## Title
Specifying multi-scale spatial heterogeneity in the rental housing market: The case of the Tokyo metropolitan area



**Authors**
Peng, Zhan
Inoue, Ryo

**Publication Date**
2021-09-01

**DOI**
10.25436/E2201T

Peer reviewed

# Specifying multi-scale spatial heterogeneity in the rental housing market: The case of the Tokyo metropolitan area

**Zhan Peng**
Graduate School of Information Sciences, Tohoku University, 6-6-06 Aramaki-Aoba, Aoba, Sendai, Miyagi 980-8579, Japan
peng.zhan.p3@dc.tohoku.ac.jp

**Ryo Inoue**
Graduate School of Information Sciences, Tohoku University, 6-6-06 Aramaki-Aoba, Aoba, Sendai, Miyagi 980-8579, Japan
rinoue@tohoku.ac.jp

**Abstract**

The urban real estate market is shaped by spatially varying environmental and social determinants, such as the valuation of green spaces, proximity to transport, and distance to central business districts. Among all the spatially varying relationships between prices and housing characteristics, some tend to vary at a global spatial scale, whereas others vary at a local spatial scale. This study applies a random model to specify multi-scale spatial heterogeneity in the rental housing market by utilizing residential rent data in the Tokyo metropolitan area from 2017. The results show that spatially varying determinants impact rental housing prices at the global, moderate, and local scales. Further, we find that the estimation is flexible because the random model determines the spatial scale of each regression coefficient.

**Funding** This research was supported by JSPS KAKENHI Grant Number JP18H01552 and JP21H01447.

**Acknowledgements** This research was the result of the joint research with CSIS, the University of Tokyo (No. 815), and used apartment rental data provided by At Home Co, Ltd. of Japan.

## 1 Introduction

The urban real estate market is a critical indicator of the local economy and social and civic development. Housing prices are shaped by structural housing features, including size, floor level, and age of the building, as well as locational and neighborhood determinants, such as the valuation of green spaces, proximity to transport, and distance to the city center.

The hedonic price model is typically used in economics to investigate and quantify the impact of such factors on housing prices. The traditional hedonic price model assumes that the impact remains unchanged throughout the study area and makes a global estimation based on a linear regression function. However, the relationship between prices and determinants may vary across regions according to residents' preferences. Furthermore, due to the uneven distribution of public resources, these inconsistent relationships tend to vary at different spatial scales. Ignoring such multi-scale spatial heterogeneity in hedonic models may lead to biased estimations and inappropriate conclusions.

Spatially varying coefficient (SVC) models have been widely applied to analyze spatial heterogeneity. Although some representative SVC models, such as multiscale geographical weighted regression [1] and Bayesian-SVC model [2], can identify the spatial scale of each relationship, their computational costs increase rapidly when the sample size increases.

Recently, the eigenvector spatially filtering-based SVC (ESF-SVC) model has attracted attention for its ability to efficiently extract multi-scale spatial heterogeneity (see Helbich and Griffith, 2016 [3]). Furthermore, Murakami et al.(2017) [4] extended it to a random effects ESF-SVC (RE-ESF-SVC) and showed that RE-ESF-SVC could increase the estimation accuracy. Considering the potential multi-scale spatial heterogeneity in the urban housing market, wherein large amounts of data are recorded, the application of RE-ESF-SVC is appropriate, and new insights into the characteristics of the market are expected. Thus, the purpose of this study is to implement the RE-ESF-SVC model to specify the multi-scale spatial heterogeneity of the urban real estate market, utilizing the residential rent data in the Tokyo metropolitan area in 2017, which contains more than 70,000 records.

# 2 Method

## 2.1 ESF-SVC model

The basis of ESF-SVC is eigenvector spatial filtering (ESF), a spatial modeling approach proposed by Griffith (1996) [5]. This approach is interpretable based on Moran coefficients ($MC$), a spatial dependence diagnostic statistic formulated as follows:

$$MC = \frac{\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N)\boldsymbol{C}(\boldsymbol{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N)\boldsymbol{z}}{\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N)\boldsymbol{z}} \tag{1}$$

where $\boldsymbol{C}$ is an $N \times N$ symmetric spatial proximity matrix between $N$ locations, $\boldsymbol{z}$ is an $N \times 1$ vector of observations, $\boldsymbol{M} = \boldsymbol{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N$ is an $N \times N$ centering matrix, $\boldsymbol{I}$ is an $N \times N$ identity matrix, and $\mathbf{1}$ is an $N \times 1$ vector of ones. $MC$ is positive if there is a positive spatial dependence in $\boldsymbol{z}$ and negative otherwise. Following Griffith (2003) [6], each eigenvector in $\boldsymbol{E} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_L](L < N)$ decomposed from the matrix $\boldsymbol{MCM}$ shows a distinct pattern of underlying spatial dependence with magnitude determined by the corresponding eigenvalue in $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_L]$.

Eigenvectors are utilized as synthetic variables and interact with explanatory variables to capture spatial heterogeneity, yielding the following ESF-SVC model [7]:

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{x}_k \circ \boldsymbol{\beta}_k^{\mathrm{ESF}} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \sigma^2\boldsymbol{I}), \tag{2}$$

$$\boldsymbol{\beta}_k^{\mathrm{ESF}} = \beta_k\mathbf{1} + \boldsymbol{E}\boldsymbol{\gamma}_k \tag{3}$$

where $\boldsymbol{y}$ is an $N \times 1$ vector of response variable, "$\circ$" denotes the column-wise product operator, $\boldsymbol{\beta}_k^{\mathrm{ESF}}$ is the vector of the spatially varying coefficients of the $k$th$(k = 1, \cdots, K)$ explanatory variable $\boldsymbol{x}_k$, $\boldsymbol{E}$ is an $N \times L$ matrix composed of $L$ eigenvectors $(L < N)$, and $\beta_k$ and $\boldsymbol{E}\boldsymbol{\gamma}_k$ refer to the fixed constant and spatially varying component, respectively.

## 2.2 RE-ESF-SVC model

Because ESF-SVC includes interaction terms between eigenvectors and predictors, utilizing many eigenvectors without selection will cause severe overfitting. Instead of using the stepwise eigenvector selection method, which can be slow for a large sample set, eigenvectors whose corresponding eigenvalues satisfy a criterion can be selected. However, eigenvectors that represent the necessary local spatial patterns can be excluded.

The RE-ESF-SVC model, assuming $\boldsymbol{\gamma}_k$ as random effects [4], is an extension of the classic ESF-SVC model to address the above limitations. This is formulated as follows:

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{x}_k \circ \boldsymbol{\beta}_k^{\text{R-ESF}} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \sigma^2 \boldsymbol{I}) \tag{4}$$

$$\boldsymbol{\beta}_k^{\text{R-ESF}} = \beta_k \mathbf{1} + \boldsymbol{E}\boldsymbol{\gamma}_k, \quad \boldsymbol{\gamma}_k \sim \mathcal{N}(\mathbf{0}_L,\ \sigma^2_{k(\gamma)} \boldsymbol{\Lambda}(\alpha_k)) \tag{5}$$

where $\boldsymbol{E}$ is the subset of $L$ eigenvectors corresponding to positive eigenvalues without selection, $\mathbf{0}_L$ is an $L \times 1$ vector of zeros, $\boldsymbol{\Lambda}(\alpha_k)$ is an $L \times L$ diagonal matrix whose $l$th ($l = 1, \cdots, L$) element is $\lambda_l(\alpha_k) = (\sum_l \lambda_l / \sum_l \lambda_l^{\alpha_k})\lambda_l^{\alpha_k}$,where $\lambda_l$ is the corresponding eigenvalue of the $l$th eigenvector $\boldsymbol{e}_l$. $\sigma^2_{k(\gamma)}$ and $\alpha_k$ are the unknown variance parameters of the random effects $\boldsymbol{\gamma}_k$.

Note that $\alpha_k$ controls the scale of the spatial dependence. A large $\alpha_k$ shrinks the coefficients of the eigenvectors with small eigenvalues toward zero; thus, the resulting $\boldsymbol{E}\boldsymbol{\gamma}_k$ of $\boldsymbol{\beta}_k^{\text{R-ESF}}$ describes a global map pattern. In contrast, a small $\alpha_k$ will yield a local map pattern. Therefore, by estimating $\alpha_k$, the RE-ESF-SVC model could automatically select eigenvectors and consider the multi-scale property of spatially varying coefficients simultaneously.

## 3 Study area and data

The metropolitan area of Tokyo, Japan is one of the largest urban areas in the world. The long history of real estate development and a well-established market mechanism make this area ideal for investigating multi-scale spatial variations in housing prices. Apartment rent data in the metropolitan area of Tokyo for the year 2017 were provided by At Home Co., Ltd. High-rise condominiums with more than 14 floors were excluded because their building structures differed from those of other low-rise apartments. Additionally, apartments before 1981 were excluded because they were built based on old seismic design codes and standards. Consequently, the dataset consisted of 72466 samples of apartments. Table 1 lists the data attributes for the modeling, where the logged rent per square meter is used as the dependent variable. Because of the skewed distribution, other attributes are also transformed into their logs and set as explanatory variables.

**Table 1** Summary of variables.
(The five major stations are Tokyo, Shinagawa, Shibuya, Shinjuku, and Ikebukuro Station)

| Variables | Mean | Minimum | Maximum | Standard deviation |
|---|---|---|---|---|
| Rent per square meter (yen/$m^2$) | 2956.79 | 504.59 | 35888.50 | 747.71 |
| Walking time to the nearest station (min) | 7.09 | 1 | 30 | 4.20 |
| Travel time to five major stations (min) | 28.33 | 9.64 | 57.36 | 8.33 |
| Apartment age (year) | 20.29 | 1 | 35 | 9.10 |
| Area of property ($m^2$) | 35.95 | 10 | 296.42 | 20.27 |
| Floor number | 3.03 | 1 | 14 | 1.99 |

## 4 Results

In this study, the $(i, j)$th element of the spatial proximity matrix $\boldsymbol{C}$ is given by the exponential distance decay function, $c_{i,j} = exp(-d_{i,j}/r)$, where $d_{i,j}$ is the Euclidean distance between points $i$ and $j$. Following [4], the range parameter $r$ is determined by the longest distance in the minimum spanning tree connecting the $N$ observation sites. Consequently, 200 eigenvectors are extracted using a fast eigendecomposition method [8].

## 4.1 Results of RE-ESF-SVC

Table 2 summarizes the estimation results of RE-ESF-SVC. The adjusted $R^2$ shows that RE-ESF-SVC explains approximately 81.7% of the variation in logged prices. The residual standard error and BIC of RE-ESF-SVC are 0.109 and -112732.2, respectively. The coefficient for the floor number is a positive constant, indicating that it has a consistent, positive relationship with housing prices. In contrast, other coefficients are estimated to vary across the region.

**Table 2** Estimation results of RE-ESF-SVC.

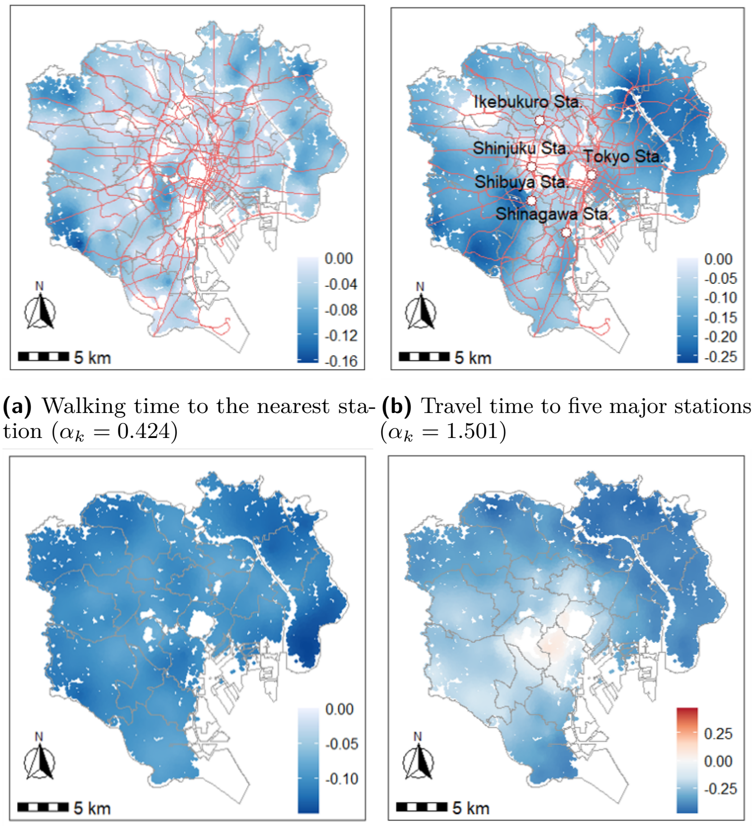
| | Minimum | Lwr Quartile | Median | Upr Quartile | Maximum | $\alpha_k$ |
|---|---|---|---|---|---|---|
| Constant | 8.577 | 9.389 | 9.568 | 9.768 | 10.400 | 1.996 |
| Walking time to the nearest station (min) | -0.164 | -0.051 | -0.035 | -0.024 | 0.021 | 0.424 |
| Travel time to five major stations (min) | -0.269 | -0.157 | -0.115 | -0.072 | 0.019 | 1.501 |
| Apartment age (year) | -0.150 | -0.107 | -0.098 | -0.090 | -0.070 | 0.873 |
| Area of property ($m^2$) | -0.476 | -0.356 | -0.277 | -0.219 | 0.093 | 1.417 |
| Floor number | 0.064 | 0.064 | 0.064 | 0.064 | 0.064 | NA |
| Residual Standard Error | 0.109 | | | | | |
| Adjusted $R^2$ | 0.817 | | | | | |
| BIC | -112732.2 | | | | | |

Figure 1 plots the spatial distribution of each varying coefficient. Only the samples whose coefficient estimates exhibit statistical significance (at least 95% confidence level) are colored. First, the walking time to the nearest station has a negative impact on prices at a local scale. Travel time to the five major stations has a global-scale impact, with its negative effect on prices evident in the north-eastern and south-western areas, indicating that the accessibility to central Tokyo encourages housing price growth. Since these two properties are commonly viewed as local and global measures of accessibility, these estimates are intuitively reasonable. Regarding the apartment age, the spatial variation of its negative impact on housing prices is relatively smooth in the whole region, whereas some apparent variations occur in the north-eastern areas, consequently yielding a moderate-scale map pattern. Coefficients for the area of properties vary at a wide spatial range, with particularly high negative values occurring in suburban areas, indicating that people who prefer to live in the suburbs of Tokyo value the size of their properties.

## 4.2 Model comparison

To investigate the flexibility of RE-ESF-SVC, we compared it with the conventional ESF-SVC. Two types of ESF-SVC models were used. The first is a preliminary model without the selection of eigenvectors $\boldsymbol{E}$. The other uses a subset of $\boldsymbol{E}$ consisting of 14 eigenvectors whose corresponding eigenvalues are beyond the threshold value of $\lambda/\lambda_{max} > 0.25$.

According to Table 3, RE-ESF-SVC achieves both the highest adjusted $R^2$ and the lowest BIC value, indicating that RE-ESF-SVC has the best explanatory power and generalization




**(a)** Walking time to the nearest station ($\alpha_k = 0.424$)

**(b)** Travel time to five major stations ($\alpha_k = 1.501$)

**(c)** Apartment age ($\alpha_k = 0.873$)

**(d)** Area of property ($\alpha_k = 1.417$)

**Figure 1** Estimated SVCs. The red lines in the first two panels represent the railways.

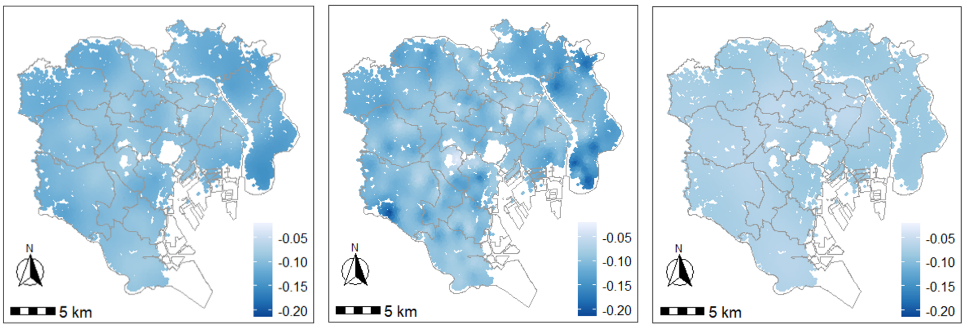
performance. As shown in Figure 2, the coefficients for the apartment age estimated by ESF-SVC without selecting eigenvectors present an uninterpretable localized map pattern because the eigenvectors that represent local spatial heterogeneity are not excluded, causing overfitting. However, after selecting eigenvectors, some valuable local spatial variations are ignored, yielding an overly smooth map pattern. In contrast, RE-ESF-SVC finds a good balance between local and global spatial scales through its flexible eigenvector selection controlled by $\alpha_k$.

**Table 3** Performances of the RE-ESF-SVC model and ESF-SVC models

| | RE-ESF-SVC | ESF-SVC (no selection) | ESF-SVC (selection) |
|---|---|---|---|
| Adjusted $R^2$ | 0.817 | 0.817 | 0.788 |
| BIC | -112732.2 | -103574.1 | -104170.3 |

## 5 Conclusion and future directions

This study utilized the RE-ESF-SVC model, a novel approach in regional analysis, to investigate the multi-scale spatial heterogeneity of the real estate market in the Tokyo




**(a)** RE-ESF-SVC **(b)** ESF-SVC(no selection) **(c)** ESF-SVC(selection)

**Figure 2** Coefficient for the apartment age estimated by RE-ESF-SVC and ESF-SVC

metropolitan area. The estimated results revealed that different housing characteristics affect prices at different spatial scales. Further, a comparison with the conventional ESF-SVC confirmed RE-ESF-SVC's ability to balance explanatory power and generalization performance through a flexible eigenvector selection mechanism.

However, the RE-ESF-SVC model cannot detect discontinuous variations at specific regional boundaries, such as the geographic segmentation of a market. Therefore, we will combine RE-ESF-SVC with discrete modeling approaches (i.e., fused lasso [9]) to simultaneously detect continuous and discontinuous spatial heterogeneity in our future work. Moreover, it is worth investigating the effects of other human and social determinants on rental prices to better understand the real estate market and assess urban living environment.

## References

1 A Stewart Fotheringham, Wenbai Yang, and Wei Kang. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6):1247–1265, 2017.

2 Alan E Gelfand, Hyon-Jung Kim, CF Sirmans, and Sudipto Banerjee. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396, 2003.

3 Marco Helbich and Daniel A Griffith. Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Computers, Environment and Urban Systems*, 57:1–11, 2016.

4 Daisuke Murakami, Takahiro Yoshida, Hajime Seya, Daniel A Griffith, and Yoshiki Yamagata. A moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*, 19:68–89, 2017.

5 Daniel A Griffith. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer/Le Géographe canadien*, 40(4):351–367, 1996.

6 Daniel A Griffith et al. *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization.* Springer Science & Business Media, 2003.

7 Daniel A Griffith. Spatial-filtering-based contributions to a critique of geographically weighted regression (gwr). *Environment and Planning A*, 40(11):2751–2769, 2008.

8 Daisuke Murakami and Daniel A Griffith. Spatially varying coefficient modeling for large datasets: Eliminating n from spatial regressions. *Spatial Statistics*, 30:39–64, 2019.

9 Ryo Inoue, Rihoko Ishiyama, and Ayako Sugiura. Identifying local differences with fused-mcp: an apartment rental market case study on geographical segmentation detection. *Japanese Journal of Statistics and Data Science*, 3:183–214, 2020.