

ビッグデータのためのローカルなオフィス賃料指数の開発

松尾和史*・村上大輔**・堤盛人***・今関豊和****

Developing a Local Office Rent Index for Big Data

Kazushi Matsuo*, Daisuke Murakami**, Morito Tsutsumi***, Toyokazu Imazeki****

In recent years, interest in spatial-temporal real estate price indexes has been growing, but most existing methods have low computational efficiency and are limited to the residential market. In this study, we propose a method of constructing a spatial-temporal real estate price index based on the Moran eigenvector-based spatially varying coefficient model, known for its high computational efficiency. This model is characterized by an approximation, which makes the computation time independent of sample size, and its additive model structure, which allows for flexible modeling. The proposed method is applied to rent data for the Tokyo office market from 2000 to 2021. The result shows that the proposed method is more valuable than existing methods in terms of computational efficiency and its ability to capture local changes over time.

Keywords: 不動産価格指数 (Real estate price index), 賃貸オフィス市場 (Rental office market), 時空間分析 (Spatial-temporal analysis), 空間可変係数 (Spatially varying coefficient), ビッグデータ (Big data)

1. はじめに

不動産市場の挙動を捉える場合、通常は各地域を代表する単一の不動産価格や賃料、すなわち、不動産価格（賃料）指数を定め、その変化や、地域間の値の違いを比較する (Zhu et al., 2022). この時、地域内においても立地によって不動産価格（賃料）の分布が不均一であることが知られており、このような空間的異質性をどのように捉えるかについて、様々な議論がなされている (e.g., Pace et al. 1998).

特に近年の議論では、ノンパラメトリックもしくはセミパラメトリックな手法を用いて、地理座標ごとに異なる不動産価格指数を付与する方法（時空間不動産価格指数）に関する研究に注目が寄せられている (e.g., Agarwal et al., 2021). このような時空間不動産価格指数の大きな特徴は、都市内の小地域ごとの挙動を詳細に追えることであり、複雑な不動産市場の構造の解明に寄与することである。また、地域

区分を定める必要がないため、可変単位地区問題 (MAUP: Modifiable Areal Unit Problem) も回避することができる (Clapp, 2004). 都市全体についてのグローバルな不動産価格指数を推定する場合においても、地点ごとのローカルな不動産価格指数から算出することで、集計バイアスを削減することができることが指摘されており (Malone and Redfearn, 2022), ローカルな視点で不動産市場を捉える重要性が増している。

しかし、既存の時空間不動産価格指数の推定法は、計算効率の面で課題がある。時空間グリッド上で柔軟に値を付与できる Local Regression Model (LRM, Clapp, 2004) に基づいた不動産価格指数は、不動産の建物ごとの属性を表す複数の説明変数に対して個別にカーネル回帰等のノンパラメトリック回帰を行うため、計算時間が説明変数の数やサンプルサイズに大きく依存する。

* 学生会員 筑波大学システム情報工学研究群 (University of Tsukuba)
〒305-8573 茨城県つくば市天王台 1-1-1 E-mail : matsuo.kazushi.sa@alumni.tsukuba.ac.jp

** 正会員 統計数理研究所 助教 (The Institute of Statistical Mathematics)

*** 正会員 筑波大学システム情報系 教授 (University of Tsukuba)

****非会員 株式会社オフィスビル総合研究所 (Commercial Property Research Institute, Inc.)

空間データの分析において、空間的異質性をとらえる方法論として、カーネル回帰を援用した Geographically Weighted Regression (GWR, Brunson et al., 1996; Fotheringham et al., 2002) や LRM の他に、M-SVC (Moran's eigenvector-based Spatially Varying Coefficients) model がある。同手法は、固有ベクトル空間フィルタリング (ESF: Eigenvector Spatial Filtering) を拡張したものであり、GWR に比べ係数が柔軟に定式化でき、計算時間がサンプルサイズに依存せず高速計算が可能であるという特徴を有している。また、不動産価格の予測において GWR よりも計算効率の面で優れていることが指摘されている (Murakami et al., 2017)。

本研究では、M-SVC を用いて、ビッグデータに対しても高い計算効率で推定可能な時空間不動産価格指数の推定方法を提案することを目的とする。なお、提案手法の実証分析には、東京 23 区における賃貸オフィスの賃料データを対象とし、提案手法と一般的な OLS (Ordinary Least Squares) に基づいた不動産価格指数、LRM に基づいた時空間不動産価格指数との比較を行う。

以下の各章では、第 2 章において、従来のモデルと提案モデルについて説明した後に、本研究における分析の内容について記述する。第 3 章では実証分析において扱うデータについて説明し、第 4 章において実証分析の結果を示す。そして、第 5 章において結論と課題を述べる。

2. モデルと分析内容

2.1. 一般的な価格指数の推定方法

現在、多くの国の公的な不動産価格指数の推定において用いられているヘドニックアプローチ (Rosen, 1974) は、不動産価格を不動産が有する様々な属性の束として捉える方法論である。ここで不動産の属性には、不動産の構造や規模、設備などの建物要因や、周辺環境などの立地属性、取引に関する情報や、経済状況などの情報が用いられ、一般的に式 1 のような形で定式化される。

$$\ln P_{i,t} = \beta_0 + \sum_{j=1}^J X_{j,i,t} \beta_j + \sum_{k=1}^T D_{k,i,t} \delta_k + \varepsilon_{i,t} \quad (1)$$

ここで、 $\ln P_{i,t}$ は時点 t における不動産 i の価格 (賃料) を表し、 $X_{j,i,t}$ は時点 t における不動産 i の j 番目の属性値を表す。また、 $D_{k,i,t}$ は $k = t$ のとき 1 をとる時点ダミーである。この時、点ダミーの係数 δ_k は、不動産の様々な属性による価格の差異をコントロールした上で得られる基準時点 ($t = 0$) と比較した時点 t の平均的な価格上昇率と解釈することができる。そのため、不動産価格指数 PI_t は、基準時点を 100 とし、式 2 で表される。

$$PI_t = \exp(\delta_t) \times 100, \quad \delta_0 = 0 \quad (2)$$

なお、式 1 の推定には、OLS が用いられることが多い。

但し、前述の方法論は、様々な属性の価値が時間的に不変であることや、実務上は新たな時点のデータが追加される度に、モデルの係数が変化するため、過去の指数が一定に定まらないという問題がある。そのため、公的な不動産価格指数の推定には、データを一定期間ごとにプールし、プールされた期間ごとにヘドニック回帰を行い、時点ダミーの係数を結合させることで、前述の 2 つの課題を克服する Rowling window approach が用いられている (Eurostat, 2013)。

2.2. LRM

時空間グリッド上に柔軟に価格指数を定める LRM は、パラメトリックな部分とノンパラメトリックな部分に分けて説明される。

$$\ln P_{i,t} = \sum_{j=1}^J X_{j,i,t} \beta_j + f(Z_{i,t}) + \varepsilon_{i,t} \quad (3)$$

ここで、 $\sum_{j=1}^J X_{j,i,t} \beta_j$ は OLS (式 1) と同様に、不動産の有する様々な属性の価値の束であり、パラメトリックに扱われる。それに対して、 $f(Z_{i,t})$ は時空間的に滑らかな曲面を表す 3 次元ベクトルであり、時空間価格指数を表す。

LRM の推定は大きく 3 つにステップに分けられる。

Step1.

はじめに、標準的なノンパラメトリック回帰手法で $E[\ln P_{i,t} | Z_{i,t}]$ と $E[X_{j,i,t} | Z_{i,t}]$ を推定する。ここで、ノ

ンパラメトリック回帰の手法には、スプライン (Hill and Scholz, 2018) や高次の局所多項式回帰 (Clapp, 2004), Nadaraya-Watson カーネル回帰 (Zhu, et al., 2022) などが用いられる。

Step2.

次に, Step1 で推定したノンパラメトリック回帰の残差を用いて, 式4をOLSで推定し, β_j を導出する。

$$\begin{aligned} \ln P_{i,t}^* &= \beta_0 + \sum_{j=1}^J X_{j,i,t}^* \beta_j + \varepsilon_{i,t} \\ \ln P_{i,t}^* &= \ln P_{i,t} - E[\ln P_{i,t} | Z_{i,t}] \\ X_{j,i,t}^* &= X_{j,i,t} - E[X_{j,i,t} | Z_{i,t}] \end{aligned} \quad (4)$$

Step3.

最後に, 推定された $\hat{\beta}_j$ を用いて, $\varphi_{i,t}$ を下記式で算出し, $\varphi_{i,t}$ の期待値をノンパラメトリック回帰で推定することで, $f(Z_{i,t})$ を得る (式5)。

$$\begin{aligned} \varphi_{i,t} &= \ln P_{i,t} - \sum_{j=1}^J X_{j,i,t} \hat{\beta}_j \\ f(Z_{i,t}) &= E[\varphi_{i,t} | Z_{i,t}] \end{aligned} \quad (5)$$

ここで, 式4における定数項 β_0 は $f(Z_{i,t})$ に含まれている点に注意されたい。

LRMの大きな特徴は, 一般的なノンパラメトリック回帰とOLSで推定できるため, 推定が容易であることである。但し, ノンパラメトリック回帰の推定に必要なパラメータ (バンド幅等) の決定において繰り返し処理を要することや, ノンパラメトリック回帰を各変数に対して個別に行う必要があるため, 計算効率がサンプルサイズやモデルに含める属性 (説明変数) の数に大きく依存する。

なお, 本研究では, LRM内のノンパラメトリック回帰の手法に, 統計分析ソフトRのnpパッケージで容易に実装可能なNadaraya-Watsonカーネル回帰を用いる。また, カーネル回帰に必要なバンド幅の選択には, 最小二乗法による交差検証を, カーネル関数にはガウスクーネルを用いる。

2.3. M-SVC

本研究で提案するM-SVCを用いたローカルな不動産価格指数は式6で表される。

$$\ln P_{i,t} = \sum_j^J X_{j,i,t} \beta_j + \sum_{k=1}^T D_{k,i,t} \delta_{k,i} + \quad (6)$$

$$f_{MC,0}(s_i) + \varepsilon_{it}, \quad \delta_{k,i} = b_k + f_{MC,k}(s_i)$$

ここで, $f_{MC,k}(s_i)$ はESFに基づいて得られる地点ごとに異なる係数 (空間可変係数) を表す。また, 時間ダミーの係数として得られる $\delta_{k,i}$ は基準時点 ($t=0$) と比較した時点 t の平均的な価格上昇率 b_k と各地点における平均的な価格上昇率からのズレ $f_{MC,k}(s_i)$ によって表される。

また, 空間的に可変な定数項と時間ダミーを用いることで, LRMと類似した形で時空間価格指数を導出することができる (式7)。

$$f(Z_{i,t}) = f_{MC}(s_i) + \sum_{k=2}^T D_{k,i,t} \delta_{k,i} \quad (7)$$

空間可変係数の推定に用いるESFは, 空間的自己相関の代表的な検定統計量であるMoran's I統計量 (Moran, 1948) における近接性の定義を用いて, 変数の空間的自己相関を考慮するアプローチである。

Moran's I統計量 I_{moran} はサンプルサイズを n , $n \times 1$ の変数のベクトルを \mathbf{y} , $n \times n$ の単位ベクトルを \mathbf{I} , $\mathbf{1}$ を各要素に持つ $n \times 1$ のベクトルを $\mathbf{1}$, $n \times n$ の隣接行列を \mathbf{C} として, 式8で表される。

$$I_{moran} = \frac{\mathbf{y}' \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} \right) \mathbf{C} \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} \right) \mathbf{y}}{\mathbf{y}' \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} \right) \mathbf{y}} \quad (8)$$

Moran's I統計量が, 正に有意な値を取るとき \mathbf{y} は正の自己相関を, 負に有意な値を取るとき \mathbf{y} は負の自己相関を持つとされる。また, Moran's I統計量は $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ を用いて近接性を定義した統計量とみなすことができる。

ESFでは, Moran's I統計量の近接性の定義の固有ベクトルを用いる。

$$\begin{aligned} \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_n\} &= \\ \mathit{evec}[(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)] & \quad (9) \end{aligned}$$

ここで $\mathit{evec}[\]$ は括弧内の行列の固有ベクトルを与える演算子である。 \mathbf{e}_1 は最大の固有値を持つ固有ベクトルであり, Moran's I統計量で説明可能な最も大域的な空間パターンを表す。 \mathbf{e}_2 は2番目に大きな固有値を持つ固有ベクトルであり, \mathbf{e}_1 と無関連な成分

のうち、Moran's I 統計量で説明可能な最も大域的な空間パターンを表す。同様に e_n は $n - 1$ 番目に大きな固有値を持つ固有ベクトルであり、 $n - 1$ 番目までの全ての固有ベクトルと無相関な成分のうち、Moran's I 統計量で説明可能な最も大域的な空間パターンを表す。

空間可変係数 $f_{MC,k}(s_i)$ は変数の影響に関する、正の空間相関パターンを説明する p 個の固有ベクトル $E_p = [e_1, \dots, e_p]$ と $p \times 1$ の係数ベクトル γ の線形結合によって表される。

$$f_{MC,k}(s_i) = E_p \gamma \quad (10)$$

このモデルでは、サンプルサイズ n に依存する行列（被説明変数や説明変数を含む $n \times m$ の行列）を内積計算によって $m \times m$ の行列に置き換えることで、大規模なデータに対しても、効率的に推定が可能である（Murakami and Griffith, 2019）。但し、LRM と異なり、時間方向の連続性は考慮されていない点に留意されたい。

なお、本提案手法の実装は、統計分析ソフト R の spmoran パッケージを用いることで容易に可能である。spmoran の詳細については Murakami(2017)を参照されたい。

2.4. 分析の概要

本研究では、前項までに示した3つのモデル(OLS, LRM, M-SVC)を東京23区におけるオフィス賃料のデータに適用し、パラメトリック部分の係数、計算効率、精度や誤差項の空間的自己相関に関する検定の結果、賃料指数の分布および推移について比較を行い、各モデルの特徴を整理する。

3. データ

本研究では、三幸エスレート株式会社から提供いただいたオフィス市場に関する業務データに対して、前章で示した3つのモデルを適用する。業務データには、2000年から2021年までの各年12月末時点における計128,720件の標準的な賃貸オフィスビルの位置情報や貸付床面積、地上階数、築年数等の基本的な情報や、募集賃料、空室面積等の賃貸借取引に関する情報が含まれている。

ここで、標準的な賃貸オフィスビルとは事務用途に限定されたものであり、自社保有のオフィスビルや、倉庫、店舗用途であるもの等はデータから除かれている。なお、オフィスビルのデータは、三幸エスレート株式会社が把握しているものに限られている点に留意されたい。

表-1 変数の概要と記述統計

変数	単位	概要	最小	最大	平均	標準偏差
募集賃料	対数・円/坪	1坪当たりの共益費込み募集賃料	8.24	11.08	9.58	0.34
地上階数	対数・階	建物の地上階数	1.10	4.09	2.02	0.35
建物属性	1フロア面積	1フロアあたりの床面積	1.10	8.16	3.82	0.90
	築年数	竣工からの経過年数	0.00	94.00	24.93	11.66
	建物内空室率	%	0.00	1.00	0.23	0.26
	冷房	ダミー	0.00	1.00	0.98	0.16
地理的属性	CBD	対数・m	5.33	9.81	8.06	0.75
	周辺貸付面積	対数・坪	2.48	13.45	11.43	1.24
	周辺空室率	%	0.00	1.00	0.06	0.04
	駅徒歩	対数・分	0.69	3.43	1.40	0.49
	建物密集度	対数・m	3.10	8.34	4.36	0.63
	周辺高さ	対数・階	1.10	3.29	1.93	0.14

なお、データの網羅率については、賃貸オフィスビルに関する公的統計が存在しないため、正確には把握することはできない。但し、本研究で扱うデータは他の民間企業が公開しているデータと比較し、同程度のオフィスビルを網羅していることが確認されている（松尾ら、2022）。

賃料に関する情報は、1棟ごとの標準的な募集賃料を用いる。これは、オフィスビル内における複数の募集情報のうち、店舗用途やシェアオフィスといった特殊事例を取り除き、標準的な事務所用途として募集された共益費込みの1坪当たりの賃料情報をそのオフィスビルの募集賃料の代表値として用いたものである。

LRMやM-SVCの推定では、地理的な変数の影響は全てノンパラメトリックな部分に含まれるため、表-1における建物属性を表す5つの変数を説明変数 $X_{j,it}$ として用いる。OLSについては、前述の5つの変数に加え、地理的な属性を表す変数や、時点ダミー、地域ダミーを説明変数として加え、推定を行う。

4. 実証分析

4.1. 各モデルの推定結果の概要

約13万件に及ぶサンプルを全て用いた推定は計算効率が低いため、ここでは無作為に抽出した2万サンプルを用いて、各モデルの推定を行った。表-2は各モデルにおける説明変数の係数と決定係数を示したものである。モデル間で説明変数の有意性に差異はあるものの、係数の符号とその大きさはおおよそ一致していることが読み取れる。なお、LRMにおけるバンド幅は緯度方向が0.000793度、経度方向が0.000706度、時間方向が3.53年と選択された。以下の分析では、地理的属性を加えた表-2の列(2)のOLSと、列(3)、(4)に示されているLRM、M-SVCの3つのモデルの比較を行う。

4.2. 賃料指数の推移と分布

次に、2000年における賃料を100とした賃料指数の推移を図-1に示す。LRMとM-SVCは各時点におけるセミパラメトリックな部分の推定値の平均値を平均的な賃料の変動値として、指数を算出した。指数の推移は、OLSとM-SVCが非常に類似した傾向

表-2 変数の概要と記述統計

model	(1)			(2)			(3)			(4)		
	OLS			OLS			LRM			M-SVC		
	coeff	t-value		coeff	t-value		coeff	t-value		coeff	t-value	
<i>(Intercept)</i>	8.747	539.63	***	8.477	97.54	***						
1フロア面積	0.113	57.30	***	0.119	59.08	***	0.149	33.80	***	0.112	61.44	***
地上階数	0.182	34.79	***	0.154	29.44	***	0.184	14.97	***	0.177	36.57	***
築年数	-0.007	-49.89	***	-0.008	-55.32	***	-0.011	-56.20	***	-0.008	-60.24	***
建物内空室率	-0.015	-2.46	**	-0.005	-0.81		-0.049	-11.22	***	-0.009	-1.60	
冷房	-0.080	-8.29	***	-0.076	-8.15	***	-0.007	-0.53		-0.071	-8.01	***
駅徒歩				-0.093	-29.08	***						
周辺貸付面積				0.049	12.45	***						
周辺空室率				-0.514	-10.24	***						
建物密集度				0.005	1.01							
周辺高さ				-0.067	-3.76	***						
CBD				0.014	1.94	*						
<i>Year dummies</i>		Yes			Yes			No			No	
<i>District dummies</i>		Yes			Yes			No			No	
<i>Observations</i>		20,000			20,000			20,000			20,000	
<i>Adjusted-R²</i>		0.628			0.654			0.819			0.702	

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

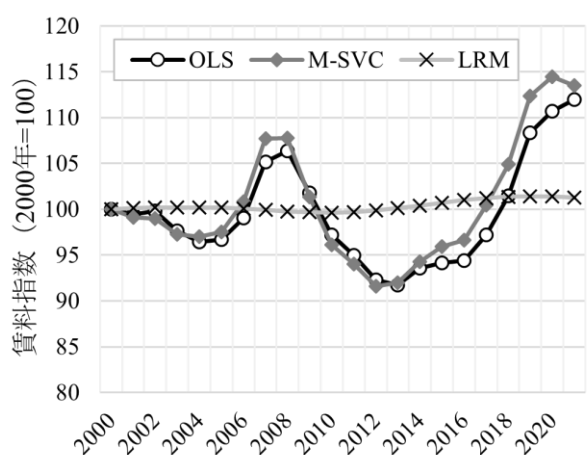


図-1 賃料指数の推移

を示している一方で、LRMは時点間の変動が小さく、極めて滑らかな推移を示していることが読み取れる。LRMは、時間方向の連続性を考慮しているため、時間方向のバンド幅が大きい場合、時点による局所的な変化を捉えにくいことが考えられる。

次に、指数の空間分布を確認するために、指数推移のピーク（2008年、2020年）と底（2004年、2012年）における250mメッシュごとの指数を算出し、可視化した（図-2）。ここで、オフィスビルが立地し

ない地域を対象地域から除外するため、メッシュの中心点を基準として半径250m圏内にオフィスビルが立地するメッシュのみを対象とした。また、賃料指数の基準値は、図-1と同様に、2000年における賃料分布の平均値を100としている。LRMとM-SVCの分布を比較すると、LRMの方が、空間的に局所的な変動を捉えていることが読み取れる。これは、緯度・経度方向におけるバンド幅が短いことに起因していると考えられる。

4.3. 各種指標の比較

次に、4つの指標を用いて、モデルの特徴を整理する。

4.3.1. 推定精度

はじめに、予測精度の比較結果を示す。本研究では、予測精度を示す指標として、絶対平均誤差（MAE: Mean Absolute Error, 式11）と対数平均二乗誤差（RMSLE: Root Mean Squared Logarithmic Error, 式12）を用いる。

$$MAE = \frac{\sum |P_{i,t} - \hat{P}_{i,t}|}{n} \quad (11)$$

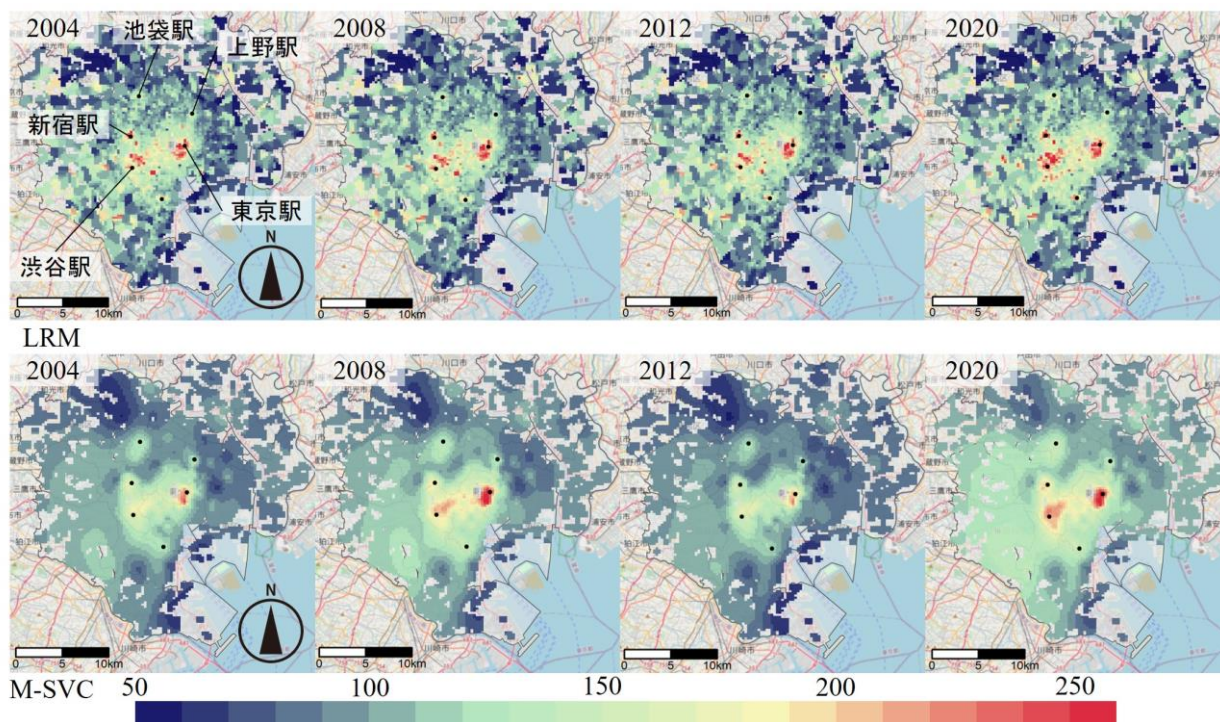


図-2 賃料指数の空間分布（上：LRM，下：M-SVC）

2000年における賃料分布の平均値を100として、指数を算出。

表-3 推定精度

		MAE	RMSLE
in-sample	OLS	2385.35	0.203
	LRM	1684.01	0.147
	M-SVC	2179.26	0.189
out-of-sample	OLS	2393.68	0.204
	LRM	2199.87	0.190
	M-SVC	2334.57	0.200

$$RMSLE = \sqrt{\frac{\sum (\ln P_{i,t} - \ln \hat{P}_{i,t})^2}{n}} \quad (12)$$

精度の算出には、モデル推定に用いた2万件のサンプル (in-sample) とモデル推定に用いていない約11万件のサンプル (out-of-sample) の両方を用いた。

その結果、いずれも LRM, M-SVC, OLS の順に精度が高いことが明らかになった (表-3)。M-SVC は近似を用いており、また時間方向の連続性を考慮していないため、全サンプルを用いて推定する LRM に精度の面で劣ることは妥当な結果である。

4.3.2. 誤差項の空間的自己相関

次に、各モデルがオフィス賃料の空間的な構造をどの程度捉えているのかを検証するために、Moran's I 統計量を用いた誤差項に対する空間的自己相関の検定を行った (図-3)。その結果、OLS の誤差項は、すべての時点において 5%水準で有意に空間的自己相関がある一方で、M-SVC は一部の時点で、LRM は

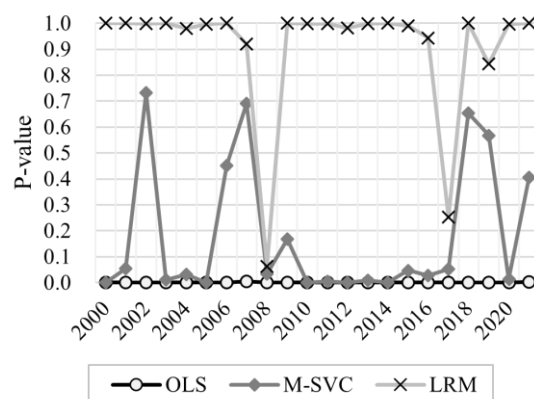


図-3 誤差項の空間的自己相関の推移

全ての時点で、空間的自己相関がないという帰無仮説が5%水準で棄却されないという結果が示された。この結果は、M-SVC や LRM は OLS では捉えていない賃料の空間的異質性を補足していることを示唆している。但し、M-SVC が捉えている賃料の空間的異質性は部分的であり、LRM には及ばないことも読み取れる。

4.3.3. J-test

各モデルの説明力を比較するために非ネスト型の J-test (Greene, 2017) を行う。これは、一方のモデルが正しければ、他方のモデルは有意な説明力を有さないという仮説を検定するものであり、一方のモデルに他方のモデルで得られた予測値を加えて推定する

表-4 J-test の結果

Model	(1) OLS	(2) OLS	(3) LRM	(4) LRM	(5) M-SVC	(6) M-SVC
<i>Predicted values from OLS</i>			-0.012 (-1.38)			0.235 (16.33)***
<i>Predicted values from LRM</i>	1.146 (151.81)***				1.074 (120.28)***	
<i>Predicted values from SVC</i>		0.975 (65.75)***		-0.019 (-1.93)*		
<i>Nonparametric portion of LRM</i>	No	No	Yes	Yes	No	No
<i>Nonparametric portion of M-SVC</i>	No	No	No	No	Yes	Yes
<i>Structural variables</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Location variables</i>	Yes	Yes	No	No	No	No
<i>District dummies</i>	Yes	Yes	No	No	No	No
<i>Year dummies</i>	Yes	Yes	No	No	No	No
<i>Observations</i>	20,000	20,000	20,000	20,000	20,000	20,000
<i>Adjusted-R²</i>	0.840	0.716	0.828	0.828	0.828	0.707

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. 括弧内は t 値を表す。

ことで検定することができる。

表-4 は OLS, LRM, M-SVC の各モデルに他の 2 つのモデルの予測値を加え、推定した結果を示している。列(1), (2)の結果は OLS のモデル式に, LRM や M-SVC の予測値が有益な情報を付与することを示している。それに対して, 列(3), (4)の結果は, LRM に対して, OLS も M-SVC も有益な情報はほとんど付与しないことを示している。また, 列(5), (6)の結果は, M-SVC に対して, OLS と LRM のいずれも有益な情報を付与するが, その係数や t 値の値に着目すると, OLS に比べ, LRM の値が極めて大きく, 付与する情報の量に大きな差があることが読み取れる。ここで, LRM が他のモデルに付与する情報は, 詳細な空間的構造であると考えれば, これらの結果は前述の誤差項の空間的自己相関の検定の結果と整合的であると考えられる。

4.4.4. 計算効率の比較

ここで, サンプルサイズや, 対象期間の長短によって, 計算効率がどのように変化するかを確認する。サンプルサイズごとの比較では, サンプルサイズを 5000 件から 20000 件まで, 対象期間による比較では, サンプルサイズを一定 (20000 件) に固定し, 対象期間のみ 11 時点 (2000 年~2010 年) から 22 時点 (2000 年~2021 年) まで変化させて, 各パターンにおける計算時間を計測した (図-4)。なお, 本研究では DELL Precision (2.20GHz, 10-Core Intel Xeon Silver 4210 with 64GB memory)を使用する。また, モデルの推定には R(version; 4.0.3)を用いており, LRM の推定には np(version; 0.60-11), M-SVC の推定には spmoran(version; 0.2.2.4)を使用する。

その結果, LRM はサンプルサイズが増えると計算時間が線形的に増加するが, M-SVC は計算時間がサンプルサイズに依存せず一定の水準であることが明らかになった。特に 2 万サンプルの場合, M-SVC の計算に要した時間は約 5.4 万秒 (0.63 日) であるのに対して, LRM の計算に要した時間は約 27 万秒 (3.13 日) であり, 両者には 5 倍程度の差があることが読み取れる。但し, 空間可変係数で捉える変数の数に対して, サンプルサイズが少ない場合, M-SVC では

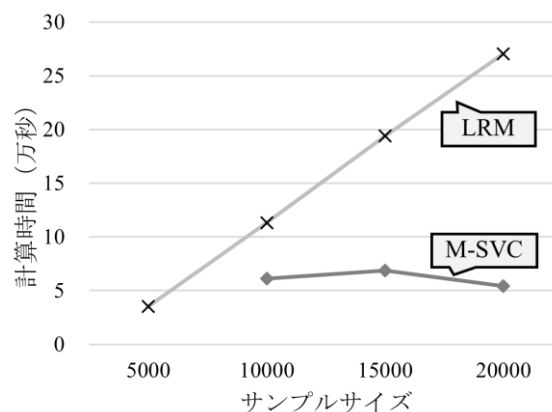


図-4 サンプルサイズによる計算時間の差異

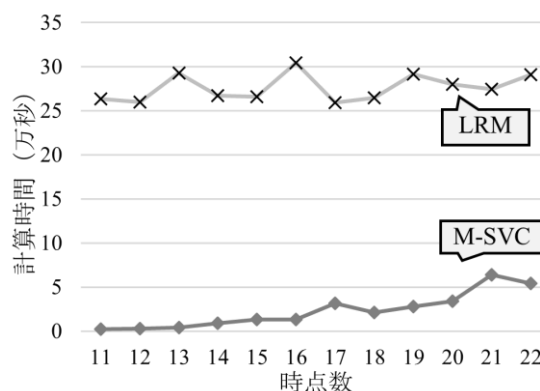


図-5 対象期間による計算時間の差異

計算できないことも明らかになった。また, サンプルサイズを 2 万に固定し, 対象期間のみを変化させた場合の計算時間の差異に着目すると, LRM は対象期間の長短に依存せず, 計算時間が一定である一方で, M-SVC は対象期間が長いほど, すなわち, 空間可変係数として推定する必要がある時間ダミーが多いほど, 計算時間が指数関数的に増加することが読み取れる (図-5)。但し, 対象期間がより長期に及ばない限り, LRM に比べ M-SVC の計算時間が短くなることが推察される。

これらの結果は, サンプルサイズが大きいほど, LRM と M-SVC の計算効率の差は広がり, M-SVC の方が効率的に推定可能であることを示している。但し, 時点ごとのサンプルサイズが小さく, 対象期間が長い場合には, LRM の方が効率的になることも示唆しており, データの特徴によって有効的なモデルが異なることを示している。

5. 結論と課題

本研究では、M-SVC を用いて、サンプルサイズに依存せず、柔軟に推定可能なローカルな不動産価格指数の構築方法を提案し、東京 23 区におけるオフィス賃料データに適用することで、既存手法との特徴の差異を明らかにした。モデルの予測精度と局所的な空間構造の把握という点では、M-SVC は LRM に劣るものの、時間方向の局所的変化の捕捉やビッグデータに対する計算効率といった面で、LRM より優れていることが示唆された。また、M-SVC は加法モデルに基づいており、非線形効果を含めたモデルへの改良等 (Murakami and Griffith, 2020) も可能であり、より柔軟にモデルを定めることができるという点でも有益である。

本研究の貢献は、ローカルな価格指数の実用性と発展の可能性を示した点にある。地域性を有する不動産市場において、行政区域等の特定の空間単位に縛られず、情報の把握が可能なローカルな価格指数は、不動産市場の空間的な構造とその変遷を理解するための有益な情報提供の一助になることが考えられる。また、本研究はオフィス市場を対象にローカルな賃料指数の推定を試みた最初の論文である。従来、オフィス市場は住宅市場に比べ、データの量と質の欠如していることが大きな課題とされている (Balemi, et al., 2021; 吉田・川井, 2020)。本研究の成果は東京オフィス市場への理解を深めるための貴重な基礎資料となることも期待される。

ただし、本研究にはいくつかの課題が残されている。一つ目は、時点数の増加に伴う計算時間の増大に関する問題である。本研究では年次データを用いているが、実用上は月次データを用いることが好ましい。しかし、時点数が増えると、いずれのモデルも計算効率が低くなり、実用性に欠けることとなる。そのため、今後は時点数の増加に対しても、計算効率を高められる方法を開発することが求められる。

二つ目は、LRM の計算効率の向上である。本研究では、最も容易に実装可能なパッケージを用いて、LRM を推定した。しかし、近年では高い計算効率でカーネル推定を行う方法論も多数研究されている (e.g., Murakami et al., 2020)。M-SVC と LRM は互い

に異なる特徴を有するモデルであり、片方のみを改善することは必ずしも適切ではない。今後の研究では、M-SVC に基づくローカルな不動産価格指数の改善だけでなく、LRM に基づくローカルな不動産価格指数の改善についても取り組んでいくことが重要である。

謝辞

本研究は JST 次世代研究者挑戦的研究プログラム JPMJS2124 の助成、及び一般社団法人三幸オフィス基金からの寄附を受けて実施した成果の一部である。ここに記して謝意を表する。

参考文献

- 三幸エステート株式会社 (2022) 2022 オフィスレントデータ. <https://www.sanko-e.co.jp/data/rentdata/rentdata_2022/>.
- 松尾和史・堤盛人・今関豊和 (2022) 空室率データで見える東京オフィス市場の空間的特性. 「GIS—理論と応用」, 30 (1), 51-63.
- 吉田二郎・川井康平 (2020) 商業用不動産に関するデータ資源：経年減価率推計を例に見る現状と課題. 「日本不動産学会誌」, 33(4), 93-97.
- Agarwal, S., Fan, Y., McMillen, D. P. and Sing, T. F. (2021) Tracking the pulse of a city—3D real estate price heat maps. *Journal of Regional Science*, 61(3), 543-569.
- Balemi, N., Füss, R. and Weigand, A. (2021) COVID-19's impact on real estate markets: review and outlook. *Financial Markets and Portfolio Management*, 35(4), 495-513.
- Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281-298.
- Clapp, J. M. (2004) A semiparametric method for estimating local house price indices. *Real Estate Economics*, 32(1), 127-160.
- Eurostat. (2013) Handbook on Residential Property Prices Indices. Available at: <https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF>
- Fotheringham A. S, Brunsdon, C. and Charlton, M. (2002)

- Geographically weighted regression: the analysis of spatially varying relationships*. Chichester, UK: John Wiley & Sons.
- Greene, W. H. (2017) *Econometric Analysis (8th Edition)*. Pearson.
- Hill, R. J. and Scholz, M. (2018) Can geospatial data improve house price indexes? A hedonic imputation approach with splines. *Review of Income and Wealth*, 64(4), 737-756.
- Malone, T. and Redfearn, C. L. (2022) To measure globally, aggregate locally: Urban land, submarkets, and biased estimates of the housing stock. *Real Estate Economics*, 50(3), 656-671.
- Moran, P. A. P. (1948) The interpretation of statistical maps. *Journal of The Royal Statistical Society, Series B(Methodological)*, 10(2), 243-251.
- Murakami, D. and Griffith, D. A. (2019) Spatially varying coefficient modeling for large datasets: Eliminating N from spatial regressions. *Spatial Statistics*, 30, 39-64.
- Murakami, D. and Griffith, D. A. (2020) Balancing Spatial and Non - Spatial Variation in Varying Coefficient Modeling: A Remedy for Spurious Correlation. *Geographical Analysis*.
- Murakami, D., Tsutsumida, N., Yoshida, T., Nakaya, T. and Lu, B. (2020) Scalable GWR: A linear-time algorithm for large-scale geographically weighted regression with polynomial kernels. *Annals of the American Association of Geographers*, 111(2), 459-480.
- Murakami, D., Yoshida, T., Seya, H., Griffith, D. A. and Yamagata, Y. (2017) A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*, 19, 68-89.
- Murakami, D. (2017) spmoran: An R package for Moran's eigenvector-based spatial regression analysis. arXiv preprint arXiv:1703.04467.
- Pace, R. K., Barry, R., Clapp, J. M. and Rodriguez, M. (1998) Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics*, 17(1), 15-33.
- Zhu, E., Wu, J., Liu, H. and Li, X. (2022) Within - city spatial distribution, heterogeneity and diffusion of house price: Evidence from a spatiotemporal index for Beijing. *Real Estate Economics*, 50(3), 621-655.