Identifying Multiple Scales of Spatial Heterogeneity in Housing Prices Based on Eigenvector Spatial Filtering Approaches

Zhan Peng * and Ryo Inoue

Abstract: Interest in studying the urban real estate market, especially in investigating the relationship between house prices and related housing characteristics, is rapidly growing. However, this increasing attention is handicapped by a limited consideration of the multi-scale spatial heterogeneity in these relationships. This study uses the rental price data of 72,466 apartments in the Tokyo metropolitan area to examine spatial heterogeneity in the real estate market at multiple spatial scales. Within the framework of spatially varying coefficient (SVC) modeling, we utilized a random effect eigenvector spatial filtering-based SVC (RE-ESF-SVC) model, an approach not previously employed in real estate studies, and compared it with the traditional ESF-SVC model, which has no random effects. Our results show that: (1) except for one housing characteristic that impacts prices consistently throughout the Tokyo metropolitan area, relationships between other characteristics and prices vary from local to global spatial scales; (2) because of the utilization of random effects, RE-ESF-SVC has the unique advantage of making estimations flexibly while maintaining a high performance.

Keywords: spatial heterogeneity; urban real estate market; spatial scale; eigenvector spatial filtering; house prices

1. Introduction

Due to the close interaction between housing prices and socioeconomic activities, the urban housing market is seen as an important indicator of a city’s development status [1–3]. The changes in housing prices and the relationships with price-determining effects have caught the attention of those in the fields of real estate and urban planning [4–7].

Based on Lancaster’s consumer behavior theory [8], Rosen [9] proposed the famous hedonic price model, a prominent approach widely applied in real estate studies to quantify the influence of associated determinants on housing prices. The basic idea of the hedonic price model is that the price of a house is determined by a bundle of values of related housing attributes (e.g., building age, overall size, floor, the distance to the nearest station, and so on). Prices and attributes are treated as the dependent variable and explanatory variables in the context of linear regression, respectively. Hence, the traditional hedonic model allows the analyst to distinguish the impact of changing a particular attribute on price fluctuations by observing the coefficient of each explanatory variable.

The traditional hedonic model assumes that the influence of housing prices’ related factors are consistent throughout the whole region. For example, no matter where the house is located, the marginal price increase associated with each additional square meter is constant. Additionally, the traditional hedonic model is estimated under the assumption that the residual of the model is uncorrelated. However, these assumptions are less reasonable for the real estate market because location plays a major role in housing price formation, leading to two special properties of housing price data: spatial autocorrelation and spatial heterogeneity [10]. Spatial autocorrelation can be explained by Tobler’s first law.
of geography, which states that “nearby things are more related than distant things” [11]. Hence, for the real estate market, houses in the same neighborhood capitalize on shared locational amenities and have similar prices, which leads to correlated model residuals and inefficient estimates. Furthermore, spatial heterogeneity describes the inconsistent relationship between associated determinants and prices over space. Differences in regional characteristics such as living environment, economic conditions, and consumer types affect the property valuation in different areas, making the impacts of associated determinants on housing prices different throughout the region [12,13].

Meanwhile, it is essential to note that spatial heterogeneity could operate at multiple spatial scales. In the urban housing market, the impacts of some housing attributes tend to gradually change throughout the whole region, resulting in a global-scale spatial pattern. In contrast, other attributes may change at a smaller scale and show a localized spatial pattern. Scale is a fundamental concept in spatial analysis, and its variation significantly affects the model performance [14]. Jang and Kang [15] suggested that analyzing the relationship between specific factors and housing prices at different scales may lead to contrary estimation results. In addition, Murakami et al. [16] suggested that capturing scale effects could make the modeling of spatial heterogeneity more stable and flexible. Therefore, it is essential to consider spatial heterogeneity at multiple spatial scales meticulously to avoid any possible misspecification.

The development of spatial analysis motivates the proposal of analyzing multi-scale spatial heterogeneity by spatially varying coefficient (SVC) models. The multi-scale geographically weighted regression (MGWR) model, proposed by Fotheringham et al. [17], has been widely used in many fields [18–20]. In addition, based on the assumption of a certain prior distribution of coefficients, the Bayesian SVC (B-SVC) model [21] computes their posterior distribution on which the estimation and inference are performed. Nevertheless, previous studies claimed that the computational costs of both MGWR and B-SVC increase significantly when the sample size increases [22]. Hence, neither MGWR nor B-SVC can analyze the urban real estate market where massive transaction data is easily collected and provided by governments or companies currently.

The eigenvector spatial filtering-based SVC (ESF-SVC) model [23] considers spatial scales for each heterogeneous relationship by introducing eigenvectors of a pre-defined spatial weight matrix into a linear regression model. Coefficients could be estimated through a computationally efficient OLS method, making ESF-SVC easily used in actual application [24–26]. Nevertheless, if all eigenvectors are used in ESF-SVC, the increasing number of variables may result in an overfitting problem. Instead, when selecting and applying only a set of eigenvectors, ESF-SVC cannot fully identify spatial scales of the data because some eigenvectors that contain important information are excluded. Currently, Murakami et al. [27] extended the traditional ESF-SVC into a random effect version, namely the random effects ESF-SVC (RE-ESF-SVC) model. The most crucial advantage of the RE-ESF-SVC is that it assumes the coefficients of the eigenvectors as random variables with unique probability distributions, which makes RE-ESF-SVC quantify a distinct spatial scale for each coefficient without selecting eigenvectors. Additionally, Murakami and Griffith [28] proposed a fast estimation algorithm to reduce the computational burdens of calculating eigenvectors and estimating parameters, making the RE-ESF-SVC capable of handling over 10,000 data points. The application of the RE-ESF-SVC may be seen in references [29,30].

Although many studies have demonstrated the spatial heterogeneity of the urban real estate market so far [31–35], the promising RE-ESF-SVC model has not been used in any previous analysis and the discussion of measuring potential multi-scale spatial heterogeneity in the price formation process remains lacking, especially for highly developed metropolitan areas, such as Tokyo. Therefore, the main objective of this study is to utilize RE-ESF-SVC to efficiently explore and characterize the multi-scale spatial heterogeneity of urban housing prices, taking the Tokyo metropolitan area real estate market as a case. This study is an extension of reference [36], which did not sufficiently explain the formation of
the spatial distribution of each housing factor. Our research is aimed at making a further and more detailed exploration.

Based on the above, the following contributions to the literature are established. First, the application of RE-ESF-SVC confirms the existence of spatially varying relationships between housing attributes and rental prices at different spatial scales. For example, the time to the nearest station affects prices locally, whereas the impact of other spatially varying housing attributes changes globally in the Tokyo metropolitan area. Second, by analyzing how to characterize, and what to characterize in, the multi-scale spatial heterogeneity, we can better understand the process that shapes the current real estate market and further promote policies related to real estate and urban planning. Finally, the advantages of RE-EFS-SVC are examined by comparing it with the traditional ESF-SVC model, which provides new insight into utilizing this novel approach in more expansive fields related to spatial analysis.

The rest of this paper is structured as follows: First, Section 2 introduces the traditional ESF-SVC model as well as the RE-ESF-SVC model. Section 3 shows the study area and the housing price data. Section 4 discusses the estimation results of the RE-ESF-SVC. Section 5 summarizes the model comparison, and Section 6 concludes this study.

2. Methods
2.1. The ESF-SVC Model

ESF-SVC is a spatial regression approach proposed by Griffith [23]. ESF-SVC is built based on a spatial autocorrelation diagnostic statistic called the Moran coefficient (MC). For the $N \times 1$ vector $z$ that records observations at $N$ locations, its MC is formulated as follows:

$$MC = \frac{z^T (I - 11^T/N) C (I - 11^T/N) z}{z^T (I - 11^T/N) z}$$

where $C$ is an $N \times N$ symmetric spatial weight matrix representing the adjacent relationship among observations. $M = I - 11^T/N$ is an $N \times N$ projection matrix for centering, where $I$ is an $N \times N$ identity matrix and $1$ is an $N \times 1$ vector of ones. MC quantifies the extent of spatial autocorrelation by comparing covariation among $N$ sample values in $z$. Therefore, positive MC indicates the presence of positive spatial autocorrelation in $z$, and negative MC in case of negative spatial autocorrelation.

Eigenvectors $E = [e_1, \ldots, e_N]$ are decomposed from the centralized spatial weight matrix $MCM = (I - 11^T/N) C (I - 11^T/N)$, which is also a part of the numerator of MC. Generally, only eigenvectors with corresponding positive eigenvalues are utilized in analyses [37]. Griffith [38] further concluded that each eigenvector $e_k$ portrays a distinct map pattern of spatial autocorrelation and the scale depends on the corresponding eigenvalue. To elaborate, an eigenvector with a larger positive eigenvalue explains stronger positive spatial autocorrelation with a larger positive MC value and shows a more global map pattern. As the eigenvalue decreases, the spatial pattern of the corresponding eigenvector will also become local.

Based on this property, ESF-SVC extracts spatial heterogeneity of data by introducing interaction terms between eigenvectors and explanatory variables as shown in the following specification:

$$y = \sum_{k=1}^{K} x_k \circ \beta^\text{ESF}_k + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I),$$

$$\beta^\text{ESF}_k = \beta_k 1 + E \gamma_k$$

where $y$ is an $N \times 1$ vector of response variable and $X$ is an $N \times K$ matrix that includes a constant variable and all explanatory variables, the $N \times L$ matrix $E$ consists of $L$ eigenvectors ($L < N$) corresponding to positive eigenvalues, “$\circ$” is the element-wise product operator, $\beta^\text{ESF}_k$ is the spatially varying coefficient of the $k$th ($k = 1, \ldots, K$) explanatory variable $x_k$, in
which $\beta_k \mathbf{1}$ is the average value, and $E\gamma_k$ is the spatially varying component representing local modifications of the average value. Since each element in $\gamma_k$ determines the weight attributed to each eigenvector in $E$, with different $\gamma_k$ being estimated, the resulting $E\gamma_k$ will describe a different scale of spatial pattern for each $\beta_k^{\text{R-ESF}}$, by which ESF-SVC addresses the multi-scale spatial heterogeneity.

Note that utilizing many eigenvectors will significantly increase the number of interaction terms between eigenvectors and explanatory variables, leading to severe overfitting. To select a smaller set of eigenvectors, Griffith and Chun [39] first retain only eigenvectors whose corresponding eigenvalues satisfy the criterion $\lambda/\lambda_{\text{max}} > 0.25$, and further conduct a stepwise variable selection. However, the selection will exclude eigenvectors representing necessary local-scale spatial heterogeneity. Thus, the reduction in the number of eigenvectors will lead to a lower estimation accuracy, and ESF-SVC cannot identify the critical local characteristics of the housing market.

2.2. The RE-ESF-SVC Model

To address the above limitation, Murakami et al. [27] improved the conventional ESF-SVC model by introducing random effects into spatially varying terms, yielding the random effects ESF-SVC (RE-ESF-SVC) model whose specification is shown as follows:

$$y = \sum_{k=1}^{K} x_k \circ \beta_k^{\text{R-ESF}} + \epsilon, \quad \epsilon \sim N(0, \sigma^2\mathbf{I})$$

$$\beta_k^{\text{R-ESF}} = \beta_k^R + E\gamma_k, \quad \gamma_k \sim N(0, \sigma_k^2 \Lambda(a_k))$$

where $\beta_k^{\text{R-ESF}}$ is an $N \times 1$ vector of spatially varying coefficients regarding the $k$th explanatory variable $x_k$ and $E$ is composed of $L$ eigenvectors with corresponding positive eigenvalues. In addition, the coefficient $\gamma_k$ is a random variable with zero mean and $\Lambda(a_k)$ is an $L \times L$ diagonal matrix with the $l$th entry being $\lambda_l(a_k) = \left(\sum \lambda_l \lambda_k^l\right)\lambda_l^a_k$, where $\lambda_l$ is the eigenvalue corresponding to the $l$th eigenvector $\epsilon_l$. $\sigma_k^2$ and $a_k$ are parameters that determine the variance and scale of $E\gamma_k$, respectively.

The RE-ESF-SVC model is a linear mixed model with the fixed effect coefficient $\beta_k^R$ and the random effect coefficient $\gamma_k$ estimated by the restricted maximum likelihood method [27,40]. The computational cost of eigen-decomposition and the restricted likelihood maximization depend on the sample size, which will increase quickly as more data are included. According to reference [28], an eigenfunction approximation based on the Nyström extension is expected to accelerate the eigen-decomposition process, and a fast parameter estimation algorithm that reduces the matrix operations is applied to further shorten the computation time.

Compared to ESF-SVC, which is a fixed effect model, RE-ESF-SVC treats $\gamma_k$ as a random effect coefficient, which makes RE-ESF-SVC more flexible. The scale parameter $a_k$ controls the spatial scale of spatial variations in the $k$th regression coefficient $\beta_k^{\text{R-ESF}}$. To elaborate, a large $\gamma_k$ will shrink coefficients of eigenvectors with small eigenvalues (local-scale spatial pattern) strongly toward 0 and intensify the impact of eigenvectors with large eigenvalues (global-scale spatial pattern) at the same time. Thus, the resulting $E\gamma_k$ describes a global map pattern for $\beta_k^{\text{R-ESF}}$. On the contrary, $E\gamma_k$ presents a local map pattern with a small $a_k$. By estimating different $a_k$ for each $\beta_k^{\text{R-ESF}}$, RE-ESF-SVC could automatically adjust the effect of eigenvectors to quantify different scales of spatial heterogeneity without additional variable selection. Therefore, RE-ESF-SVC is a flexible model that could overcome the limitations of ESF-SVC. We will compare ESF-SVC and RE-ESF-SVC in the following application to further illustrate this issue.

3. Study Area and Data

3.1. Study Area and Rental Housing Data

This study focuses on the Tokyo metropolitan area, which is known as Japan’s political and economic center and one of the world’s most highly developed urban areas. The
Tokyo metropolitan area consists of 23 municipalities (wards, “ku” in Japanese) with a total area of about 619 km$^2$. The considerable total population of about 9.572 million assures a substantial demand for the real estate market and the distinct regional characteristics of each ward give rise to possible spatial heterogeneity in housing prices.

We use the Tokyo metropolitan area rental apartment prices data for the year of 2017, which is collected by At Home Co., Ltd. (Tokyo, Japan). Rental apartments are suitable for this study because the large sample size and wide spatial distribution of the transaction data are helpful to reflect the overall features of the market and reduce the estimation error. We removed apartments above 14 floors because the pricing strategy of high-rise condominiums differs from that of low-rise apartments. We also removed apartments built before 1981 because the old seismic design standards of buildings are not applicable after 1981. Consequently, the data set consists of 72,466 observations with geographic coordinates and detailed dwelling characteristics. Figure 1 shows the study area and spatial distribution of the apartment rent data.

![Figure 1. The study area and distribution of the apartment rent data.](image)

3.2. Variables

Table 1 shows the summary statistics of the variables. Following the specification used in the traditional hedonic model, the dependent variable in this study is the natural logarithm of rent per square meter, and the explanatory variables consist of structural and locational characteristics.

In particular, structural characteristics are apartment age, apartment size, and floor number. Apartment age generally has a negative impact on rent per square meter because of the property depreciation over time, while the floor number is commonly expected to have a positive impact because higher floors are considered quieter and have a better view. In addition, locational characteristics include the time to the nearest station and the
accessibility to major stations. Specifically, the accessibility to major stations is defined as the average commute time from the nearest station of each house to five major stations, including Tokyo, Shinagawa, Shibuya, Shinjuku, and Ikebukuro Station, which are located in the central part of the Tokyo metropolitan area. Because commute time increases as traffic conditions worsen, both these locational attributes should negatively affect rental housing prices. All of the above explanatory variables are logarithmically transformed to reduce the influence of a skewed distribution on estimation.

Table 1. Summary of variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rent per square meter (JPY/m²)</td>
<td>2956.79</td>
<td>504.59</td>
<td>35,888.50</td>
<td>747.7</td>
</tr>
<tr>
<td>Time to the nearest station (min)</td>
<td>7.09</td>
<td>1</td>
<td>30</td>
<td>4.20</td>
</tr>
<tr>
<td>Accessibility to major stations (min)</td>
<td>28.33</td>
<td>9.64</td>
<td>57.36</td>
<td>8.33</td>
</tr>
<tr>
<td>Apartment age (year)</td>
<td>20.29</td>
<td>1</td>
<td>35</td>
<td>9.10</td>
</tr>
<tr>
<td>Apartment size (m²)</td>
<td>35.95</td>
<td>10</td>
<td>296</td>
<td>20.27</td>
</tr>
<tr>
<td>Floor number</td>
<td>3.03</td>
<td>1</td>
<td>14</td>
<td>1.99</td>
</tr>
</tbody>
</table>

Note: major stations are Tokyo, Shinagawa, Shibuya, Shinjuku, and Ikebukuro Station.

4. Results of RE-ESF-SVC

This study first examined the estimation results derived from the RE-ESF-SVC model. The estimates and distributions of the coefficients revealed the fact that some housing attributes impact prices on different spatial scales, which highlighted the importance of considering the multi-scale spatial heterogeneity of the Tokyo metropolitan area real estate market.

We extracted eigenvectors from the centralized spatial weight matrix $MCM$. The $(i,j)$th element of the matrix $C$ is calculated based on the following exponential distance decay function:

$$c_{i,j} = \begin{cases} 
\exp(-d_{i,j}/r), & \text{if } i \neq j \\
0, & \text{otherwise}
\end{cases}$$

where $d_{i,j}$ is the Euclidean distance between observation sites $i$ and $j$. Following references [27,41], we fixed the range parameter $r$ as the longest distance in the minimum spanning tree that connects all sample points. Using the fast eigen-decomposition method [28], 200 eigenvectors corresponding to positive eigenvalues were extracted and applied in RE-ESF-SVC. The "spmorans" package in R [42] was used to estimate RE-ESF-SVC parameters.

4.1. Summary of Estimates

Table 2 summarizes the estimated coefficients and diagnostic information of RE-ESF-SVC. The Adjusted $R^2$ suggests that explanatory variables can explain 81.6% of the variation in logged rental prices in the context of RE-ESF-SVC. No significant residual correlation is detected by the Moran coefficient (0.005).

As shown in Table 2, coefficients of the floor number were estimated to be constant, indicating no significant spatial variations in this attribute in the study area. The zero estimate of $\sigma^2_k$ confirms that the impact of floor number on rental prices is consistent. On the contrary, coefficients of all other attributes are varying on different spatial scales, with each having a distinct positive $\alpha_k$ value. This result verifies that the average estimation by the traditional hedonic price model may be flawed. Among attributes with spatially varying coefficients, the apartment age is negatively associated with rental prices across the study area. Regarding two variables having locational characteristics, although their estimates reveal both negative and positive effects on rent per square meter, those positive values are not significant at the 95% confidence level. It is interesting that the apartment size shows both significant negative and positive effects on prices. The above results provide strong
evidence that the multi-scale spatial heterogeneity model is not negligible in the Tokyo metropolitan area real estate market.

Table 2. Results of the RE-ESF-SVC model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>LQ</th>
<th>Median</th>
<th>UQ</th>
<th>Max</th>
<th>$\sigma^2_k$</th>
<th>$a_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.577</td>
<td>9.389</td>
<td>9.568</td>
<td>9.768</td>
<td>10.400</td>
<td>0.021</td>
<td>1.996</td>
</tr>
<tr>
<td>Time to the nearest station</td>
<td>−0.164</td>
<td>−0.051</td>
<td>−0.035</td>
<td>−0.024</td>
<td>0.021</td>
<td>0.002</td>
<td>0.424</td>
</tr>
<tr>
<td>Accessibility to major stations</td>
<td>−0.269</td>
<td>−0.157</td>
<td>−0.115</td>
<td>−0.072</td>
<td>0.019</td>
<td>0.004</td>
<td>1.542</td>
</tr>
<tr>
<td>Apartment size</td>
<td>−0.476</td>
<td>−0.356</td>
<td>−0.277</td>
<td>−0.219</td>
<td>0.093</td>
<td>0.005</td>
<td>1.557</td>
</tr>
<tr>
<td>Apartment age</td>
<td>−0.150</td>
<td>−0.107</td>
<td>−0.098</td>
<td>−0.090</td>
<td>−0.070</td>
<td>0.001</td>
<td>1.309</td>
</tr>
<tr>
<td>Floor number</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Residual standard error</td>
<td>0.109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual Moran coefficient</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.817</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>−112,732.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: major stations are Tokyo, Shinagawa, Shibuya, Shinjuku, and Ikebukuro Station.

4.2. Distributions of Coefficients

Figure 2 visualizes four spatially varying coefficients. We note that only samples whose coefficient estimates are statistically significant at the 95% confidence level are colored.

In Figure 2a, we see that as the time from the property to the nearest station increases, rent per square meter declines, which is applicable across the whole region. This negative influence presents a relatively localized spatial pattern because of the small $a_k$ value (0.424).

From Figure 2b, compared to the time to the nearest station, coefficients regarding the accessibility to major stations change at a global scale because of the large $a_k$ value (1.542).

The estimates are intuitively reasonable because these two properties are commonly used local and global measures to reflect the traffic accessibility. Although the accessibility to major stations encourages rental housing price growth across the Tokyo metropolitan area, significant negative values are mainly located in the northeastern and southwestern suburbs, where land use is mainly for residential and industrial purposes. In contrast, the impact is less pronounced in central Tokyo, which has a high concentration of public, political, and commercial facilities within a short distance. Thus, the high convenience of living and the diversity of demand diminish the impact of commute time on rental housing prices.

Figure 2c shows a global-scale spatial variation of the coefficients regarding the apartment size. Negative impacts are observed in the surroundings of the study area, especially in the three northeastern wards. Conversely, positive impacts can be found in Minato ward, where the increase in the apartment size contributed to higher rent per square meter. This result seems to be against our expectations, but it is reasonable when regional characteristics are taken into consideration. As one of the most luxurious areas of Tokyo, Minato ward is where the majority of business and diplomatic activities are concentrated. Many extra-large premium residences, whose price valuations are different from other apartments, are located there, leading to higher rents with increased property sizes.

Finally, the apartment age is statistically significant across the entire study area. Although the large $a_k$ value (1.309) suggests a global-scale spatial heterogeneity of the apartment age, the small $\sigma^2_k$ value (0.001) indicates that its negative impacts do not vary as drastically as other variables, consequently yielding a relatively smooth map pattern shown in Figure 2d.
5. Model Comparison

As mentioned in Section 2, compared to ESF-SVC, RE-ESF-SVC could analyze multi-scale spatial heterogeneity more flexibly while maintaining a high accuracy at the same time. This section verifies the advantages of RE-ESF-SVC by comparing these two models. Two ESF-SVC models are compared with RE-ESF-SVC. The first one uses the same 200 eigenvectors (hereinafter referred to as the full ESF-SVC) as RE-ESF-SVC. The other one uses only 14 eigenvectors, whose corresponding eigenvalues are beyond the threshold of $\lambda / \lambda_{\text{max}} > 0.25$ (hereinafter referred to as the selected ESF-SVC).

As shown in Table 3, which summarizes the estimates of the two ESF-SVC models, all housing characteristics were estimated to have spatially varying effects on prices. In contrast, RE-ESF-SVC could identify the consistent effect of the floor number and quantify the distinct scale of spatial variation for other spatially varying coefficients through parameter $\alpha_k$ (see Table 2). The Adjusted $R^2$ value of the full ESF-SVC is the same as RE-ESF-SVC, but the lower BIC of RE-ESF-SVC suggests a lower information redundancy and better generalization performance, which is critical for applications to other data. Additionally, among the three models, the selected ESF-SVC has the lowest adjusted $R^2$ value and the highest residual standard error, indicating that the selection of eigenvectors in the ESF-SVC model leads to the decrease in explanatory power and estimation accuracy.
Table 3. Results of ESF-SVC models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>The Full ESF-SVC</th>
<th>The Selected ESF-SVC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>LQ</td>
</tr>
<tr>
<td>Time to the nearest station</td>
<td>-0.241</td>
<td>-0.052</td>
</tr>
<tr>
<td>Accessibility to major stations 1</td>
<td>-3.646</td>
<td>-0.222</td>
</tr>
<tr>
<td>Apartment age</td>
<td>-0.213</td>
<td>-0.110</td>
</tr>
<tr>
<td>Apartment size</td>
<td>-0.519</td>
<td>-0.357</td>
</tr>
<tr>
<td>Floor number</td>
<td>-0.010</td>
<td>0.053</td>
</tr>
<tr>
<td>Residual standard error</td>
<td>0.109</td>
<td></td>
</tr>
<tr>
<td>Residual Moran coefficient</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.817</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>-103,574.1</td>
<td></td>
</tr>
</tbody>
</table>

1 Note: major stations are Tokyo, Shinagawa, Shibuya, Shinjuku, and Ikebukuro Station.

For illustrative purposes, taking the coefficient of the apartment age as an example, Figure 3 visualizes the estimates given by the RE-ESF-SVC and ESF-SVC models. When all eigenvectors were input into the full ESF-SVC, unnecessary local spatial variations represented by some eigenvectors with small eigenvalues are redundant, leading to an overly localized spatial pattern that is hard to interpret. Instead, directly excluding all eigenvectors with small eigenvalues through selection will lose necessary local information, so the result of the selected ESF-SVC displays an overly smooth map pattern where almost no spatial variation can be observed. On the contrary, RE-ESF-SVC could find a good balance between the global and local spatial variations. Although all eigenvectors were used, RE-ESF-SVC adjusted appropriate eigenvectors for the variable to represent its distinct spatial heterogeneity, resulting in an understandable map pattern where the global-scale negative impacts are more evident in the surrounding area than the central area. The above comparison confirms the flexibility and higher accuracy of RE-ESF-SVC.

Figure 3. Distributions of the coefficient for the apartment age estimated by RE-ESF-SVC and ESF-SVC. (a) RE-ESF-SVC. (b) The full ESF-SVC. (c) The selected ESF-SVC.

6. Conclusions

There is growing interest in real estate and urban planning studies to investigate the relationship between housing prices and related housing characteristics. However, this increasing attention is handicapped by the limited consideration of multiple scales of spatial heterogeneity in the data. In this study, we collected data on 72,466 rental apartments in the Tokyo metropolitan area and utilized the RE-ESF-SVC model to analyze how different housing characteristics are related to prices on different spatial scales. The analysis results lead us to the following findings.

First, it is necessary to consider multiple scales of the spatial heterogeneity in the Tokyo metropolitan area real estate market. The different customer preferences and distinct
regional characteristics of each ward lead to varying relationships between prices and associated attributes. RE-ESF-SVC could extract these inconsistent relationships by allowing coefficients to vary according to location. According to the estimates and distributions of the coefficients of the RE-ESF-SVC, the time to the nearest station negatively affects prices locally, whereas accessibility to major stations, apartment age, and apartment size have global-scale impacts. The various spatial scales are quantified by different estimated values of $a_k$.

Second, the comparison with the traditional ESF-SVC model, which has no random effects, suggests that RE-ESF-SVC is more flexible and effective. We compared the RE-ESF-SVC model with two ESF-SVC models, one using the same number of eigenvectors as RE-ESF-SVC and the other one using only a selected set of eigenvectors. The highest adjusted $R^2$ and the lowest BIC value of RE-ESF-SVC suggest that RE-ESF-SVC statistically outperforms ESF-SVC. In addition, RE-ESF-SVC treats the coefficients of eigenvectors as random variables, with each having its own possibility distribution. This allows RE-ESF-SVC to automatically adjust eigenvectors to determine a distinct spatial scale for each spatially varying coefficient without a previous variable selection, which is necessary for ESF-SVC, resulting in interpretable spatial patterns and a more accurate estimation.

Third, we attempted to provide an application of RE-ESF-SVC in a spatial analysis. RE-ESF-SVC is a novel method whose application is still limited, especially in urban real estate analysis. RE-ESF-SVC has its benefits by efficiently handling large amounts of data and flexibly identifying the spatial heterogeneity of data, which give rise to the possibility of applying RE-ESF-SVC in more expansive fields, such as ecology and transportation. We believe our findings could serve as inspiration for future studies.

There are, however, some limitations to this research. The explanatory variables of the models only consist of five structural and locational factors due to the limitations of the data collection. It is suggested to investigate how neighborhood factors, such as the diversity of land use, the distribution of public services, and the concentration of people, impact rental prices at the same time. These determinants are becoming increasingly influential in the context of livable and attractive cities. Therefore, we hope to incorporate complementary indicators for a more general explanation of the spatial variability in the urban real estate market in the future. In addition, we note that spatial heterogeneity in the real estate market could operate on a discrete spatial scale. For example, significant price changes may occur in specific local regions with well-defined boundaries, such as the brand-name area, a reputable school district, and the surroundings of major stations. Future work will focus on investigating discrete spatial heterogeneity at the same time.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijgi11050283/s1, A part of the dataset is openly available in Mendeley Data at 10.17632/fjtbpb628j.1, reference number [43].

Author Contributions: Conceptualization, Zhan Peng; methodology, Zhan Peng; software, Zhan Peng; validation, Zhan Peng; formal analysis, Zhan Peng; investigation, Zhan Peng; resources, Ryo Inoue; data curation, Zhan Peng; writing—original draft preparation, Zhan Peng; writing—review and editing, Ryo Inoue and Zhan Peng; visualization, Zhan Peng; supervision, Ryo Inoue; project administration, Ryo Inoue; funding acquisition, Ryo Inoue. All authors have read and agreed to the published version of the manuscript.

Funding: This research was founded by the Japan Society for the Promotion of Science KAKENHI, Grant Numbers JP18H01552 and JP21H01447.

Data Availability Statement: The original real estate rent data provided by At Home Co., Ltd. cannot be made publicly available due to the data disclosing policy of the data owner. Therefore, we disclose a part of the dataset recorded on 8000 properties that is openly in Supplementary Materials. The dependent variable, rent per square meter of properties, was transformed logarithmically and the explanatory variables, the attributes of properties, were also transformed logarithmically and then normalized to mean 0 and variance 1. Noises were added to the geographic coordinates of the data.
Acknowledgments: This research was the result of joint research with the Center of Spatial Information Science, The University of Tokyo (No. 815). We thank At Home Co, Ltd. for providing the apartment rental data used in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References


32. Li, H.; Wei, Y.D.; Yu, Z.; Tian, G. Amenity, accessibility and housing values in metropolitan USA: A study of Salt Lake County, Utah. *Cities* 2016, 59, 113–125. [CrossRef]


42. Murakami, D. *Spmoran: Moran’s Eigenvector-Based Spatial Regression Models*; R Package Version 0.2.2.2; 2022. Available online: https://CRAN.R-project.org/package=spmoran (accessed on 18 March 2022).