

# データ分布を考慮した空間的異質性分析 —固有ベクトル空間フィルタリングに基づく 可変係数分位点回帰モデル—

彭 湛<sup>1</sup>・井上 亮<sup>2</sup>

<sup>1</sup>学生会員 東北大学 大学院情報科学研究科 (〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06)

E-mail: peng.zhan.p3@dc.tohoku.ac.jp (Corresponding Author)

<sup>2</sup>正会員 東北大学准教授 大学院情報科学研究科 (〒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

空間現象の生成過程が場所により異なることに注目する空間的異質性分析は、空間分析分野で関心を集める重要な課題である。その従来手法は、説明変数が目的変数の条件付き平均に与える影響に関する空間的変動を評価することに限られ、平均以外の部分に対する説明変数の影響の違いを分析できない。そこで本研究は、固有ベクトル空間フィルタリングに基づく空間的可変係数 (ESF-SVC) モデルと条件付き分位点回帰 (CQR) を組み合わせ、目的変数の条件付き分布の複数の分位点において空間可変係数を推定する分析手法 (ESF-SVQR) を提案する。提案手法は、目的変数の様々な分位点に対する説明変数の影響の空間的変動を分析可能である。2017年の東京都7区の不動産募集賃料データに対して提案手法を適用し、従来の分析手法が評価できない異なる価格層の住宅市場の空間的異質性を分析できることを確認した。

**Key Words:** *spatial heterogeneity, eigenvector spatial filtering, quantile regression*

## 1. はじめに

空間的異質性とは、空間現象の生成過程が場所によって異なるという性質で、多くの空間現象がこの性質を有するとされる<sup>1)</sup>。通常の線形回帰や空間計量経済モデル、空間統計モデルは分析対象領域全域で共通の生成過程しか分析できないのに対し、空間的異質性分析は、共通の生成過程とは異なる生成過程を有する地域を把握できる。社会経済活動分析において、注目する空間現象の生成要因の地域差を解明することは、各地域の特徴に応じた効果的な政策決定に根拠を提供できると期待される。

空間的異質性分析手法の中でも、説明変数と目的変数との関係を空間全域で連続的に変動する回帰係数として推定する、空間的可変係数 (Spatially varying coefficients, SVC) モデルが最もよく知られている。その代表的手法である地理的加重回帰モデル (Geographically weighted regression, GWR)<sup>2)</sup>は、近傍観測値の影響の加重を用いて各地点で係数値を推定するノンパラメトリック手法である。また、ベイズ推定に基づく SVC (Bayesian SVC, BSVC) モデル<sup>3)</sup>と固有ベクトル空間フィルタリングに基

づく SVC (Eigenvector spatial filtering-based SVC, ESF-SVC) モデル<sup>4)</sup>は、空間的異質性を、全域で共通の傾向を表す生成過程からの局所的な偏差として推定するアプローチである。

これらの SVC モデルは幅広く応用されているが<sup>5,6)</sup>、いずれも条件付き平均回帰の枠組みに、説明変数の変動に対応した、目的変数の条件付き分布の平均値 (中心) の空間的変動だけに着目している。しかし、平均値は、実際のデータに存在する外れ値や歪み分布の影響を受けやすい。また、多くの分析では、関心の対象がそもそも分布の中心とは限らない。例えば、従来手法は駅開業が不動産市場全体に与える平均影響の地域差を捉えることができるが、「市場下位層や上位層の住宅の価格への影響は地域によってどう違うのか」といった、ターゲットを絞った地域経済政策にとって重要な問題に答えられない。したがって、目的変数の分布の平均以外の部分の特徴を考慮する分析手法が不可欠である。

さて、条件付き分位点回帰 (Conditional quantile regression, CQR) は、目的変数の分布の中心ではなく、複数の分位点における説明変数の影響を測る手法として注目されて

いる<sup>9)</sup>。CQR は統計学では古くから知られている最小絶対偏差 (Least Absolute Deviation, LAD) 法をもとに発展している統計的方法である。最小二乗法を用いて「平均」に焦点を当てた従来の回帰分析と比べ、CQR は平均以外の様々な水準での分析を可能にし、データ分布全体を包括的に理解することができる。さらに、CQR は外れ値や歪み分布に対するロバスト性を有する<sup>9)</sup>。

CQR の特徴を空間的異質性分析に活かすために、GWR を CQR の形式に拡張した空間的加重分位点回帰モデル (Geographically weighted quantile regression, GWQR) が提案されており、米国各郡の死亡率の水準の違いと最も関連する影響要因の特定に応用された<sup>10)</sup>。しかし、GWR を拡張した GWQR は、GWR と同じ限界を持つ。まず、各場所で推定に利用する近傍観測点の数は、繰り返し計算による交差検証で決定するため、データ数が多いと膨大な時間がかかり、推定が困難になる。さらに、GWR は複数の空間的可変係数を推定する場合、すべての係数に対して近傍観測点の数が共通であるため、異なる空間スケールで変化する係数が同時に存在しても、そのスケールを区別できない<sup>11)</sup>。BSVC モデルと CQR を結合した手法も提案されたが、この手法も計算複雑度の問題に直面している<sup>12)</sup>。一方、ESF-SVC モデルは、中心化された空間重み行列の固有ベクトルを用いることで、空間的異質性の異なるスケールを考慮できる。また、基本的な線形回帰の枠組みに基づいているため、容易に拡張して効率的に実装することが可能である。そのため、ESF-SVC モデルの活用が期待されている。

そこで、既往手法の課題を克服しながら、データ分布を考慮して空間的異質性を分析するために、本研究では、ESF-SVC モデルを CQR と組み合わせた ESF-based spatially varying quantile regression (ESF-SVQR) モデルを提案する。ESF-SVQR モデルは、目的変数の条件付き分布の異なる分位点で効率的に空間的異質性を分析することができる。また、Least absolute shrinkage and selection operator (Lasso) 正則化<sup>13)</sup>を導入し、提案手法の頑健性を向上させる。提案手法は、不動産価格データの分析に適用し、ESF-SVQR モデルの性能を従来手法と比較して評価する。

以降、第 2 章で ESF-SVC モデルと CQR を紹介し、第 3 章で提案手法とその推定方法を示す。第 4 章では実データ分析の設定と結果を示し、ESF-SVQR モデルの性能を考察する。最後に、第 5 章で本研究の結論を示す。

## 2. 既往手法

### (1) ESF-SVC モデル

ESF-SVC モデルの基本は、空間相関を検定する Moran coefficient 統計量 (MC) である。 $N$  地点で得られた観測値

ベクトル  $\mathbf{y}$  の MC は式(1)で定式化する。

$$MC(\mathbf{y}) = \frac{\mathbf{y}'\mathbf{M}\mathbf{C}\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}} \quad (1)$$

$\mathbf{C}$  は  $N \times N$  の空間重み行列、各要素  $c_{ij}$  は地点 (地域)  $i, j$  ( $i \neq j$ ) の間の距離に基づく距離減衰関数あるいは隣接関係によって決められる。 $\mathbf{M} = \mathbf{I} - \mathbf{1}\mathbf{1}'/N$  は  $N \times N$  の centering matrix,  $\mathbf{I}$  は  $N \times N$  の単位行列、 $\mathbf{1}$  は要素が 1 の  $N \times 1$  のベクトルである。中心化された空間重み行列  $\mathbf{M}\mathbf{C}\mathbf{M}$  を式(2)に従って分解して得られた固有ベクトルは、データ間の潜在的な空間的相関関係を表している。

$$\mathbf{M}\mathbf{C}\mathbf{M} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}' \quad (2)$$

$\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_N]$  は固有ベクトルの  $N \times N$  の行列、 $\mathbf{\Lambda}$  は対角要素が固有値  $\lambda_1, \dots, \lambda_N$  の  $N \times N$  の対角行列である。

各固有ベクトル  $\mathbf{e}_l$  ( $l = 1, \dots, N$ ) は独特な空間構造を、対応する固有値  $\lambda_l$  は空間スケールを表す<sup>14)</sup>。固有ベクトルの線形結合を合成変数として線形回帰モデルに導入することで、下記の ESF-SVC モデルが得られる。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{E}}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \quad (3)$$

$\mathbf{y}$  は目的変数の  $N \times 1$  のベクトル、 $\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_K]$  は定数項を含む説明変数の  $N \times (K+1)$  行列、 $\tilde{\mathbf{E}} = [\mathbf{E}, \mathbf{x}_1 \circ \mathbf{E}, \dots, \mathbf{x}_K \circ \mathbf{E}]$  は各説明変数と全ての固有ベクトルとの相互作用で作られた  $N \times (K+1)L$  行列、 $\circ$  は要素ごとの積、 $\boldsymbol{\varepsilon}$  は正規分布に従う誤差項である。回帰係数  $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_K]'$  と  $\boldsymbol{\gamma} = [\gamma_0', \dots, \gamma_K']'$ ,  $\gamma_k = [\gamma_{1,k}, \dots, \gamma_{L,k}]'$  の推定値を用い、空間的異質性を表す  $k$  番目の空間的可変係数  $\boldsymbol{\beta}_k^{ESF}$  を以下のように構築する。

$$\boldsymbol{\beta}_k^{ESF} = \beta_k \mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k \quad (4)$$

定数  $\beta_k$  は平均値、固有ベクトルの線形結合  $\mathbf{E}\boldsymbol{\gamma}_k$  は平均から乖離した局地的な空間的変動を表す。 $\boldsymbol{\gamma}_k$  は各固有ベクトルを重み付けて空間的変動のスケールを制御する。

### (2) CQR

分位点は、データ分布上の位置を、分布の下端から数えたときの累積比率を用いて示す値である。データ  $y_1, \dots, y_N$  の  $\tau$  ( $0 \leq \tau \leq 1$ ) 分位点  $q_\tau(\mathbf{y})$  は、 $\mathbf{y}$  の累積分布関数  $F_Y(\mathbf{y})$  の逆関数

$$q_\tau(\mathbf{y}) = F_Y^{-1}(\tau) = \inf\{\mathbf{y}: F_Y(\mathbf{y}) \geq \tau\} \quad (5)$$

で表す。例えば、 $y_1, \dots, y_N$  の下位から 50% までの観測値が含まれる変数の値が中央値 (0.5 分位点) となる。

$y_1, \dots, y_N$  の平均値  $\bar{y}$  は二乗損失関数  $\sum_{i=1}^N (y_i - \mu)^2$  を最小にする  $\mu$  で求められるが、 $\tau$  分位点はチェック関数

$$\rho_\tau(u) = \tau u \mathbf{1}(u > 0) + (\tau - 1)u \mathbf{1}(u \leq 0) \quad (6)$$

で評価した非対称絶対損失

$$\begin{aligned} & \sum_{i=1}^N \rho_{\tau}(y_i - a) \\ &= \sum_{i: y_i \geq a} \tau |y_i - a| + \sum_{i: y_i < a} (1 - \tau) |y_i - a| \end{aligned} \quad (7)$$

を最小にする値  $a$  として推定される。

説明変数ベクトル  $\mathbf{x}_i = [1, x_{i1}, \dots, x_{iK}]'$  が与えられた時、目的変数の条件付き分布の  $\tau$  分位点を

$$q_{\tau}(y_i | \mathbf{x}_i) = \inf \{y: F(y | \mathbf{x}_i) \geq \tau\} \quad (8)$$

とする。この時の分位点回帰は、説明変数  $\mathbf{x}_i$  の変化に応じた目的変数の条件付き分布の  $\tau$  分位点  $q_{\tau}(y_i | \mathbf{x}_i)$  の変化を説明する回帰モデルであり、式(9)と式(10)のように表現される。

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}_{\tau} + \boldsymbol{\varepsilon}_{\tau} \quad (9)$$

$$q_{\tau}(y_i | \mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta}_{\tau} \quad (10)$$

$\boldsymbol{\beta}_{\tau} = [\beta_{0,\tau}, \beta_{1,\tau}, \dots, \beta_{K,\tau}]'$  と  $\boldsymbol{\varepsilon}_{\tau}$  はそれぞれ、 $\tau$  分位点における推定係数と分布型が未知な誤差項である。

条件付き平均に関する係数を推定する通常の最小二乗法による回帰では、残差の二乗和  $\sum_{i=1}^N (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2$  を最小化する回帰係数  $\boldsymbol{\beta}$  の推定値  $\hat{\boldsymbol{\beta}}$  を求める。しかし CQR は、チェック関数で評価した残差絶対値の加重和

$$\begin{aligned} & \sum_{i=1}^N \rho_{\tau}(y_i - \mathbf{x}_i' \boldsymbol{\beta}_{\tau}) \\ &= \sum_{i: y_i \geq \mathbf{x}_i' \boldsymbol{\beta}_{\tau}} \tau |y_i - \mathbf{x}_i' \boldsymbol{\beta}_{\tau}| + \sum_{i: y_i < \mathbf{x}_i' \boldsymbol{\beta}_{\tau}} (1 - \tau) |y_i - \mathbf{x}_i' \boldsymbol{\beta}_{\tau}| \end{aligned} \quad (11)$$

を最小化することで、分位点  $\tau$  に対応した回帰係数  $\boldsymbol{\beta}_{\tau}$  の推定値  $\hat{\boldsymbol{\beta}}_{\tau}$  を得る。この推定値が有する性質は、 $\mathbf{y}$  の累積分布関数を用いると、下式で表せる。

$$F_Y(\mathbf{x}_i' \hat{\boldsymbol{\beta}}_{\tau}) = \tau \quad (12)$$

また、式(11)の最適化問題は線形計画問題に帰着することが知られている<sup>15)</sup>。分布の任意の分位点をターゲットとすることで、CQR は、外れ値や分散不均一の影響を回避しながら、頑健な分析を行うことができる。

### 3. 提案手法

#### (1) ESF-SVQR モデル

目的変数の分布の様々な分位点で空間的異質性を分析するために、ESF-SVC モデルを CQR の枠組みに拡張して下記の ESF-SVQR モデルを提案する。

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}_{\tau} + \tilde{\boldsymbol{\varepsilon}}_i' \boldsymbol{\gamma}_{\tau} + \boldsymbol{\varepsilon}_{\tau} \quad (13)$$

$\tilde{\boldsymbol{\varepsilon}}_i'$  は  $\tilde{\mathbf{E}} = [\mathbf{E}, \mathbf{x}_1 \circ \mathbf{E}, \dots, \mathbf{x}_K \circ \mathbf{E}]$  の  $i$  番目の観測値ベクトル、 $\boldsymbol{\beta}_{\tau} = [\beta_{0,\tau}, \beta_{1,\tau}, \dots, \beta_{K,\tau}]'$ 、 $\boldsymbol{\gamma}_{\tau} = [\boldsymbol{\gamma}_{0,\tau}', \boldsymbol{\gamma}_{1,\tau}', \dots, \boldsymbol{\gamma}_{K,\tau}']'$ 、 $\boldsymbol{\gamma}_{k,\tau} = [\boldsymbol{\gamma}_{1,k,\tau}, \dots, \boldsymbol{\gamma}_{L,k,\tau}]'$  は推定係数である。そして、 $\tau$  分位点における  $k$  番目の説明変数  $\mathbf{x}_k$  に対応する空間的可変係数は

$$\boldsymbol{\beta}_{k,\tau}^{SVC} = \beta_{k,\tau} \mathbf{1} + \mathbf{E} \boldsymbol{\gamma}_{k,\tau} \quad (14)$$

となる。この係数ベクトル  $\boldsymbol{\beta}_{k,\tau}^{SVC}$  の第  $j$  要素は、地点  $j$  において説明変数  $\mathbf{x}_k$  の 1 単位の変化に対応する、目的変数  $y_j$  の分布の  $\tau$  分位点の変化を表す。分位点による係数の違いを無視した通常の ESF-SVC モデルと比べ、提案モデルは、目的変数と関連要因の関係が場所や分位点によってどのように変化するかを分析できる。例えば、不動産市場の分析では、住宅価格が下位層の物件と上位層の物件のそれぞれに対して、各地域の価格決定要因を明らかにすることができる。

#### (2) 係数推定

提案モデルの係数  $\boldsymbol{\beta}_{\tau}$  と  $\boldsymbol{\gamma}_{\tau}$  は、チェック関数で評価した損失関数の最小化で推定できる。しかし、ESF-SVC モデルは、多くの固有ベクトルの導入によって過剰適合が生じる問題を持つ。そのため、本研究では、既往研究<sup>16)</sup>に基づき、 $\lambda_l / \lambda_{max} > 0.25$  を満たす固有ベクトルのみを使用し、更に Lasso 正則化で係数推定値が非ゼロとなる固有ベクトルを選択してモデルを構築する。Lasso 付き提案手法の最適化問題は式(15)で表される。

$$\min_{\boldsymbol{\beta}_{\tau}, \boldsymbol{\gamma}_{\tau}} \sum_{i=1}^N \rho_{\tau}(y_i - \mathbf{x}_i' \boldsymbol{\beta}_{\tau} - \tilde{\boldsymbol{\varepsilon}}_i' \boldsymbol{\gamma}_{\tau}) + \lambda \|\boldsymbol{\gamma}_{\tau}\|_1 \quad (15)$$

$\lambda$  は正則化の強さを決めるハイパーパラメータである。 $\lambda$  の値が増加すると、より多くの  $\boldsymbol{\gamma}_{\tau}$  の要素が 0 に推定され、より多くの固有ベクトルがモデルから除外される。各分位点での推定に対する最適な  $\lambda$  の値を、BIC を基準とする格子点探索で決定する。スパース解が BIC の評価にもたらす偏差を防ぐために、本研究では、Lasso を適用した上で、推定係数が非ゼロの変数のみを式(13)に導入し、式(16)で定義される BIC が最小になる結果を採用する。

$$\begin{aligned} BIC(\lambda) &= \log \left( \sum_{i=1}^N \rho_{\tau}(y_i - \mathbf{x}_i' \boldsymbol{\beta}_{\tau} - \tilde{\boldsymbol{\varepsilon}}_i' \boldsymbol{\gamma}_{\tau}) \right) \\ &+ df \times \frac{\log(N)}{2N} \end{aligned} \quad (16)$$

$df$  は非ゼロに推定された係数の数、 $N$  はデータ数である。

#### 4. 集合住宅賃料データ分析への適用

本章では、提案した ESF-SVQR モデルを用い、住宅市場の賃料水準毎の空間的異質性分析を行う。推定結果を、ESF-SVC による推定結果と比較し、ESF-SVQR モデルの性能を考察する。

##### (1) 分析対象とモデル

###### a) 分析対象地域と集合住宅賃料データ

本分析は、アットホーム株式会社が収集した 2017 年東京都の 7 区（千代田、中央、港、新宿、目黒、世田谷、渋谷）のマンションの募集賃料データを使用する。なお、14 階以上の階がある高層建物の物件と、旧耐震基準に従って建築された築年が 1981 年以前の物件は賃料傾向が異なるため分析対象から外し、異常値・欠損値を除いた 20,445 件を分析対象とする。分析対象を図-1 に示す。

本分析では、目的変数を単位面積当たりの賃料の対数値、説明変数を最寄り駅までの徒歩時間、専有面積、物件の所在階層、築年数それぞれの対数値とした賃料関数を定義する。なお、築年数には 0 があるため、全ての観測値に 1 を加える。説明変数をさらに平均 0、分散 1 に標準化する。以上の変数の基本統計量を表-1 に示す。

###### b) 分析モデルの設定

本分析では、指数型の距離減衰関数  $\exp(-d_{i,j}/r)$  を用いて空間重み行列を構築する。 $d_{i,j}$  は観測点  $i, j$  の間のユークリッド距離、レンジパラメータ  $r$  は全ての地点を結ぶ最小全域木の最長距離で与えた。構築した空間重み行列に基づき 200 個の固有ベクトルを近似計算<sup>18)</sup>で抽出する。そのうち  $\lambda_i/\lambda_{max} > 0.25$  を満たす 13 個の固有ベクトルを更に選択し、65 個の相互作用項  $\tilde{\mathbf{E}}$  を ESF-SVC モデルと ESF-SVQR モデルに導入した。

さらに、第 3 節で記述した通り、過剰適合を防ぐために、Lasso で ESF-SVQR モデルに用いる固有ベクトルを選択する。各分位点では、正則化項のハイパーパラメータ  $\lambda$  の値を 0 から 25 までの範囲内に 0.01 ずつ増加させ、BIC が最小になる結果を採用する。ESF-SVC モデルに対しても Lasso で  $\lambda$  の値を交差検証で決定し固有ベクトルを選択する。変数選択の結果を表-2 に示す。

##### (2) 結果と考察

本分析では、賃料単価と関連する物件属性との関係を 4 つの分位点  $\tau = [0.25, 0.50, 0.75, 0.95]$  で考察する。

###### a) 係数推定値の基本統計量

表-3 は、ESF-SVC モデルと ESF-SVQR モデルの推定値の基本統計量を示す。

ESF-SVC モデルは、所在階の係数と他の属性の係数をそれぞれ定数と空間的可変係数として推定した。また、

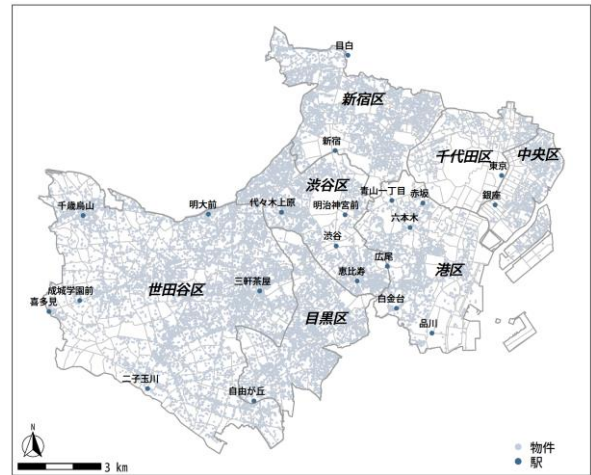


図-1 分析対象の地域と物件位置

表-1 数値属性の基本統計量

変数名	平均値	最小値	最大値	標準偏差
賃料単価 (円/㎡)	3456.31	504.62	35888.50	771.61
最寄り駅までの徒歩時間 (分)	6.66	1.00	28.00	3.850
専有面積 (㎡)	35.95	10.00	445.00	25.897
築年数 (年)	19.43	1.00	35.00	9.460
所在階 (階)	3.07	1.00	14.00	2.132

表-2 Lasso 変数選択の結果

	$\lambda$ の値	非ゼロの係数の数
ESF-SVC モデル	$1.91 \times 10^4$	37
ESF-SVQR モデル		
$\tau = 0.25$	19.39	39
$\tau = 0.50$	22.12	39
$\tau = 0.75$	20.00	40
$\tau = 0.90$	9.02	44

最寄り駅までの徒歩時間の係数について、予想外の正の推定値が存在する。一方、ほぼ全ての ESF-SVQR 推定値は各分位点で空間的に変化している。0.5 分位点だけで、所在階の係数は空間的に変化しない定数として推定された。また、分位点に関わらず、最寄り駅までの徒歩時間と築年数は賃料単価と負の相関を有する。0.5 分位点以下の価格水準では、専有面積も賃料と負の相関を持つものの、0.75 と 0.9 分位点では、専有面積が大きいほど賃料が高くなる傾向がある。以上により、各物件属性が賃料単価に与える影響は、空間的な位置や賃料単価の分位点によって異なることが明らかになる。

表-3 ESF-SVC モデルと ESF-SVQR モデルの係数推定値の基本統計量

	最小値	第 1 四分位点	中央値	第 3 四分位点	最大値	標準偏差
ESF-SVC モデル						
切片	7.898	8.054	8.114	8.175	8.402	0.105
最寄駅までの徒歩時間 (分)	-0.050	-0.0297	-0.021	-0.014	-0.002	0.011
専有面積 (m <sup>2</sup> )	-0.145	-0.107	-0.088	-0.062	0.033	0.040
築年数 (年)	-0.085	-0.072	-0.065	-0.061	-0.055	0.007
所在階 (階)	0.030	0.030	0.030	0.030	0.030	0.000
$\tau = 0.25$						
切片	7.849	7.982	8.036	8.091	8.286	0.094
最寄駅までの徒歩時間 (分)	-0.036	-0.024	-0.020	-0.016	-0.008	0.006
専有面積 (m <sup>2</sup> )	-0.135	-0.122	-0.107	-0.095	-0.035	0.023
築年数 (年)	-0.076	-0.072	-0.070	-0.067	-0.064	0.003
所在階 (階)	0.029	0.030	0.031	0.031	0.032	0.001
$\tau = 0.50$						
切片	7.914	8.052	8.109	8.168	8.367	0.097
最寄駅までの徒歩時間 (分)	-0.035	-0.026	-0.021	-0.018	-0.011	0.006
専有面積 (m <sup>2</sup> )	-0.135	-0.111	-0.096	-0.079	-0.006	0.029
築年数 (年)	-0.075	-0.068	-0.065	-0.063	-0.058	0.004
所在階 (階)	0.031	0.031	0.031	0.031	0.031	0.000
$\tau = 0.75$						
切片	7.972	8.126	8.185	8.247	8.459	0.105
最寄駅までの徒歩時間 (分)	-0.033	-0.025	-0.021	-0.017	-0.010	0.005
専有面積 (m <sup>2</sup> )	-0.127	-0.097	-0.082	-0.062	0.014	0.033
築年数 (年)	-0.064	-0.062	-0.061	-0.060	-0.058	0.001
所在階 (階)	0.030	0.031	0.031	0.031	0.031	$9.227 \times 10^5$
$\tau = 0.90$						
切片	8.031	8.202	8.260	8.331	8.571	0.116
最寄駅までの徒歩時間 (分)	-0.038	-0.023	-0.019	-0.015	-0.006	0.006
専有面積 (m <sup>2</sup> )	-0.121	-0.088	-0.065	-0.043	0.043	0.039
築年数 (年)	-0.062	-0.056	-0.055	-0.054	-0.048	0.003
所在階 (階)	0.026	0.028	0.029	0.031	0.033	0.002

## b) 推定係数の空間分布

賃料単価と属性との関係をより直観的に考察するために、ESF-SVC モデルと ESF-SVQR モデルの係数推定値の空間分布をそれぞれ図-2 に可視化して比較する。

最寄駅までの徒歩時間について、二つの手法が類似の空間分布を示し、空間全域で賃料単価に負の影響を与えると推定した。これは、家から駅に着く時間が長いほど、単位面積の賃料が安くなることを示唆している。特に小田原線しか通っていない世田谷区の南西部は、他の地域よりも公共交通機関へのアクセスが悪く、このエリアで賃貸を選ぶ人にとっては、駅の近接性がより重要な要因となるため、この属性が家賃相場に大きく影響している。また、各分位点における ESF-SVQR 推定値の空間分布も似ており、当該属性の影響が賃料単価の水準によって大

きな違いがないことが分かる。

ESF-SVC モデルでは、専有面積の係数がほとんどの地域で負に推定されているが、渋谷区と港区では符号が反転して正の値として推定されている領域が存在する。ところが、ESF-SVQR モデルの推定結果では、下位層の 0.25 と 0.5 分位点では、専有面積が賃料単価を上昇させる効果は見られず、面積の大きい物件ほど賃料単価が安くなる一般的な結果が得られた。しかし、上位層の 0.75 と 0.9 分位点では、特に渋谷区の渋谷・神宮前エリアと港区の赤坂・六本木・青山エリアでは、面積の増加によって賃料単価が高くなる傾向がある。その理由として、賃貸市場の上位に評価の異なる高級住宅が存在することが考えられる。プレミアム物件の規模が大きくなると、大型バルコニーなどの多様な設備がより充実するため、

賃料単価の上昇に寄与している。

さらに、ESF-SVC モデルと ESF-SVQR モデルとも、築年数の係数を研究対象地域の全域で負に推定している。これは築年数の増加に伴って賃料単価が安くなることを示し、新築に対する選好性が高いことを反映している。一方、賃貸市場の上位物件では、築年数が賃料単価の影響が小さくなる傾向が見られる。この結果について、市場上位の高級マンションの築年数の古さを感じさせないことが多く、付加価値をつけられた部屋も多いため、年

月が経過しても価値が落ちにくいと考えられる。

最後に、ESF-SVC モデルでは、所在階の賃料単価への影響は全域で共通だと推定されたが、ESF-SVQR モデルでは、賃貸市場の中央層（0.50 分位点）以外は、場所により異なる推定値が得られた。表-3 に示したように、各分位点での所在階の係数推定値の変動幅がもともと小さく、中央層では全域で共通の影響があるため、ESF-SVC モデルは所在階の影響を定数として推定したと推測する。しかし中央層以外では空間的異質性が存在し、

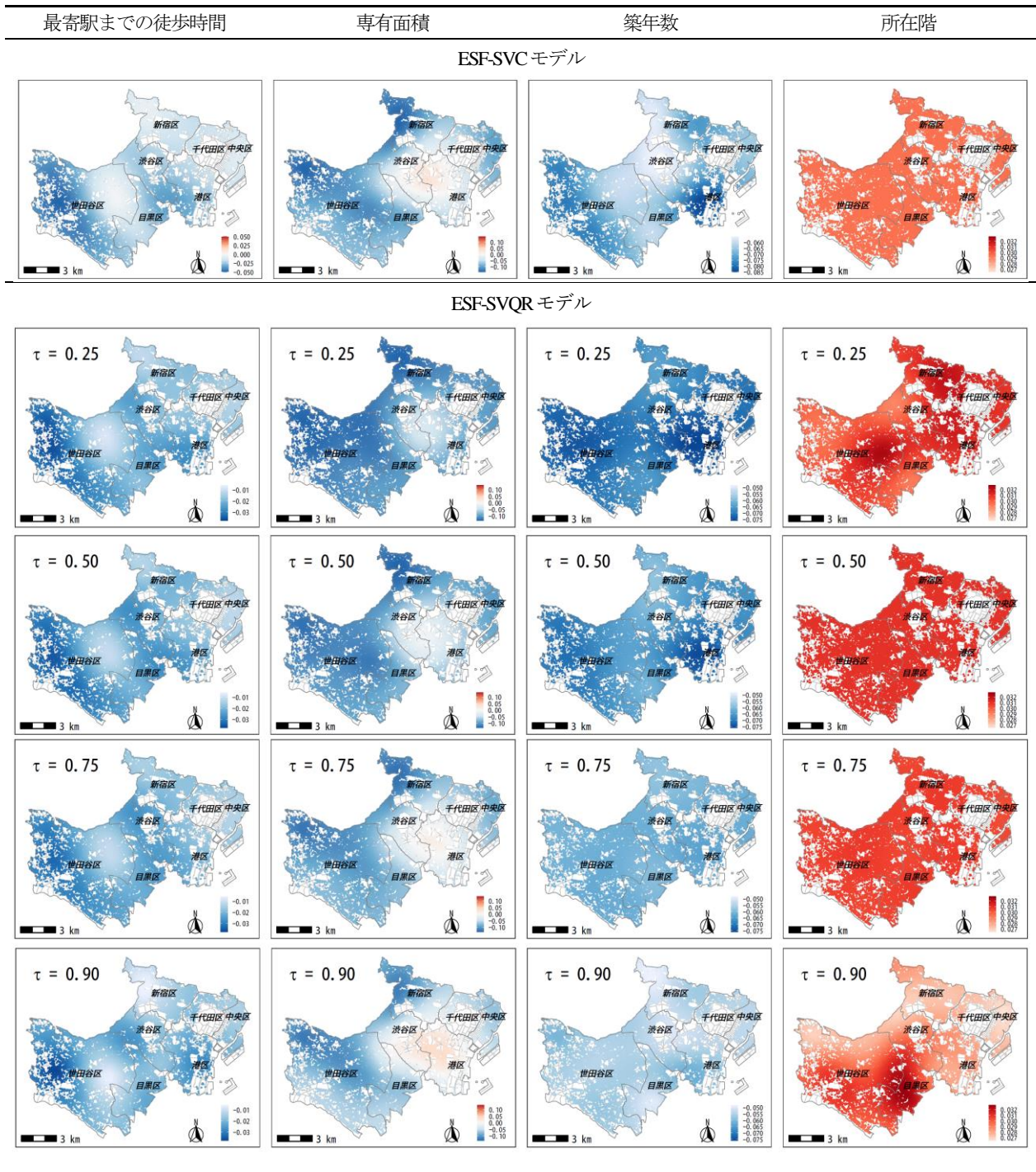


図-2 ESF-SVC モデルと ESF-SVQR モデルの係数推定値の空間分布

特に 0.9 分位点では目黒区の賃料単価が所在階から強い正の影響を受けることが明らかになった。目黒区では第一種低層住居専用地域が指定されているところが多く、中高層住宅を建設できる土地は極めて限られている。そのため、中高層階の賃料が高額になっている可能性がある。また、周辺に高層マンションがほぼないため、西側からは富士山、東側からは東京都心部の高層ビル群が望める。これらは、所在階が目黒区の賃貸市場の重要な価格決定要因になる可能な原因だと推測される。

以上のことから、物件属性が賃貸市場の平均水準の賃料に与える影響の空間的変動しか推定できない ESF-SVC モデルと比べ、提案手法である ESF-SVQR モデルはその影響が価格水準によってどのような差異があるのかをより詳しく考察できることが分かる。特に、市場上位にある高級物件はより特殊な価格評価を持つことが確認された。

## 5. 結論

本研究では、ESF-SVC モデルと CQR を組み合わせた ESF-SVQR モデルを提案し、目的変数の条件付き分布上の分位点の特徴を考慮可能な、新しい空間的異質性分析手法を構築した。

従来の条件付き平均回帰に基づく空間的異質性分析手法は、目的変数の分布の中心に関する生成要因が有する空間的異質性だけに着目しているため、分布の中心以外の部分の生成要因について見落とされている。それに対し、本研究の提案手法は、目的変数の分布全体を対象とし、複数の分位点において説明変数の空間的可変影響を包括的に把握でき、特定の市場やユーザーのグループを対象とした戦略や政策を立てる際に有用である。東京都 7 区の賃貸マンション価格データを用いた分析により、提案モデルの適用可能性が確認された。

近年、Random effect ESF-SVC (RE-ESF-SVC) モデル<sup>19)</sup>が提案され、柔軟性の点で ESF-SVC より優れていることが確認された<sup>20)</sup>ため、今後の展開として RE-ESF-SVC に基づく分位点回帰手法の開発を検討する必要がある。

**謝辞**：本研究は、JSPS 科研費 21H01447 の助成を受けた。また、本研究は、東京大学 CSIS 共同研究 (No. 815) による成果である (利用データ: 不動産データライブラリー 戸データ 全国 2013-2017 データセット (アットホーム株式会社提供))。

## REFERENCES

- 1) Goodchild, M. F.: The validity and usefulness of laws in geographic information science and geography, *Ann. Am. Assoc. Geogr.*, Vol. 94, No. 2, pp. 300-303, 2004.
- 2) Brunson, C., Fotheringham, A. S. and Charlton, M.: Geographically weighted regression, *J. R. Stat. Soc. Series D STA.*, Vol. 47, No. 3, pp. 431-443, 1998.
- 3) Gelfand, A. E., Kim, H. -J., Sirmans, C. F. and Banerjee, S.: Spatial modeling with spatially varying coefficient processes, *J. Am. Stat. Assoc.*, Vol. 98, No. 462, pp. 387-396, 2003.
- 4) Griffith, D. A.: Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR), *Environ. Plan A.*, Vol. 40, No.11, pp. 2751-2769, 2008.
- 5) Mollalo, A., Vahedi, B. and Rivera, K. M.: GIS-based spatial modeling of COVID-19 incidence rate in the continental United States, *Sci. Total Environ.*, Vol. 728, pp. 138884, 2020.
- 6) Xu, P., Huang, H., Dong, N. and Wong, S. C.: Revisiting crash spatial heterogeneity: A Bayesian spatially varying coefficients approach, *Accid. Anal. Prev.*, Vol. 98, pp. 330-337, 2017.
- 7) McCord, M. J., Davis, P. T., Haran, M. and Bidanset, P.: House price estimation using an eigenvector spatial filtering approach, *Int. J. Hous. Mark. Anal.*, Vol. 13, No.5, pp. 845-867, 2019.
- 8) Koenker, R. and Bassett, G.: Regression quantiles, *Econometrica.*, Vol. 46, No.1, pp. 33-50, 1978.
- 9) 元山 齊: Commentary 分位点回帰について: 理論と応用, 社会と調査, 第 14 号, pp. 92-97, 2015. [Hitoshi Motoyama: Commentary Quantile regression: Theory and applications, *Advances in social research*, No. 14, pp. 92-97, 2015.]
- 10) Chen, V. Y. -J., Deng, W. -S., Yang, T. -C. and Matthews, S. A.: Geographically weighted quantile regression (GWQR): An application to U.S. mortality data, *Geogr. Anal.*, Vol. 44, No. 2, pp. 134-150, 2012.
- 11) Wolf, L. J., Oshan, T. M. and Fotheringham, A. S.: Single and multiscale models of process spatial heterogeneity, *Geogr. Anal.*, Vol. 50, No. 3, pp. 223-246, 2018.
- 12) Reich, B. J., Fuentes, M. and Dunson, D. B.: Bayesian spatial quantile regression, *J. Am. Stat. Assoc.*, Vol. 106, No. 493, pp. 6-20, 2011.
- 13) Griffith, D. A.: *Spatial Autocorrelation and Spatial Filtering*, Springer Berlin, 2003.
- 14) Koenker, R. and Hallock, K. F.: Quantile regression, *J. Econ. Perspect.*, Vol. 15, No. 4, pp. 143-156, 2001.
- 15) Griffith, D. A. and Chun, Y.: *Handbook of Regional Science.*, Springer Berlin, pp. 1477-1507, 2014.
- 16) Tibshirani, R.: Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Series B. Stat. Methodol.*, Vol. 58, No. 1, pp. 267-288, 1996.
- 17) Seya, H., Murakami, D., Tsutsumi, M. and Yamagata, Y.: Application of LASSO to the eigenvector selection problem in eigenvector-based spatial filtering, *Geogr. Anal.*, Vol. 47, No.3, pp. 284-299, 2015.
- 18) Murakami, D. and Griffith, D. A.: Spatially varying coefficient modeling for large datasets: Eliminating N from spatial regressions, *Spat. Stat.*, Vol. 30, pp. 39-64, 2019.
- 19) Murakami, D., Yoshida, T., Seya, H., Griffith, D. A. and Yamagata, Y.: A Moran coefficient-based mixed effects

1) Goodchild, M. F.: The validity and usefulness of laws in

approach to investigate spatially varying relationships, *Spat. Stat.*, Vol. 19, pp. 68-89, 2017.

20) Peng, Z. and Inoue, R.: Identifying multiple scales of spa-

tial heterogeneity in housing prices based on eigenvector spatial filtering approaches, *ISPRS Int. J. Geo-Inf.*, Vol. 11, No.5, pp. 283, 2022.

## SPATIAL HETEROGENEITY ANALYSIS FROM THE PERSPECTIVE OF DATA DISTRIBUTION: AN EIGENVECTOR SPATIAL FILTERING-BASED SPATIALLY VARYING QUANTILE REGRESSION MODEL

Zhan PENG and Ryo INOUE

Spatial heterogeneity has long been an important issue in spatial analysis. Although several modeling approaches have been proposed and applied to date, they only consider spatial variation in the center of the response distribution because they are based on conditional mean regression. However, research often aims to observe how the entire distribution is affected and varies with spatial location. By combining the eigenvector spatial filtering-based SVC (ESF-SVC) model with the conditional quantile regression (CQR), we developed an ESF-based spatially varying quantile regression (ESF-SVQR) model to estimate spatially varying coefficients at multiple quantiles of the conditional distribution of the response variable. This proposed model can provide a comprehensive perspective to investigate how covariates affect different parts of the response distribution at different locations, which is crucial for designing strategies or policies targeting different markets or groups. Furthermore, an empirical application of rental price data from the Tokyo metropolitan area in 2017 suggests that the proposed model can capture local variations across quantiles that may be ignored by the conventional spatial heterogeneity modeling techniques.