大きさの空間領域別に賃料形成が行われている実態を確認した。

Key Words: housing rental market, geographical segmentations, sparse modeling, generalized fused lasso, Tokyo metropolitan area

1. はじめに

不動産市場は、需要層や地区特性、物件種別・属性などの違いにより複雑に細分化され、各市場で異なる価格形成が行われている。その実態把握は、ある地域の不動産市場構造を理解する上で重要である。この分析アプローチの一つは、物件位置に着目して不動産価格形成要因およびその影響が共通である地域を抽出する分析である。

同一の都市圏内でも、不動産価格形成要因に対する評価が地域により異なることはよく知られている。例えば、都心居住を指向する人々は交通便利性や都市が提供する様々なサービスを受けられる生活環境を重視するのに対し、郊外居住を指向する人々はより広い住宅や自然環境への近接性などを重視する傾向にあると考えられる。物件の環境要因から生じる各地域の不動産市場に参加する需要層の違いは、各地域の物件属性に対する評価の違いに反映され、地理的な市場細分化を産む。

また、住宅地として人気の自治体や鉄道路線・鉄道駅、

高級住宅地を表すブランドと認識されている町丁目などが存在する。それらの地域・地区では、その人気故により高品質な物件が供給され、高い価格が形成される。

このように、様々な要因によって生じる不動産市場の細分化の多くは、地理的な位置と関連して生じている。地理的細分化の実態を実務者は経験的に把握し、地域別の不動産情報提供や、各物件の評価などに反映している。また、研究者は、細分化の定量的な把握やその成因の分析に大きな関心を寄せており、これまでに多くの研究がなされてきた（例えば、Bourassa *et al.*¹⁾, Goodman and Thibodeau²⁾, Islam and Asami³⁾, 黒田ら⁴⁾など）。しかし地理的細分化は、市町村や鉄道沿線、最寄り駅、町丁目単位など、スケールの異なる様々な空間領域単位に影響を受けると考えられ、その構造を考慮した分析は容易ではない。

また、不動産市場の地理的区分の構造を把握するには、現在公開の公的な不動産価格情報は十分な空間解像度を有さないという課題もある。まず、地価公示や地価調査は、不動産取引の指標となるように周辺地域の標準的な

属性を持つ土地の価格を半年毎に公表するが、件数は必ずしも多くない。比較的密に情報提供される都心部でも数百メートル程度の間隔を置いており、町丁目の持つブランド力の評価などには十分な空間解像度とはいえない。地価公示や地価調査のみによる不動産市場動向の把握には限界があることを受けて、平成 18 年から不動産取引価格が公開されている。件数は公示地価よりも多いものの、取引事例は空間上で均等には分布しない上、取引当事者間の個別事情などが反映された比較が難しいという特徴を有する。取引事例の件数は、地価公示などの公的地価指標よりは多いものの、不動産市場の地理的市場分割の状況を把握することは難しい。

一方、賃貸市場では売買市場よりもはるかに多くの物件が取引されており、特に都市部では取引物件の空間密度は高い。賃貸住宅市場に限定すれば、募集賃料は賃貸情報サイトなどを通して広く公開されており、また、多数の市場参加者が関わる中で取引が成立しているため、売買価格よりも賃料には個別事情が反映される余地が小さい。また、集合住宅では、面積や間取りなど物件の属性が規格化されており、集合住宅の賃料は分析が比較的容易であるという利点がある。賃貸市場に集合住宅の部屋が多数出ている都市部においては、賃料分析を通して地理的市場細分化の実態を分析できる可能性が高い。

また、近年、 L_1 正則化による変数選択が可能なスパースモデリング手法の拡張として、generalized fused lasso が提案されている。指定したパラメータ間の差に対する正則化項を設定した分析ができる手法で、地理分析においては、隣接地域のパラメータの差に対して正則化項を設定すると、隣接地域で共通のパラメータを推定することが可能である。この手法は、地理的市場細分化の分析において、価格形成が共通の領域を柔軟に抽出できる可能性があると期待される。

そこで、本研究は、東京都区部を対象に、マンションの賃料データに基づいて、不動産市場の地理的細分化の実態把握を行う。地域・地区毎に異なる賃料形成を表現できるように、地域的要因を表す大量の説明変数を設定した賃料モデルを設定した上で、generalized fused lasso を用いた変数選択を通して、賃貸マンション市場における賃料形成の地域的差異の把握を目指した分析を行う。

以下に、本論文の構成を示す。第 2 章では、不動産価格分析に関する既往研究をまとめ、地理的区分を把握する際の問題を点を明らかにする。第 3 章では、第 2 章で明らかにされた問題点に対するアプローチとして、本研究で用いるスパースモデリング手法である、fused lasso について説明する。第 4 章では、実際に東京都区部の賃貸マンションデータを用いた分析を行い、賃貸市場の地理的区分について考察する。最後に第 5 章では、本研究の成果を整理し、展望を記す。

2. 不動産市場における地理的分割に関する既往研究

本章では、不動産市場の地理的細分化に関心を寄せた既往研究を整理する。

Wu and Sharma⁵⁾は、様々な物件属性を利用した主成分分析によるクラスタリングを通して設定したサブマーケットを用いて分析している。また、不動産価格が持つ空間的な特性を踏まえ、それらを構造化したモデルによる分析もなされており、例えば、Montero *et al.*⁶⁾は空間回帰モデル、大庭ら⁷⁾、Kestens *et al.*⁸⁾は、地理的加重回帰モデル(GWR)を用いている。

一方、地域分割をモデルで明示的に考慮する方法として、分割単位ごとのパラメータ推定や分割単位をダミー変数としたモデル推定など考えられる。例えば、藤澤・隅田⁹⁾は、マンション価格形成における属性や立地場所の影響を明らかにするため、間取りタイプや駅までの所要時間などの属性に関するパラメータを沿線や地域ごとに推定し、その比較・考察を行っている。Bourassa *et al.*¹⁾、Goodman and Thibodeau²⁾、Islam and Asami³⁾は、複数の地域区分による分析を通して、予測精度を基準にモデルを選択している。しかし、これらの研究では、予め決められた地域区分に依存した分析に留まっており、地域区分の単位について検討が不十分である。どのように地域区分を設定し、変数に組み込むのかという変数選択上の問題が生じていると考えられる。黒田ら⁴⁾は、地域分割の分析最小単位を設定した上で、隣接地域の結合と価格モデルの推定を焼き鈍し法を用いて同時に行う手法を提案している。しかし、異なるスケールの空間領域による影響を考慮できていないことや、分割数を推定することはできないなどの課題がある。

3. 変数選択手法

本研究では、地域によって異なる価格形成を表現できるように、大量の説明変数候補を与えた上で、変数選択を通して賃料モデルを推定し、共通のパラメータが推定される領域を抽出する分析アプローチを取る。

大量の説明変数候補を与えると、可能なモデルの数は膨大となる。例えば、説明変数の候補が M 個ある場合に、漏れなく探索するには、 $2^M - 1$ 種類のモデルから選ばなくてはならない。そこで、本章では効率的に変数選択を行う手法を説明する。

(1) 評価基準

説明変数を増やせばデータへの適合度は増すが、新しいデータに対する予測(汎化)性能は低下することが知

られており、機械学習分野では過学習と呼ばれる。そこで、予測性能の評価方法として、交差検証法 (Cross Validation, CV) や、赤池情報基準 (Akaike's Information Criterion, AIC) などが提案されている。

CV とは、データを訓練データと検証用データに分け、訓練データからモデルの推定を行い、検証データで予測の良さを推定する方法で、応用範囲が広い。CV で用いられる評価基準にはモデルの対数尤度や平均二乗誤差 (MSE) などがある。

一方、AIC とは、モデルの良さを測るための指標である。 θ をパラメータの最尤推定量、 $L(D|\theta)$ をデータ D が所与の下で、尤度関数を最大化する θ を得たときの尤度関数の値、 k をモデルの未知パラメータ数と表すと、AIC によるモデル選択は、式(1)を最小化するモデルを選ぶ操作となる。

$$AIC = (-2)[L(D|\theta) - k] \quad (1)$$

AIC は、パラメータ数で表現されるモデルの複雑さとデータへの適合度のバランスを取りモデルを選択する。

(2) 古典的な変数選択手法

説明変数の候補となる変数が多い場合、すべての説明変数の組み合わせを試しモデルを評価するのは現実的ではない。古典的な方法としては、一定の規則に従って変数を逐次選択していく逐次選択法がある。その代表的な方法である変数増減法¹⁰⁾は、適当なルールで説明変数を増減して AIC の最小となるモデルを探す方法である。しかし、説明変数が増えれば増えるほど、局所的最小や過大な計算時間への不安は免れない。

(3) L_1 正則化(lasso): スパースモデリング

L_1 正則化 (lasso : least absolute shrinkage and selection operator) はスパースモデリングの代表的な手法である。スパースモデリングとは、回帰分析に限らず、対象がスパース性 (疎性) を持つという仮定を積極的に用いて、与えられたデータに対し本質的な部分を自動的に抽出する、という広い概念を持ったものである¹¹⁾。

スパース性 (疎性) とは、まばら・わずかという言葉で表されるような、情報の大部分が 0 である性質を指す。スパース性は、画像、動画、音声といった、様々なデータに存在している。近年、スパースモデリングはビッグデータの解析、あるいは不良設定問題を解く手法として注目されている。具体的には、ゲノムなどのデータ解析や MRI などの画像処理技術、天文学の物理計測の解析、マーケティング分析など、幅広い分野で実用が期待される技術である¹¹⁾。

回帰分析におけるスパースモデリングは、多数の説明変数の中から、少数の必要なものを取り出すことのでき

る効率的な変数選択手法と言える。

lasso¹²⁾は、回帰モデルの損失関数にパラメータの L_1 ノルムに基づく正則化項を加えた正則化損失関数を最小化するパラメータを推定する方法である。なお、以下の説明は、式(2)の関数を用いて示す。ただし、 $n \times 1$ の被説明変数ベクトルを \mathbf{y} 、 $n \times k$ の説明変数行列を \mathbf{X} 、 $k \times 1$ のパラメータベクトルを $\boldsymbol{\beta}$ 、平均 0・分散 σ^2 ・共分散 0 の正規分布に従う $n \times 1$ の誤差項ベクトルを $\boldsymbol{\varepsilon}$ とする。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

このとき、lasso は次の最適化問題として定式化される。

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \quad \text{subject to } \|\boldsymbol{\beta}\|_1 \leq t \quad (3)$$

ただし L_1 、 L_2 ノルムの定義は以下の通りである。

$$\|\mathbf{z}\|_1 = \sum_i |z_i| \quad (4)$$

$$\|\mathbf{z}\|_2 = \sqrt{\sum_i z_i^2} \quad (5)$$

また、 t は正の定数で、正則化パラメータである。

lasso の最適化問題では、OLS の最適化問題に L_1 正則化項で表される制約条件が加わることで、パラメータが 0 に推定されやすいという特徴を有する。一方、式(6)で表される L_2 正則化 (Ridge 回帰) は、推定値の分散は抑えられるがパラメータは 0 に推定されにくい(図-1)。

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \quad \text{subject to } \|\boldsymbol{\beta}\|_2^2 \leq t \quad (6)$$

最適値を与える $\boldsymbol{\beta}$ は、正則化パラメータ t により変化し、 t を小さくすると多くの成分が 0 となる。一方、 t を大きくすると、 $\boldsymbol{\beta}$ の非 0 成分の数が増える(図-2)。

Lagrange 未定乗数を使って定式化すると、式(7)の正則化項付き最小化問題と等価になり、正則化パラメータ t は、 t が小さくなるにつれて大きくなる。

$$\min_{\boldsymbol{\beta}} \left[\frac{1}{2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right] \quad (7)$$

Ridge 回帰と同様に、 t や λ は先験的に与える必要があり、CV などを用いて最適な値を選択することが多い¹²⁾。

(4) Generalized fused lasso

Fused lasso¹³⁾とは、効率的に前後のパラメータの差の有無を判断することができる手法で、時系列データにおける変化時点の抽出などに用いられる。 y_i, x_i^j, β_j を、それぞれ前出の $\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}$ の各成分とし、Fused lasso は、lasso に新たな L_1 正則化項を加えた次式で表される。

$$\min_{\boldsymbol{\beta}} \left[\frac{1}{2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^k \beta_j x_i^{(j)} \right)^2 + \lambda \sum_{j=1}^k |\beta_{j+1} - \beta_j| + \gamma \lambda \sum_{j=1}^k |\beta_j| \right] \quad (8)$$

ただし、 γ は 0 以上の定数で、2つの正則化項間の重みを表すハイパーパラメータである。

Fused lasso では、前後のパラメータ間の差の絶対値の

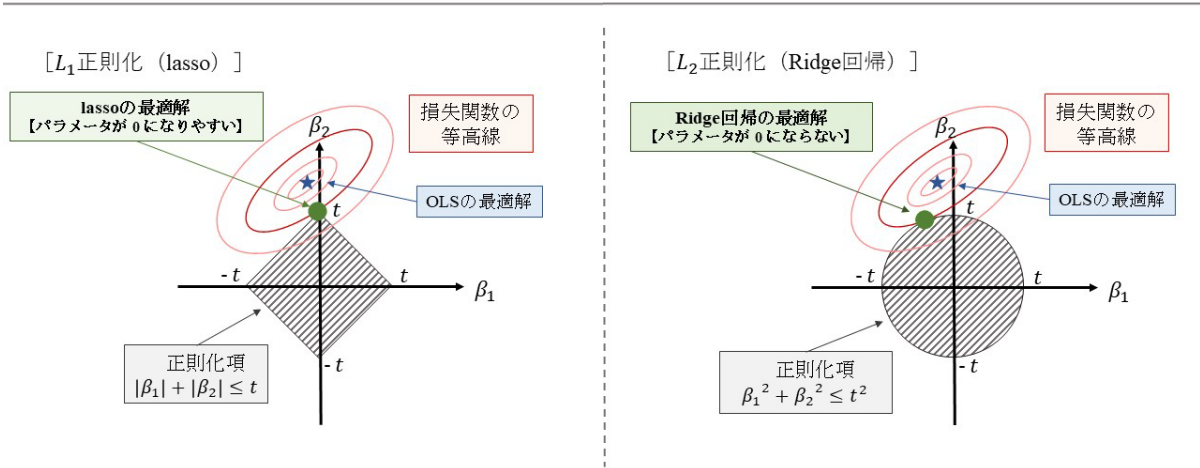


図-1 lasso と Ridge 回帰の最適化

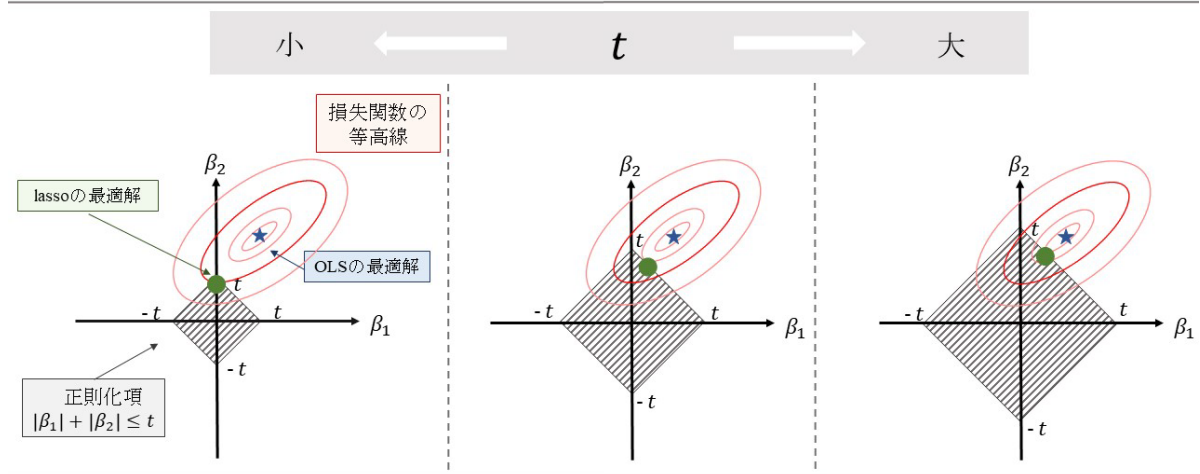


図-2 正則化パラメータの大小による最適値の変化

和を取った L_1 正則化項が加わり、パラメータの差も 0 に近づき、共通のパラメータを推定することができる。

Generalized fused lasso は、パラメータ間の差について、前後のパラメータだけではなく、任意の隣接関係に基づくパラメータの組み合わせに拡張した式(9)で表される。

$$\min_{\beta} \left[\frac{1}{2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^k \beta_j x_i^{(j)} \right)^2 + \lambda \sum_{(m,n) \in E} |\beta_m - \beta_n| + \gamma \lambda \sum_{j=1}^k |\beta_j| \right] \quad (9)$$

ただし、 E は、設定する隣接関係に対応した各成分の組み合わせを表す集合である。

Generalized fused lasso は、隣接地域のパラメータ内に L_1 正則化項を設け、共通のパラメータを推定することで、地域区分の推定を行うなど、地理的分析への応用もされている。例えば Wang and Rodriguez¹⁴⁾ は、フロリダにおける小児がんの発症率に関するパラメータを地域ごとに設け、共通のパラメータを推定することで、高い発症率と考えられる地域を特定することを試みている。

本研究には、Generalized fused lasso を用いて、隣接地域のパラメータ間に L_1 正則化項を設け、共通のパラメータを推定することで、より柔軟に東京都区部の賃貸住宅

市場の地理的区分を把握できると期待される。区や町丁目など地域区分の最小単位は与えるものの、隣接地域内で共通のパラメータを推定することで、その区分に依らないより広域的な地理的区分を捉えることができると考えられる。また、パラメータ自体の L_1 正則化項も加わっているため、重要ではない説明変数については、パラメータが 0 に推定されやすく、効率的に変数選択が行える。

4. 東京都区部の賃貸マンション市場の地理的分割の分析

本章では、東京都区部の賃貸マンション市場における地理的な市場分割に関して、Generalized fused lasso を用いた分析を示す。

賃料形成要因の評価の違いを分析する最小空間単位を区とし、経過年数、専有面積、最寄り駅までの徒歩時間、所在階数、部屋数の基本的な物件属性に対するパラメータを区毎に推定する。さらに、鉄道沿線や、駅勢圏、町丁目目で設定した地区毎の賃料水準の違いをダミー変数で

表現するモデルを立て、地域的差異の有無を推定する。

(1) 使用データ

アットホーム株式会社が収集した、2015・2016 年の東京都区部のマンションの募集賃料データを用いる。

管理費・共益費を含む実質賃料を専有面積で割った 1 平方メートル当たりの賃料を分析対象とする。なお、以後、賃料単価と記す。また、築年・築月とデータの最終更新日から算出した経過年数と、住所から抽出した町丁目名、専有面積、最寄り駅名、最寄り駅までの徒歩時間、所在階層、部屋数を分析に使用する。なお、高層マンションでは、設備等の違いにより、低層マンションとは異なる賃料形成が行われていると見込まれたため、本分析では、地上階層 15 階以下の建物の部屋を対象とする。その中から、上記データ項目に欠損があるデータを除いた 270,605 件を分析に用いる。

また、ダミー変数は下記の通り設定した。物件の属する鉄道沿線を表すダミーは、Yahoo!路線情報を参考に、最寄り駅を通る複数の鉄道沿線に関するダミー変数を設定した。ただし、駅勢圏ダミーと全く同値となる沿線ダミーは除外した。駅勢圏ダミーと町丁目ダミーは、それぞれ平均賃料単価が中央値となる平和島駅と南砂 3 丁目を基準として、ダミー変数を設定した。

使用データの概要を表-1、表-2 に示す。

(2) 分析モデル

分析では、賃料単価の対数を被説明変数、経過年数の対数、専有面積、最寄り駅までの徒歩時間、所在階層、部屋数の数値属性と町丁目ダミー、鉄道沿線ダミー、駅勢圏ダミーを説明変数とし、数値属性に関して各区のパラメータを設定した賃料モデルで分析する。なお、数値属性は、平均 0・分散 1 に標準化した値を用いる。\$P_{ward}\$ を区毎に推定するパラメータの集合、\$Ward\$ を区の集合 (23 区)、\$cho\$ を町丁目の集合 (2932 町丁目)、\$line\$ を鉄道路線の集合 (59 路線)、\$station\$ を駅勢圏の集合 (474 駅) と表し、

\$\beta_0\$ を定数項、\$\beta_{pw,w}^{ward}\$ を区毎に推定するパラメータ \$p_w\$ の \$w\$ 区のパラメータ、\$\beta_c^{cho}\$ を町丁目ダミー \$c\$ のパラメータ、\$\beta_l^{line}\$ を鉄道沿線ダミー \$l\$ のパラメータ、\$\beta_s^{station}\$ を駅勢圏ダミー \$s\$ のパラメータと表すと、賃料モデルは式(10)で表せる。

$$y_i = \beta_0 + \sum_{p_w \in P_{ward}} \sum_{w \in Ward} \beta_{p_w,w}^{ward} x_{ip_w,w}^{ward} + \sum_{c \in cho} \beta_c^{cho} d_{ic}^{cho} + \sum_{l \in line} \beta_l^{line} d_{il}^{line} + \sum_{s \in station} \beta_s^{station} d_{is}^{station} + \varepsilon_i \quad (10)$$

\$Trans\$ を賃貸物件の集合、\$Neighbor_{ward}\$ を隣接区の組み合わせ 55 組の集合、\$Neighbor_{cho}\$ を隣接町丁目の組み合わせ 5,029 組の集合、正則化パラメータを \$\lambda, \gamma\$ とすると、Generalized Fused lasso に基づくパラメータ推定を表す最適化問題は式(11)と書ける。

$$\min_{\beta} \left[\frac{1}{2} \sum_{i \in Trans} \left(y_i - \sum_{p \in P} \beta_p x_{ip} \right)^2 + \lambda \sum_{p_w \in P_{ward}} \sum_{(a,b) \in Neighbor_{ward}} \left| \beta_{p_w,a}^{ward} - \beta_{p_w,b}^{ward} \right| + \lambda \sum_{(c,d) \in Neighbor_{cho}} \left| \beta_c^{cho} - \beta_d^{cho} \right| + \gamma \lambda \sum_{p \in P} \left| \beta_p \right| \right] \quad (11)$$

第 2 項は、区毎に推定する、経過年数、専有面積、最寄り駅までの徒歩時間、所在階層、部屋数に対する 5 種類のパラメータについて、隣接区のパラメータとの差に制約を課している。第 3 項は、町丁目ダミーについて、隣接町丁目とのパラメータの差に対する制約を課す。また、第 4 項は、すべてのパラメータの絶対値に関する制約を課しており、第 2・3 項との重み付けを正則化パラメータ \$\gamma\$ で行う。

正則化パラメータ \$\lambda, \gamma\$ の両方を変化させ、AIC 最小となるモデルを探索する。\$n\$ を標本数、\$S_e\$ を残差 2 乗和、\$p\$ をパラメータ数とし、線形回帰モデルの誤差項が正規分布に従うとした場合、AIC は式(12)と書ける。

$$AIC = n \left\{ \log \left(\frac{2\pi S_e}{n} + 1 \right) \right\} + 2p \quad (12)$$

表-1 使用データ 被説明変数と数値属性

変数	平均値	標準偏差	最大値	最小値
賃料単価(円/㎡)	3123.34	804.19	6999.50	1000
経過年数(年)	21.63	11.99	69.17	0
専有面積(㎡)	36.33	19.03	440	10
最寄り駅までの徒歩時間(分)	6.75	4.18	60	0
所在階層(階)	3.77	2.58	15	1
部屋数(室)	1.41	0.66	8	1

表-2 使用データ ダミー変数

ダミー変数	対象	個数
町丁目ダミー	賃貸取引物件が存在したすべての町丁目 ※基準町丁目：南砂 3 丁目	2,932
鉄道沿線ダミー	すべての鉄道沿線	59
駅勢圏ダミー	賃貸取引物件が存在したすべての駅勢圏 ※基準駅勢圏：平和島駅	474

(3) 分析結果

本分析では、 γ に関する 0.001, 0.1, 1, 10 の 4 通りの設定に対して推定を行った。 $\gamma=1$, パラメータ数 673 の時に AIC が最小となり、自由度調整済み決定係数は 0.758 であった。この推定結果の詳細を以下に示す。なお、標準化した数値属性に対するパラメータは、推定値を標準偏差で除して、元の数値に対するパラメータの大きさを表している。推定されたパラメータ数を表-3 に示す。Generalized fused lasso によって、約 95%の町丁目ダミーと約 23%の駅勢圏ダミーに関して、パラメータが 0 に推定され、また、多数の隣接する町丁目で共通のパラメータ推定値が得られた。

一方、区毎に推定した物件属性に対するパラメータは、経過年数と最寄り駅までの徒歩時間については各区ですべて異なる推定値が得られ、専有面積、所在階層、部屋数については、一部の隣接区のみで共通の推定結果となった。専有面積と所在階層に関するパラメータの推定結果を図-3・4 に示す。斜線部は、推定結果が等しい範囲を表す。それぞれ都心区と郊外区で推定パラメータの水準に違いがあるようだが、共通の推定値となったのは一部の区に限られる。賃料の評価が、地理的に異なる実態が存在し、一部の区で共通の賃料市場が形成されている可能性が示唆される。ただし、これらの物件属性に対する評価は、区よりも細分化されているため、区単位に設定した分析ではその実態が抽出できていない可能性も考えられる。

次に、鉄道路線および駅勢圏のダミー変数に関するパラメータ推定結果を示す。鉄道沿線ダミーに対するパラメータ推定値を図-5、鉄道沿線ダミーと駅勢圏ダミーのパラメータ推定結果を足し合わせ、駅勢圏に対する評価と表した結果を図-6 に示す。

まず、lasso によって、一部の鉄道路線・駅勢圏のパラメータを 0 と推定することができており、鉄道路線に対する評価を駅勢圏に対する評価から分離することができた。鉄道路線に対する評価については、都心路線や南西部に伸びる路線の評価が高く、北東部の路線よりも高いことが推定されている。東京都区部の不動産市場では、

南東側に価格の高い領域が広がっていることが知られており、その特徴が鉄道路線に関するパラメータ推定結果に現れていることが確認できる。

次に、駅勢圏に対する評価に注目する。業務・商業地域が広い東京などの都市を対象とした分析では、交通利便性を表す説明変数の設定が難しい。そこで、本分析では交通利便性を明示的に表す説明変数を入れずに、駅勢圏ダミーで表現している。図-6 から、港区や渋谷区周辺の駅勢圏が高く評価されている様子が推定され、また郊外に向かうにつれて評価が下がる実態が表現されている。本分析のように、大量の賃料・価格情報を利用した分析を行うと、交通利便性に対する評価をデータから抽出できる可能性があることが確認できた。

最後に、区や鉄道沿線、駅勢圏よりも、より局所的な地区における、賃料水準の違いを表現することを目指した町丁目ダミーについて、そのパラメータ推定結果を図-7 に示す

都心区以外については、ほとんどの町丁目ダミーのパラメータが 0 に推定され、個別の物件属性や鉄道沿線・駅勢圏ダミーで賃料水準を表現できていることが確認された。一方、都心区では、正・負のパラメータが推定された地域が混在している様子が観察され、局所的な範囲で賃料水準が異なることが確認できる。

とりわけ港区や渋谷区では、多くの町丁目ダミーについて 0 以外の値が推定された。その一例として広尾・白金周辺の推定値を図-8 に示す。なお、緑線は共通のパラメータとして推定された範囲を表し、駅勢圏と町丁目のパラメータ値の凡例は図-5・6 に準ずる。広尾や白金など高級住宅地として評価の高い都心地域では、同一駅勢圏内でも地区によって異なる賃料水準を有することが示されており、町丁目ダミーで表現可能な局所的な賃料形成が行われている実態が確認できた。また、一部の隣接町丁目では共通の値が推定され、本研究で提案した、大量の地区ダミーを設定したモデルに対して、地区の隣接関係に基づく制約を課した上で、Generalized fused lasso を用いて分析するアプローチによって、賃貸マンション市場の地理的分割の実態を把握できる可能性が確認された。

表-3 推定されたパラメータ数

説明変数	パラメータ数			
	設定	0以外に推定	共通の推定値	0に推定
ln(経過年数(年))	23	23	0	0
専有面積(m ²)	23	23	3	0
最寄り駅までの徒歩時間(分)	23	23	0	0
所在階層(階)	23	23	2	0
部屋数(室)	23	23	1	0
町丁目ダミー	2,932	262	47	2,670
鉄道沿線ダミー	59	48		11
駅勢圏ダミー	474	363		111

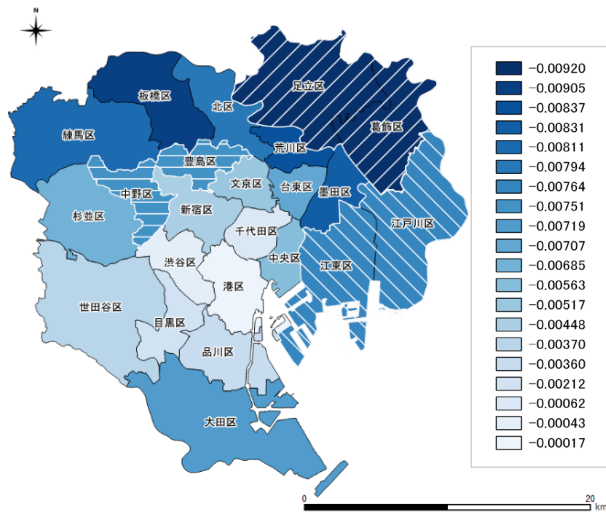


図-3 面積に対するパラメータ値の分布

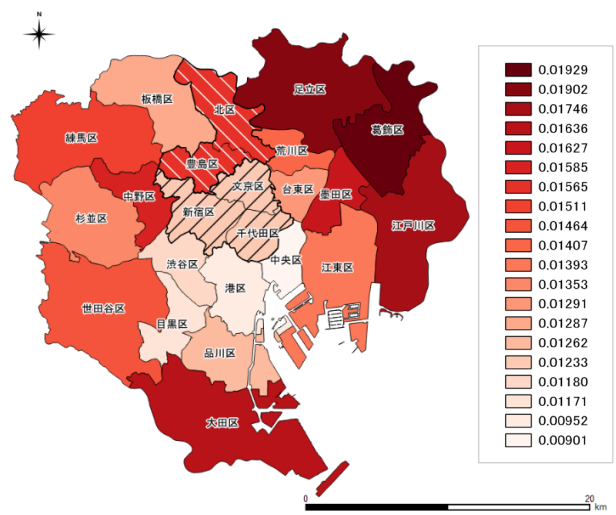


図-4 所在階層に対するパラメータ値の分布

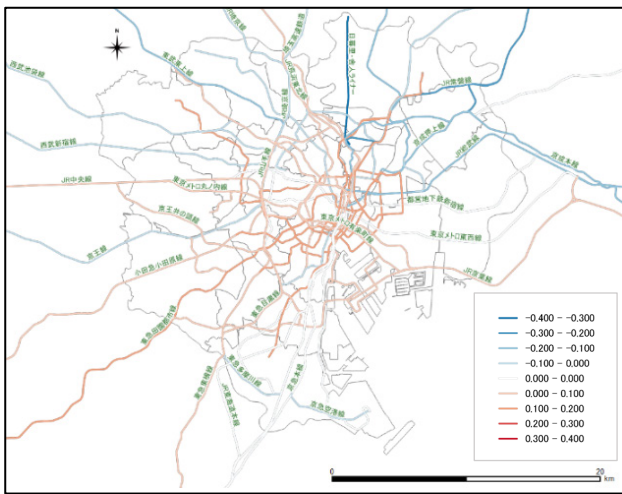


図-5 鉄道沿線ダミーに対するパラメータ値の分布

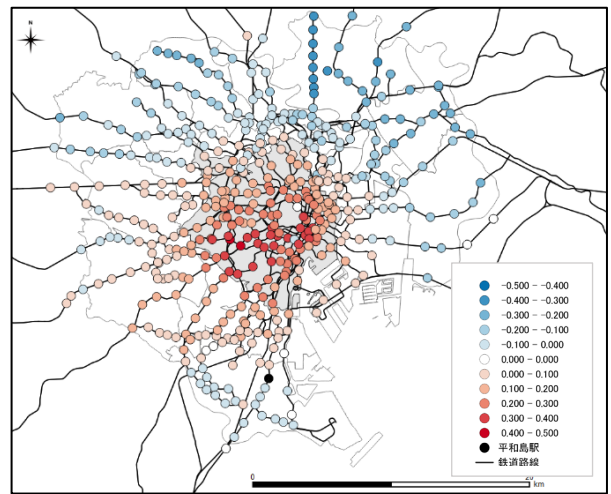


図-6 駅勢圏の総合的な評価 (駅勢圏+沿線)

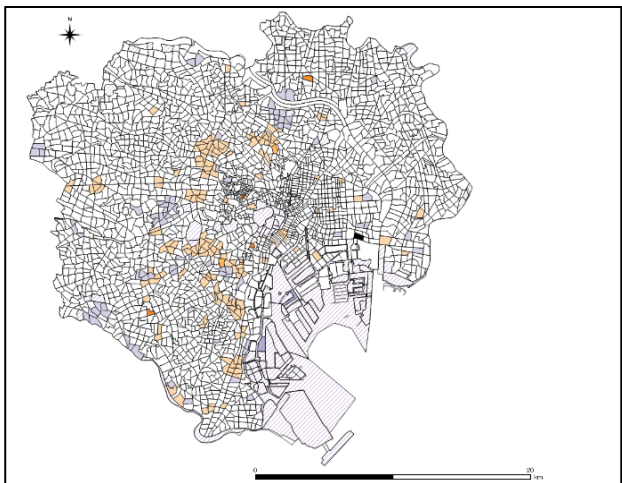


図-7 町丁目ダミーに対するパラメータ値の分布

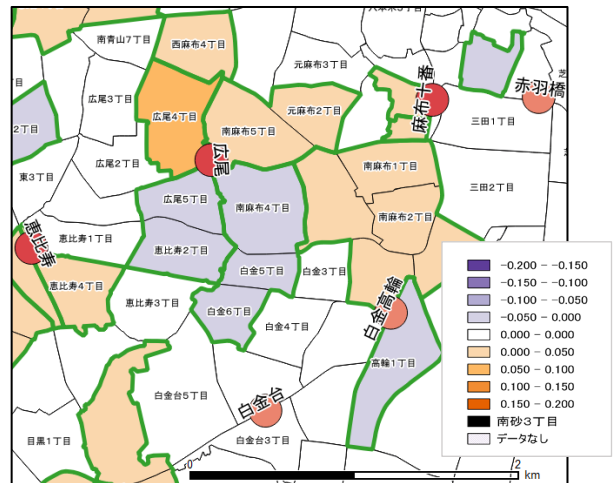


図-8 広尾・白金周辺の町丁目ダミー

(4) まとめ

東京都区部の賃貸マンションデータに対し、Generalized Fused lasso を適用した分析を行った。区と町丁目について隣接関係に基づく制約を課し、共通のパラメータ

ータを推定することで、賃料の評価が、地理的に類似あるいは異なる実態が存在し、東京都区部の不動産市場の地理的分割の把握に有効であることが確認できた。また、効率的にパラメータを 0 に推定することで、鉄道路線に

対する評価を駅勢圏に対する評価から分離することが可能であり、都心区以外については、個別の物件属性や鉄道沿線・駅勢圏ダミーで賃料水準を表現できていることが確認された。鉄道沿線ダミーのパラメータ値より、不動産市場において知られている南東部の評価の高さを、データから定量的に表現することができたとと言える。加えて、大量の賃料・価格情報を利用した分析を行うことで、交通利便性に対する評価をデータから抽出できる可能性があることが確認できた。

5. おわりに

本研究では、件数が多く形成過程に透明性が高い住宅賃料を用いて、地理的に分割された賃料形成を表現できるよう地域別のパラメータを設定した賃料モデルに対し、generalized fused lasso による変数選択を通して、賃貸住宅市場における地理的区分の把握を目的とした分析を行った。地区の隣接関係に基づく制約を課した上で、generalized fused lasso を用いて分析するアプローチによって、パラメータが共通の値や 0 に推定され、大きさの異なる様々な空間領域から影響を受け、価格形成が異なる実態を把握することができた。対象とした東京都区部において、賃料の評価が、地理的に異なる実態が存在し、一部の区で共通の賃料市場が形成されている可能性が示唆され、さらに都心部では町丁目単位の局所的な範囲で賃料水準が異なることを確認した。また、大量の賃料・価格情報を利用することで、交通利便性に対する評価をデータから抽出できる可能性があることが確認できた。

今後の課題として、本研究では正則化パラメータを AIC を用いて決定したが、決定方法によって分析結果が大きく変わる可能性があるため、さらなる検討が必要である。また、物件属性に対する評価を区毎に推定したが、実際はもっと小さい空間単位で評価が変わる可能性が考えられる。さらに、本研究では間取りや設備など物件の種別については扱っていない。時系列分析と合わせて、これらを将来の展望とする。

謝辞：公益社団法人東京都不動産鑑定士協会との共同研究を通して、アットホーム株式会社より募集賃料データを提供いただいた。また、分析結果の解釈に関して、田中敏夫氏・阿部 隆志氏・林典之氏・菊池 光貴氏より有

益なコメントをいただいた。ここに記し感謝の意を表す。

参考文献

- 1) Bourassa, S.C., Hoesli, M., and Peng, V.S.: Do housing submarket really matter? *Journal of Housing Economics*, **12**(1):12-28, 2003.
- 2) Goodman, A.C. and Thibodeau, T.G.: Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, **12**(3):181-201, 2003.
- 3) Islam, K.S. and Asami, Y.: Addressing structural instability in housing market segmentation of the used houses of Tokyo, Japan. *Procedia - Social and Behavioral Sciences*, **21**:33-42, 2011.
- 4) 黒田 翔, 堤 盛人, 奈良岡 浩二: 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. *ジャーナル・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, 2014.
- 5) Wu, C. and Sharma, R.: Housing submarket classification: The role of spatial contiguity. *Applied Geography*, **32**(2):746-756, 2012.
- 6) Montero, J.M., Fernández-Avilés, G., and Mínguez, R.: Estimating environment impacts on housing prices. *Environmetrics*, DOI: 10.1002/env.2453, 12p., 2017.
- 7) 大庭哲治, 柄谷友香, 中川大, 青山吉隆: 京町家集積の近隣外部効果に関する研究. *土木学会論文集 D*, **62**(2):227-238, 2006.
- 8) Kestens, Y., Thériault, M., and Des Rosiers, F.: Heterogeneity in hedonic modeling of house prices: Looking at buyer's household profiles. *Journal of Geographical System*, **8**(1):61-96, 2006.
- 9) 藤澤美恵子, 隅田和人: 東京大都市圏における新築マンション価格のヘドニック分析. *日本都市計画学会学術研究論文集*, **36**:943-948, 2001.
- 10) 川井隆司, 小田浩司, 枝村俊郎: ヘドニック地価関数モデルによる都市の地価構造分析. *土木計画学研究・論文集*, **9**:269-276, 1991.
- 11) 池田思朗: 計測技術におけるスパースモデリングの応用について. *Medical Imaging Technology*, **32**(3):176-181, 2014.
- 12) Tibshirani, R.: Regression shrinkage and selection via the LASSO. *Journal of Royal Statistical Society B*, **58**(1):267-288, 1996.
- 13) Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K.: Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society: Series B*, **67**(1):91-108, 2005.
- 14) Wang, H. and Rodríguez, A.: Identifying pediatric cancer clusters in Florida using log-linear models and generalized lasso penalties. *Statistics and Public Policy*, **1**(1):86-96, 2014.

(2018. 4. 27 受付)

Identification of Geographical Segmentation of Rental Housing Market in Tokyo Metropolitan Area

Ryo INOUE, Rihoko ISHIYAMA,