学位論文

Spatial Statistical Analysis and Its Temporal Variation of Tokyo Land Prices

東京都地価の空間統計分析とその経年変化

Kanno Yuta

Graduate School of Engineering

Tokyo University of Science

March 2023

Preface

This thesis is a product of my Ph.D. studies at the Department of Management Science, Graduate School of Engineering, Tokyo University of Science. It has greatly benefited from the generous support of several individuals.

This thesis consists of three accepted papers. It is, therefore, a study that analyzes the polarization and dispersion of land prices through three papers—collaborations between Yuta Kanno and Takayuki Shiohama, a Professor at Nanzan University, Department of Data Science, Faculty of Science and Engineering.

I would like to express my sincere gratitude to Prof. Takashi Seo, Prof. Takako Akakura, Prof. Yukinobu Taniguchi, Prof. Takashi Sozu, Masaya Fujisawa, and Takayuki Shiohama for their patience, constructive guidance, and valuable advice.

In particular, I would like to thank Prof. Takashi Sozu for his help in the preparation of the judging committee. I would also like to thank our secretaries: Ms. Hideko Otobe, Ms. Masayo Ida, and Ms. Nozomi Inaba.

All omissions and errors are my own.

Yuta Kanno

Contents

List of Figures

List of Tables

1

Introduction

From the viewpoint of area studies and city planning, the polarization and dispersion of land prices have become problematic in recent years. Here, "polarization of land prices" is defined as the differentiation and concentration of land prices toward two or more centers. "Increased dispersion" is defined as widening price differentials on a global area and failure to maintain price homogeneity on a regional area. These problems impact several groups of people and things, such that the Japan Real Estate Appraisers Association, the contractor for the public notice of land prices, is considering changing the appraisal method (Ministry of Land, Infrastructure, Transport and Tourism, 2020). This thesis evaluated land price analysis related to the problem awareness of city planning and area studies, as described above. From the problem awareness, the existing spatial statistical models are used to analyze the temporal change of its regression coefficient and residual structure. By following the changes over time, we investigated the cause of polarization and widening gap in land prices and analyzed how the dispersion increases.

This thesis comprises three papers. In Kanno and Shiohama (2020) corresponding to Chapter 3 in this thesis, a spatial process model in the field of geostatistics was used to estimate the price model of public land prices. The regression coefficients' temporal changes and residual structure were analyzed. In Kanno and Shiohama (2021), which corresponds to Chapter 4, the land price model was estimated using a spatial econometric model in econometrics. Temporal changes in the regression coefficients were analyzed, and the estimated

distributions of the regression coefficients were visualized. In Kanno and Shiohama (2022) corresponding to Chapter 5, the economic situation of each municipality was incorporated into the model; using the spatial process model of land prices, the temporal changes of the estimated regression coefficients and residual structure were analyzed from 1997 to 2021. Then, we discussed the impact on land prices in the early days of the Corona disaster.

Statistical analyses of real estate prices are typically investigated using the hedonic approach—a technique for measuring non-market asset values using regression models. Examples of explanatory variables in the regression models include environmental and social capital. Through land and house models, it is possible to observe how differences in environmental conditions can result in differences in land and house prices. Early studies on the hedonic approach include the housing market analyses of Rosen (1974) and Goodman (1978).

However, constructing a land price model with accurate forecasting performance using the ordinary least squares (OLS) method requires collecting a large volume of land price data together with sufficiently large explanatory variables. As Kato (2005) stated, not all factors that have an impact on land prices are observable, and the error of land price models has spatial and time-series correlations because real estate evaluation depends on the surrounding environment and time trend.

In this thesis, the following two spatial statistical models were the focus of analysis.

- **Spatial process model**

  In geostatistics, there is a technique called "kriging" which originated from Matheron (1963). It assumes second-order stationarity in a random field of a spatial process. It expresses the covariance of the observed data as a function of distance, structures the spatial correlation, and makes a spatial prediction of the value of the random field at an arbitrary point. Kriging is a statistical method that uses a structured covariance function to obtain a spatial predictor that is the best linear unbiased prediction. Universal kriging is a method of spatial prediction based on the regression residuals of a land price model.

There have been many studies estimating real estate price functions using spatial process models. In the early studies, Dubin (1998) and Basu and Thibodeau (1998) used a generalized least squares method to estimate land price trends and parameters for the spatial correlation structure of the residuals. Hengl et al. (2007) used the geographical information of Croatia to explain the characteristics and limits of universal kriging. Li and Heap (2011) presented 18 comparative studies on spatial prediction accuracy and affirmed that the prediction performance of kriging is superior to that of other spatial interpolation methods. Real estate price analyses using co-kriging, which performs spatial interpolation using covariates, were carried out by Chica-Olmo (2007), Montero and Larraz (2011), Chica-Olmo et al. (2013) and Kuntz and Helbich (2014). In Japan, for example, Inoue et al. (2016) used the official land price and standard land price data of the 23 wards in Tokyo over 15 years (between 2000 and 2014) and set a land price model for each use district. Then, they assumed a spherical model of a spatio-temporal correlation structure in both spatial and time series directions and adopted that model to perform spatio-temporal interpolation of land prices. With this procedure, they performed spatial interpolation with approximately 80% of all interpolation points with an interpolation accuracy within 5% and approximately 95% with accuracy within 10%.

- **Spatial econometric model**

  In spatial econometrics, the GWR model was proposed by Brunsdon et al. (1996) and Fotheringham et al. (1998) as a spatial statistical model considering spatial heterogeneity. To build a land price model, it is necessary to model the complex relationships between the real estate market and macroeconomic variables such as inflation, economic growth, GDP, and the unemployment rate. The hedonic approach measures the economic value of the environment based on the capitalization hypothesis (i.e., the differences in environmental conditions are reflected in real estate prices). Additionally, it is necessary to consider spatial influences, such as spatial autocorrelation, spatial heterogeneity, and non-stationarity in the spatial

statistical analysis of land prices. For land price analysis using a second-order stationary spatial model, we assume that the covariance of the observed values only affects the distance between points. However, such an isotropic assumption does not generally hold for land price analysis. Moreover, we often use data in which points with large local variations and those with less variation are mixed. The GWR model is a local regression model that captures spatial heterogeneity or non-stationarity by estimating spatially varying regression coefficients.

Cho et al. (2006) estimated the GWR model using housing data from Knox County, Tennessee, showing that the proximity of water areas and parks to housing is reflected in the price. One of the disadvantages of the GWR model is that the multicollinearity between explanatory variables occurs when using a common bandwidth for spatial kernels for all explanatory variables, which yields similar or unstable regression coefficients for the target area. Hence, various models that extend the GWR model have been proposed and applied in the literature. Lu et al. (2017) overcome the shortcomings of the GWR model by using the MGWR model, which estimates the local regression coefficients using variable-specific bandwidth for spatial kernels. Various global GWRapplications also exist in the literature. Helbich et al. (2014) estimated a mixed GWR model that distinguishes between local and global explanatory variables based on Australian residential data. Using sample sizes based on distance measures for spatial kernels, Lu et al. (2015) argued that the fitting performance is better than that for the usual distance-based kernels and constructed a parameter-specific, distance-metric GWR (PSDM-GWR) model using both distinct bandwidth and metric functions of each explanatory variable. Additionally, they proposed back-fitting algorithms to fit the generalized linear model with the parameter estimation of PSDM-GWR models. In Japan, for example, under a land price model using Bayesian GWR and ordinary least squares (OLS), Furuya (2004) conducted land price analysis for Yokohama City and revealed that the estimated regression coefficients might differ significantly by estimation method. Uesugi (2012) estimated a multiple regression model, a spatial autoregressive model (SAM), a spatial

error model (SEM), and a GWR model using official land price data for Saitama City and compared the parameters of the GWR model with those with the best estimation accuracy among the four models, showing that the effect of the explanatory variable on land price differs by region.

Using the above spatial models, we will analyze and elucidate the following issues.

1. **Superiority of Spatial Process Model and Spatial Econometric Model**

   First, in Chapter 3, we compare the estimated spatial process model with the non-spatial model using the OLS method. In Chapter 4, we compare the GWR model, the MGWR model, and the non-spatial model. By comparing the residuals of land price models, spatial correlation of the residuals, and the adjusted contribution rate, etc., we confirm the spatial models' goodness of fit using the IRWGLS method and the MGWR model.

2. **Temporal changes in the regression coefficients of the spatial process model constructing the residual structure**

   Next, in Chapters 3 and 5, we analyze the temporal changes in the spatial process model's estimated regression coefficients and the residual structure's temporal changes. While many studies have described how useful kriging is, not many have looked at the annual changes in spatial process model coefficients or spatial variograms over 25 years. Masunari (2007) analyzed the annual changes in the variograms and regression coefficients of land price models. Chica-Olmo et al. (2019) investigated a hedonic regression model to estimate housing prices and the spatial variability of prices over multiple years in Granada, Spain. All of these studies have mentioned the difficulty of assuming isotropy in the spatial process of land prices in Tokyo, but in this study, this issue is dealt with by dividing Tokyo into four regions using a clustering technique for land price data for which the assumption of isotropy does not hold. As a result, unlike previous studies, we were able to extract the characteristics of land price formation in each region of Tokyo.

3. **Temporal changes in the regression coefficients of the spatial econometric model**

   Third, in Chapter 4, we analyze the temporal changes in the estimated regression coefficients of the spatial econometric model. In this field of research, there are many studies aimed at extracting environmental factors at some point in time, but not many studies have investigated annual changes in about a quarter of a century. We assume independence between the different time points and estimate secular changes under the GWR and MGWR models. This study clarifies the interannual variability of geographical and environmental factors for land prices by applying the GWR and MGWR models. Lu et al. (2017) estimated GWR and MGWR models using housing transaction prices in London in 2001 and revealed that the MGWR model is superior in terms of fitting and prediction accuracy. Recently, several studies have extended the GWR and MGWR models to space-time dimensions, including Huang et al. (2010). LeSage and Pace (2009) derived estimates focusing on the results of long-term spatiotemporal equilibrium with regard to the use of cross-sectional data and focusing on the dynamics embodied by time-dependent parameters regarding the use of spatiotemporal data.

4. **In models in two fields, each region has a preference depending on the area of use**

   Fourth, in Chapters 3–5, we analyze districts' use. In the business world, areas commonly referred to as "luxury residential areas" are regulated by low-rise residential areas. It tends to be more expensive than other use districts. We analyze the actual situation.

5. **Addition of newest data and the impact of the Covid-19 pandemic**

   Finally, in Chapter 5, we estimated the spatial process model with the addition of data up to 2021 when the effects of covid-19 appear and discussed the effects of corona damage on land prices.

The following summarizes what we learned from the three accepted papers that constitute this thesis. The spatial process model was found to fit better

than the non-spatial model. Of the GWR, MGWR, and non-spatial models, the MGWR model was found to have the best fit. Due to the secular changes in environmental factors that indicate convenience of the spatial process and spatial econometric models, land price between the central or southern parts of the 23 wards including surrounding and other areas, polarize and disperse. In particular, in the western part of Tokyo, the estimates of two models show that the land prices increase to the east. Over the past 20 years, spatial dispersion of local area in land prices in Tokyo except the city center has increased, judging from in the spatial process model, the temporal changes in the spatial structure of the estimated regression residuals, and in the spatial econometric model, the increasing trend of the range of the estimated regression coefficients. The increase in spatial dispersion can also be explained by the increase in the importance of individual factors, such as hazard map caution areas, crime, local noises, and sunshine conditions, not included in this study. The preference for brands has seen the official land prices in the central part of the 23rd ward increase, a tendency that grows stronger every year. There are also regional differences in land preferences judging from the secular changes in environmental factors that indicate usage area of the spatial process and spatial econometric models. For example, from the central to the northern areas of the 23 wards and from the western area of Kitatama to the eastern area of Nishitama (Western Tama Area), low-rise residential areas with an emphasis on the living environment are preferred. Conversely, from the Minamitama (Southern Tama) area to the western area of Nishitama (Western Tama Area), residential and semi-residential areas with an emphasis on convenience and commerciality are preferred. Each influence became stronger over time. Further, the effects mentioned above changed significantly before the 2008 financial crisis and remained stable after the crisis. These factors have not been explained in previous studies and can be the factors in the polarization of land prices. Finally, it was found that there are areas and periods when the spatial correlation of residuals becomes stronger, without the effect of covariates that influence official land prices. Although the official land price in Tokyo in 2021 fell because of the Covid-19 outbreak, no significant changes were confirmed in the determinants of land prices and the structure of regression residuals. However, in the 23rd district, it can be confirmed that the

spatial correlation of the regression residuals is weakened, and the influence of the outbreak is likely to continue.

The changes in land policies since the collapse of the bubble economy have increased the imbalance in population and accelerated the polarization and the widening gap in land prices. These results suggest the need to switch from previous land policies that stimulated polarization and widening gap in land prices to new policies that spatially stabilize land price fluctuations through the cooperation of various related organizations that can control the dispersion of land price fluctuations.

2

Urban Structure of Tokyo

2.1 Urban Structure

Tokyo's urban structure has been studied in many fields from different perspectives. Since urban form significantly impacts urban sprawl and transportation, changes in the urban structure of Tokyo have garnered attention not only in the field of urban planning but also in demography, transportation, environment, and economics. For example, Hatta and Ohkawara (1993), Sakai et al. (2016), Zheng (2007), Wang et al. (2018), and Zhou and Gao (2020). Bagan and Yamagata (2012) studied how Tokyo became the world's largest megacity using remote sensing and census data.

As a typical case of the world's largest cities, chronic traffic jams and commuting hells have been persistent features of Tokyo since the 1960s. Currently, Tokyo has a good transportation network that connects urban and suburban locations and enables the city population to commute long distances (Nozawa and Higuchi, 2001). Tokyo is the capital of Japan and the center of politics, economy, imperial family, and media; it is unlike other capitals that separate political functions from economic zones, such as Canberra in Australia, Brazil, and Washington DC in the United States. Many companies have their headquarters concentrated in Tokyo because they think it is more advantageous for their business to be based there, where central ministries and agencies are situated. While the concentration in Tokyo is highly convenient, the problem of over-concentration in metropolitan Tokyo has been pointed out by several studies

over concerns of maintaining sustainability in the region. Decentralization, as stated in the monograph of Hein and Pelletier (2006) and Mochida (2008), is a key governmental policy that has been emphasized for reducing concentration within the metropolitan Tokyo area. The expansion of Tokyo as a city was not due to administrative control but rather due to the private sector's economic activities accompanying Tokyo's overconcentration. Several urban policies have been implemented, but population concentration surpasses all regulations (Omura, 2010).

The Tokyo metropolitan area is divided into three parts: central Tokyo, suburbs along the railways, and suburbs distant from the railways. Each area has its unique characteristics, along with railway lines and stations. Tokyo's 23 wards are central to Tokyo. The study of the development of the urban fringe of the Tokyo metropolitan area has attracted significant attention because sustainable rural systems are key concepts of urban development and new perspectives, such as curbing the expansion of urban areas and the need for compact urban development (Li and Monzur, 2018; Kikuchi, 2008; Yokohari et al., 1994).

Many studies exist on Tokyo's urban structure and planning; however, there are only a few studies on Tokyo's spatial and temporal land price distributions. This study analyzes the spatial structure of Tokyo land prices with temporal variations. The next subsection describes the study area and data used in this research. The findings provide helpful suggestions for planning a sustainable urban form of development for Tokyo.

2.2 Overview of the Data

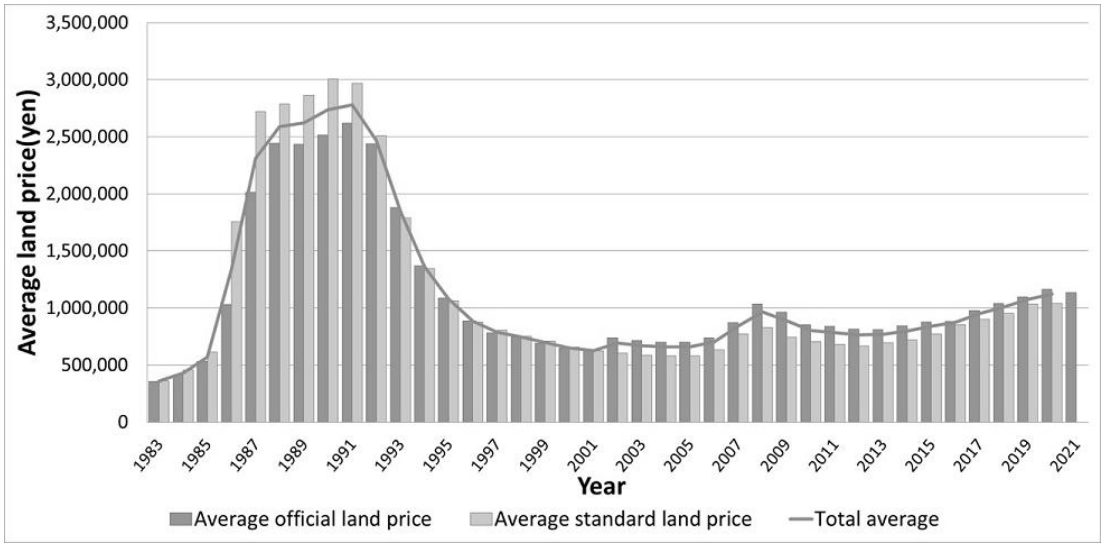


Figure 2.1: Time series plots for average land prices from 1983 to 2021.

Figure 2.1 plots the transition of the average official/standard land prices in all use districts in Tokyo over 39 years between 1983 and 2021. During the bubble economy between 1986 and 1991, real estate prices rose rapidly but dropped sharply after the bubble burst. Although prices rose momentarily from 2006 to 2008, they entered a downward trend again because of the 2008 financial crisis. More recently, prices have sustained a gradual upward trend for seven consecutive years, approximately two years ahead of the national land price average. The average land price in 2021 fell for the first time in eight years due to the Covid-19 outbreak.

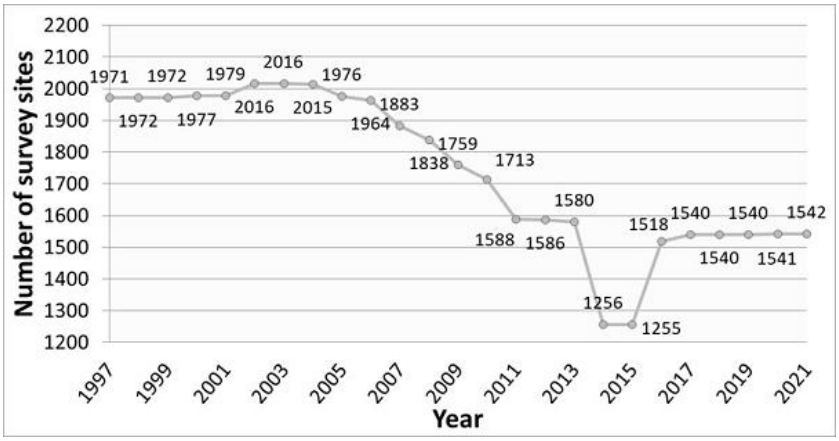


Figure 2.2: Time series plots of numbers of survey sites from 1997 to 2021.

Figure 2.2 shows the transition in the number of official land price survey sites in residential areas in Tokyo (excluding the Tokyo Islands) between 1997 and 2021[1]. Survey sites are examined and changed annually, and those deemed ineligible are replaced. The abnormal price fluctuations during the bubble economy until its collapse emphasize the importance of identifying land price trends in more detail. As a result, the number of survey sites in the mid-1990s almost doubled compared to the 1980s. It stayed close to 2,000 sites until around 2005, but it has been gradually decreasing for various reasons, such as a decrease in real estate appraisers. In 2014, it decreased by more than 20% from the previous year. However, in the 2016 official land prices, the number of survey sites reached the pre-2013 level and has been hovering at about 1,500 sites in recent years.

[1]Here, "residential districts" refer to seven types of districts: low-rise exclusive residential districts of categories 1 or 2, medium-to-high rise exclusive residential districts of categories 1 or 2, residential districts of categories 1 or 2, and quasi-residential districts.

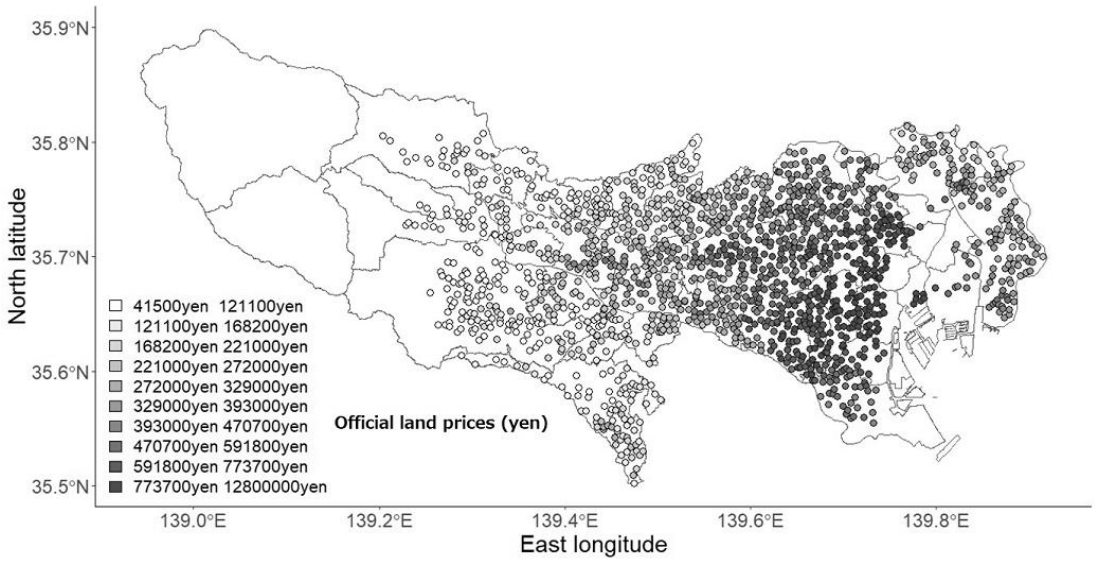


Figure 2.3: Spatial distribution of Tokyo official land prices in 2021.

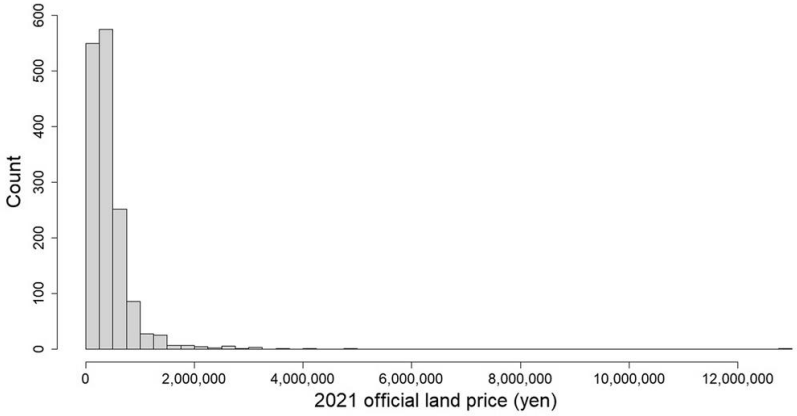


Figure 2.4: Histogram of Tokyo official land prices in 2021.

Figure 2.3 shows the spatial distribution of official land prices in Tokyo's residential areas in 2021. The number of official survey sites was 1,542, and the price per square meter varied from JPY 41,500 to JPY 12,800,000. Official land prices are higher near the center of the 23 wards and decrease concentrically outwards. However, the official land prices in the 23 wards are not necessarily higher than in other areas, as some sites in the cities of Musashino and Mitaka (western Tokyo) are more expensive than those in Adachi, Katsushika, and

Edogawa wards in northeast Tokyo. The number of official points and their position are highly biased by region. The histogram of official land prices in Figure 2.4 shows that the land prices in some areas are very high, forming a right-skewed distribution.

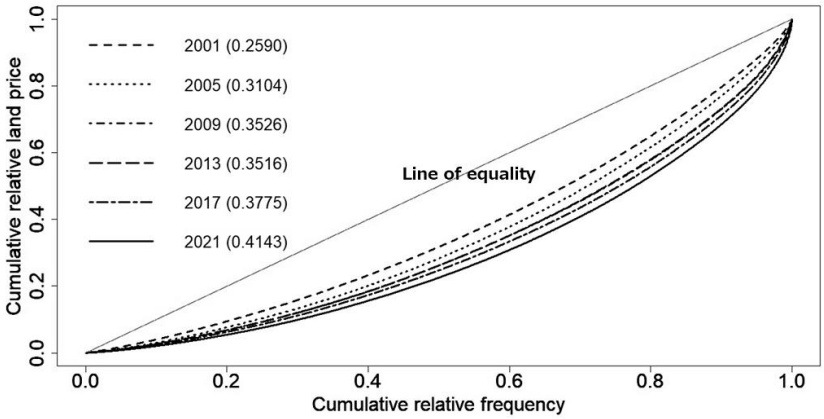


Figure 2.5: Plots of Lorenz curve together with Gini indexes in Tokyo land prices for 2001, 2005, 2009, 2013, 2017, 2021.

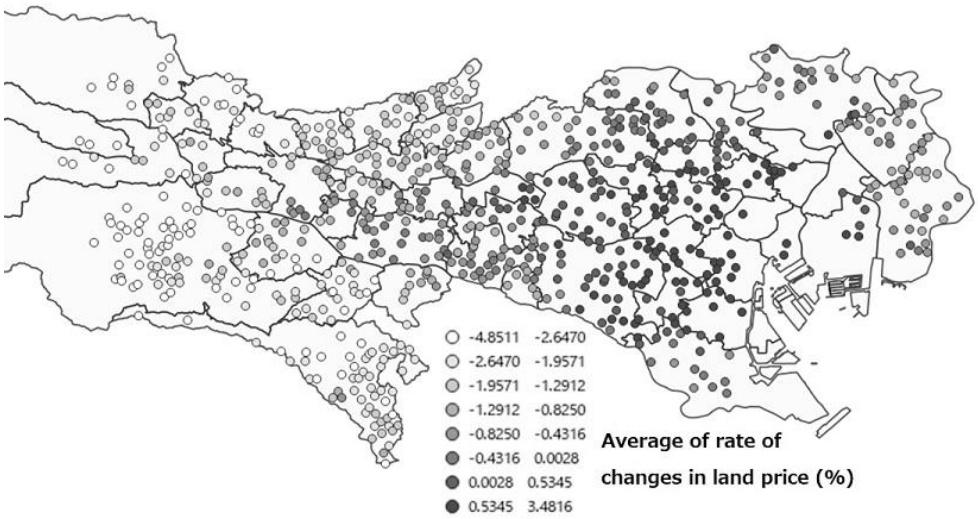


Figure 2.6: Spatial distribution of the average of the rate of changes in Tokyo land prices.

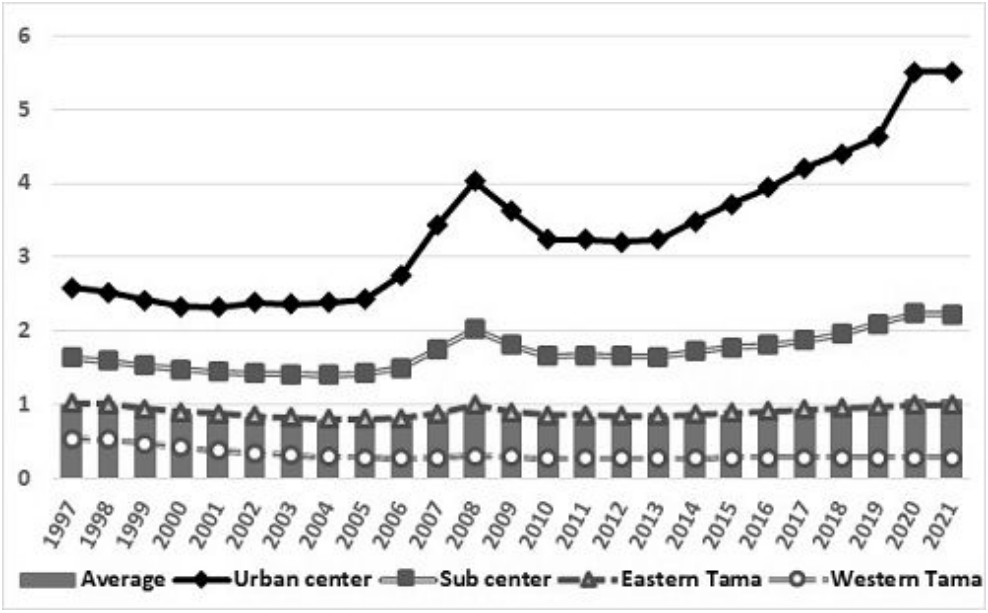


Figure 2.7: Spatial distribution of the average of the rate of changes in Tokyo land prices.

To see how inequality changes in the land price distribution in Tokyo, we plot a Lorenz curve for every four years after 2001 in Figure 2.5. The Gini indices are shown in the figure below. We can see that spatial inequality increases with time. The Japanese Gini index after income redistribution was around 0.372 in 2017, which is similar to the Tokyo land price distribution value. Figure 2.6 shows the spatial distribution of the average rate of change in Tokyo's land prices. The average rate of change is the average year-on-year percentage of 857 locations in the study with official land prices for 22 years from 1997 to 2018. As can be seen from this figure, the rate of changes in land prices appears in the clusters: central Tokyo remains high compared with other areas such as western Tokyo. Figure 2.7 shows the transitions of average land prices in the city center (Chiyoda Ward, Chuo Ward, Minato Ward), sub-centers (Shinjuku Ward, Shibuya Ward, Toshima Ward), eastern Tama (Musashino City, Mitaka City, Chofu City), and western Tama (Hachioji City) from 1997 to 2021, with the average land price in residential areas in Tokyo in 1997 set to 1. From this figure, we can see that the polarization of land prices and the widening disparity in land prices are progressing in Tokyo as well. We found that land price dispersion occurs in regional clusters. The

average official land price in the target area was 412,200 yen per square meter in 1997 and 432,900 yen in 2021, but the standard deviation was 182,900 yen in 1997 and 506,000 yen in 2021. These facts indicate that Tokyo's spatial inequality and dispersion have increased over the last two decades.

3

Study of Land Price Model in Geostatistics

3.1 Methodology

The universal kriging method used in this study is explained in detail in Cressie (1993). This section briefly describes the model and method used for parameter estimation.

The logarithmic land price at discrete sites $\boldsymbol{s}_i\,(i = 1, 2, \ldots, n)$ is denoted by $Y_t(\boldsymbol{s}_i)$ at time $t = 1, 2, \ldots, T$. It is assumed that official land prices can be expressed as asset characteristics of these sites $m_t(\boldsymbol{s}_i) = \boldsymbol{X}_t^{'}(\boldsymbol{s}_i)\boldsymbol{\beta}_t$ and the error term. For the land price process, we consider a continuous spatial process, $\{Y_t(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, where D is the target region. With universal kriging, $Y_t(\boldsymbol{s})$ can be expressed as

$$Y_t(\boldsymbol{s}) = \boldsymbol{X}_t^{'}(\boldsymbol{s})\boldsymbol{\beta}_t + u_t(\boldsymbol{s}). \tag{3.1.1}$$

Assuming first- and second-order spatial stationarity on the residual of universal kriging $u_t(\boldsymbol{s}_i)$, the spatial variogram is defined as follows:

$$2\gamma_t(\boldsymbol{h}) = Var[u_t(\boldsymbol{s}_i + \boldsymbol{h}) - u_t(\boldsymbol{s}_i)], \quad \forall \boldsymbol{s}_i, \boldsymbol{h} \in D. \tag{3.1.2}$$

Here, we assume first- and second-order intrinsic stationarity for the difference between two arbitrary sites; that is,

$$E[u_t(\boldsymbol{s}_i + \boldsymbol{h}) - u_t(\boldsymbol{s}_i)] = 0. \tag{3.1.3}$$

We also assume that the residual process $u_t(\boldsymbol{s})$ is independent of the time points; that is,

$$Cov[u_s(\boldsymbol{s}_i), u_t(\boldsymbol{s}_i + \boldsymbol{h})] = 0, \quad \text{for } \ t \neq s. \tag{3.1.4}$$

If we assume second-order stationarity of the residuals for spatial processes, then the covariance function, as the expected value of the residuals $m_t(u_{s_i})$, is defined as

$$\begin{aligned} C_t(\boldsymbol{h}) &= Cov[u_t(\boldsymbol{s}_i), u_t(\boldsymbol{s}_i + \boldsymbol{h})] \\ &= E[\{u_t(\boldsymbol{s}_i) - m_t(u_{s_i})\}\{u_t(\boldsymbol{s}_i + \boldsymbol{h}) - m_t(u_{s_i+h})\}]. \end{aligned} \tag{3.1.5}$$

Where the equation 3.1.2 can be transformed as

$$\begin{aligned} Var[u_t(\boldsymbol{s}_i + \boldsymbol{h}) - u_t(\boldsymbol{s}_i)] &= Var[u_t(\boldsymbol{s}_i + \boldsymbol{h})] + Var[u_t(\boldsymbol{s}_i)] - 2Cov[u_t(\boldsymbol{s}_i + \boldsymbol{h}), u_t(\boldsymbol{s}_i)] \\ &= 2C_t(0) - 2C_t(\boldsymbol{h}). \end{aligned} \tag{3.1.6}$$

The relationship between the spatial variogram and the covariance function is expressed as $\gamma_t(\boldsymbol{h}) = C_t(\boldsymbol{0}) - C_t(\boldsymbol{h})$. We call $\gamma_t(\boldsymbol{h})$ a semi-variogram.

There are various proposals for the theoretical model of the spatial variogram corresponding to the covariance function. The shape of a spatial variogram is determined by three parameters: the nugget, sill, and range. The nugget is the semi-variance when the distance between points is close to $\boldsymbol{0}$, the sill represents the variance of a spatial process, and the value obtained by subtracting the nugget from the sill is called the partial sill. The range is the minimum $\boldsymbol{h}$ since the correlation between $u_t(\boldsymbol{s}_i)$ and $u_t(\boldsymbol{s}_i + \boldsymbol{h})$ is 0.

In this study, the following spherical model was adopted as the theoretical variogram model[2].

$$\gamma_t(\boldsymbol{h}) = \begin{cases} \tau_t^2 + \sigma_t^2 & (\text{if } \|\boldsymbol{h}\| > 1/\phi_t), \\ \tau_t^2 + \sigma_t^2 \left[\dfrac{3}{2}\phi_t \|\boldsymbol{h}\| - \dfrac{1}{2}(\phi_t \|\boldsymbol{h}\|)^3 \right] & (\text{if } 0 < \|\boldsymbol{h}\| \leq 1/\phi_t), \\ 0 & (\text{otherwise}). \end{cases} \tag{3.1.7}$$

where τ_t^2 denotes a nugget, σ_t^2 denotes a partial sill, $\tau_t^2 + \sigma_t^2$ denotes a sill, and $1/\phi_t$ denotes the range.

Assuming that the isotropy assumption holds in the spatial process, the empirical semi-variance in interval h_r, $\gamma_t^*(h_r)$ is given by

$$\gamma_t^*(h_r) = \frac{1}{\#N_r} \sum_{(i,j)\in N_r} \gamma_{t,i,j}^*(\boldsymbol{h}) = \frac{1}{2\#N_r} \sum_{(i,j)\in N_r} [\hat{u}_t(\boldsymbol{s}_i) - \hat{u}_t(\boldsymbol{s}_j)]^2, \tag{3.1.8}$$

where N_r is a set of sample pairs satisfying $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \approx h_r$ and $\#N_r$ is the number of sample pairs satisfying $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \approx h_r$. $\gamma_{t,i,j}^*$ measures the dissimilarity between the sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. To estimate the empirical variogram using the observed data, we use the Cressie-Hawkins robust estimator (Cressie and Hawkins, 1980):

$$\gamma_t^*(h_r) = \frac{\dfrac{1}{2}\left\{ \dfrac{1}{\#N_r} \displaystyle\sum_{(i,j)\in N_r} \sqrt{|\hat{u}_t(\boldsymbol{s}_i) - \hat{u}_t(\boldsymbol{s}_j)|} \right\}^4}{0.475 + \dfrac{0.494}{\#N_r}}, \tag{3.1.9}$$

which is found to be robust to outliers.

[2]The following are the names and model formulas of typical theoretical variograms other than spherical models:

$$\text{Exponential model;} \quad \gamma_t(\boldsymbol{h}) = \begin{cases} \tau_t^2 + \sigma_t^2[1 - \exp(-\phi_t\|\boldsymbol{h}\|)] & (\text{if } \|\boldsymbol{h}\| > 0), \\ 0 & (\text{otherwise}). \end{cases}$$

$$\text{Linear model;} \quad \gamma_t(\boldsymbol{h}) = \begin{cases} \tau_t^2 + \sigma_t^2 & (\text{if } \|\boldsymbol{h}\| > 1/\phi_t), \\ \tau_t^2 + \sigma_t^2\|\boldsymbol{h}\| & (\text{if } 0 < \|\boldsymbol{h}\| \leq 1/\phi_t), \\ 0 & (\text{otherwise}). \end{cases}$$

$$\text{Gaussian model;} \quad \gamma_t(\boldsymbol{h}) = \begin{cases} \tau_t^2 + \sigma_t^2[1 - \exp(-\phi_t^2\|\boldsymbol{h}\|^2)] & (\text{if } \|\boldsymbol{h}\| > 0), \\ 0 & (\text{otherwise}), \end{cases}$$

In this study, we adopt a spherical model with the smallest residual sum of squares.

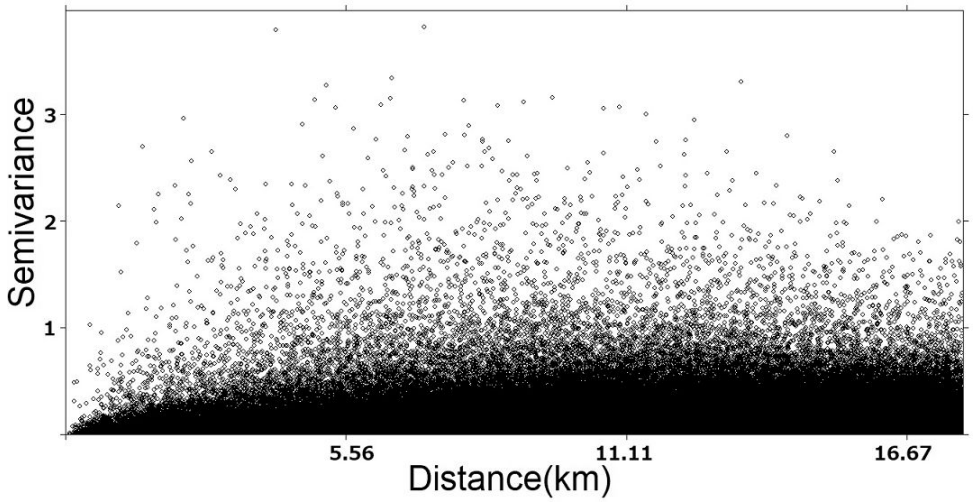


Figure 3.1: Variogram cloud of logarithmic Tokyo official land prices in 2018.

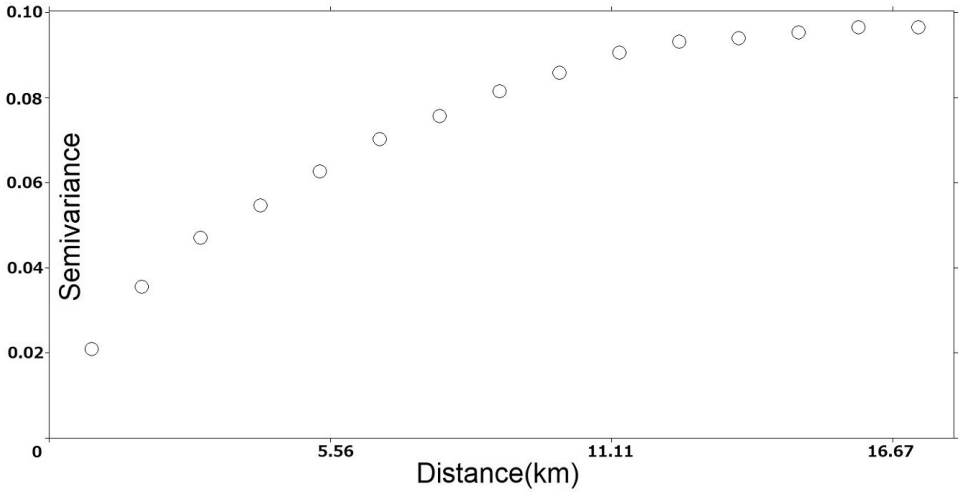


Figure 3.2: Plots of the empirical variogram in 2018.

We obtained a variogram cloud by plotting dissimilarity as a function of distance h. Figures 3.1 and 3.2 show the variogram cloud and empirical variogram of the regression residuals of the 2018 official land price data. From these figures, we can see that the larger the semi-variance, the larger the dissimilarity between the observation points. From the variogram cloud, we can see that as the distance between the observation points increases, the semi-variance also increases, and

its variation has a nonlinear relationship with the distance. In addition, the empirical variogram shows that, as the distance increases, the dissimilarity of the observed values converges to a constant value.

The vector of the parameters of the theoretical variogram is denoted by $\boldsymbol{\theta}_t = (\tau_t, \sigma_t, \phi_t)^{'}$, and the parameters of the theoretical variogram are estimated using the nonlinear weighted least squares (NWLS) method introduced by Cressie (1985). The theoretical and empirical variograms are expressed as γ_t and γ_t^*, respectively.

$$\hat{\boldsymbol{\theta}}_{t,\mathrm{NWLS}} = \mathrm{argmin}_{\boldsymbol{\theta}_t} \sum_{r=1}^{R} [Var\{\gamma_t^*(h_r)\}]^{-1} [\gamma_t^*(h_r) - \gamma_t(h_r|\boldsymbol{\theta}_t)]^2. \tag{3.1.10}$$

The weighting of the variance term is applied in the following equation:

$$Var\{\gamma_t^*(h_r)\} \approx 2\gamma_t(h_r|\boldsymbol{\theta}_t)^2/\#N_r. \tag{3.1.11}$$

To estimate the regression coefficients and variogram parameters in universal kriging, we use the iterative re-weighted generalized least squares (IRWGLS) method, introduced by Schabenberger and Gotway (2005) and Seya and Tsutsumi (2014).

IRWGLS method

Step 1 We denote the log land price and data matrix by $\boldsymbol{y}_t$ and $\boldsymbol{X}_t$, respectively, for time $t(1 \leq t \leq T)$, and obtain the estimated value of $\boldsymbol{\beta}_t$ by least squares method $\hat{\boldsymbol{\beta}}_{t,\mathrm{OLS}}$.

Step 2 To derive the residual vector $\boldsymbol{y}_t - \boldsymbol{X}_t^{'}\hat{\boldsymbol{\beta}}_{t,\mathrm{OLS}}$, we calculate the empirical variogram from the obtained residuals.

Step 3 Using the nonlinear weighted least squares method, we apply the theoretical variogram to the empirical variogram to obtain $\hat{\boldsymbol{\theta}}_{t,\mathrm{NWLS}}$.

Step 4 We calculate the covariance function from the estimated theoretical variogram, assume that the variance-covariance function is known, and use the Estimated GLS (EGLS) method to estimate the trend parameter $\hat{\boldsymbol{\beta}}_{t,\mathrm{IRWGLS}}$.

Step 5 We repeat steps [Step 3] through [Step 5] until $\boldsymbol{\beta}_t$ and $\boldsymbol{\theta}_t$ converge.

The IRWGLS estimator of $\boldsymbol{\beta}_t$ is the EGLS estimator, which is given by the following equation as a result of iterative calculations.

$$\hat{\boldsymbol{\beta}}_{t,\mathrm{IRWGLS}} = [\boldsymbol{X}_t \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_{t,\mathrm{IRWGLS}})^{-1} \boldsymbol{X}_t^{'}]^{-1} \boldsymbol{X}_t \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_{t,\mathrm{IRWGLS}})^{-1} \boldsymbol{y}_t, \tag{3.1.12}$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix.
The variance-covariance function is given by

$$Var(\hat{\boldsymbol{\beta}}_{t,\mathrm{IRWGLS}}) = [\boldsymbol{X}_t \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_{t,\mathrm{IRWGLS}})^{-1} \boldsymbol{X}_t^{'}]^{-1}. \tag{3.1.13}$$

3.2 The Data

The 2018 official land prices were obtained from 26,000 sample points distributed across all 47 prefectures of Japan. Of this, 20,572 points are located in urban areas, 1,394 points in urbanization control areas, 4,015 points in other city planning areas, and 19 points outside the city planning areas. Of the 2,602 points in Tokyo, 1,540 were in residential areas, which were the focus of this study, excluding the Tokyo Islands. We used 38,914 points in the Tokyo residential areas between 1997 and 2018.

Table 3.1: List of variables and their overviews.

Dependent variable	Summary
Official land prices (yen/m^2)	Official land price data for residential areas in Tokyo from 1997 to 2018.
Explanatory variables	
1, Access index (minutes)	Access index of the target points. A weighted average of the commuting time from the closest station to the 6 terminal stations by the number of passengers.
2, Distance to station (m)	Distance to the closest major station.
3, Front road width (m)	Front road width. For two more roads, I selected the maximum width road.
4, Floor area ratio (decimal)	Floor area ratio.
5, Land area (m^2)	Acreage of the target area.
6, South-headed	Dummy variable. Front road facing south, southeast or southwest: 1, others: 0.
7, Driveway	Dummy variable. The front road is a driveway: 1, others: 0.
8, Gus equipment	Dummy variable. Presence: 0 or absence: 1 of gas facility.
9, Sewerage	Dummy variable. Presence: 0 or absence: 1 of sewerage facility.
10, Low-rise residential	Dummy variable. Low-rise residencial area: 1, others: 0.
11, Residential/quasi-residential	Dummy variable. Residential or quasi-residencial area: 1, others: 0.
12, Quasi-fire prevention	Dummy variable. Quasi–fire prevention area: 1, others: 0.
13, Fire prevention	Dummy variable. Fire prevention area: 1, others: 0.

To exclude the influence of the bubble economy period and the years after its collapse, we used the official land prices of residential areas on the 1st of January every year between 1997 and 2018 to perform a spatial statistical analysis of the coefficients and residuals in the land price model. The explanatory variables were selected based on Hasegawa et al. (2006): the access index of the target points (Access), distance to the closest major station (Distance), front road width (Width), floor area ratio (Floor), and acreage of the target area (Area). In addition, the presence or absence of gas and sewerage facilities were added as dummy variables alongside a few other dummy variables. All variables, except for the floor-area ratio and dummy variables, were transformed to logarithmic values to estimate the land price model. Table 3.1 provides the list of explanatory variables and their overview[3].

Table 3.2 shows the summary statistics of the access index (min), distance from the closest station (m), front road width (m), and land area (m^2) as of 2018. The access index shows the convenience of access to central Tokyo by train. For this variable, we used the shortest average travel time from the closest station to the six major stations, Shinjuku, Ikebukuro, Tokyo, Shibuya, Ueno, and Shinagawa[4]. Data on the distance from the station, front road width, and land area were obtained from official land price announcements published by the MLIT. As shown in Table 3.2, the average time from Tokyo to each terminal station was 42min; the average distance from the closest station was 1,100m; the average front road width was 5.8m; the average land area was $212m^2$. In addition, the skew to the right distributions of distance from the closest station, front road width, and land area was observed with several large outliers.

[3]We do not use population-related explanatory variables because they make it difficult to interpret the causal relationships of the land price model. In addition, for gas and sewerage dummies, the case where there is no equipment is set to 1 to pay attention to the decline in land prices because they do not correspond.

[4]Weighted average of the average number of passengers getting on and off at each station each year. As a specific derivation method, the arrival at the terminal station is set to 8:45 am on weekdays, and where there are multiple routes, the route with the shortest time, including transfer, is selected. "YAHOO! JAPAN route information" is used for the railway travel time to each terminal station.

Table 3.2: Summary statistics for access index, distance from the closest station, front road width, and land area in 2018.

	Access	Distance	Width	Area
Number of samples	1,540	1,540	1,540	1,540
Mean	42.046	1096.5077	5.8255	212.0265
Standard error	0.3501	19.4006	0.06686	5.2694
Median	39	800	5	165
Mode	61.2157	1200	4	165
Standard deviation	17.8511	989.2427	3.4094	268.6872
Kurtosis	-0.455	15.2321	11.6238	235.4916
Skewness	0.4778	3.2265	25.9309	12.1622
Max	106.9304	9500	40	7367
Min	8.6933	90	2	40

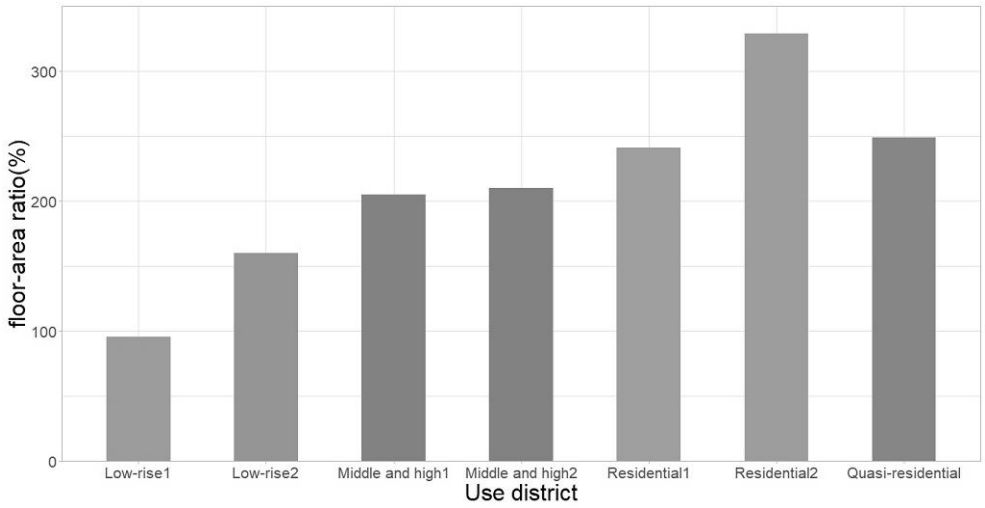


Figure 3.3: Average floor-area ratio by use district in 2018.

Figure 3.3 is a graph of the average floor-area ratio (%) by use district in 2018. The floor-area ratio of low-rise exclusive residential districts of categories 1 and 2, represented by upscale residential areas, was low. Conversely, the residential and quasi-residential districts of categories 1 and 2 have high floor-area ratios. If the regression coefficient for the floor-area ratio is positive, the higher the floor-area ratio, the higher the land prices, and vice versa.

The dummy variables used in this study include the south-headed (southD), driveway (dwD), gas equipment (gD), sewerage (sD), low-rise residential (lowrD), residential/quasi-residential (rD), quasi-fire prevention (qfireD) and fire prevention (fireD). These data were obtained from the information published along with the official land price data released by the MLIT. If the regression coefficient of the "South-headed dummy" is positive, then good sun exposure increases land prices. For the "Driveway dummy", if the front road is a driveway, then the land rights relationship between land owners become complicated; hence, if the regression coefficient is negative, land prices reduce. For the "Gas equipment dummy" and "Sewerage dummy," if their regression coefficient is negative, the respective area requires infrastructure development; therefore, the land prices drop. The "Low-rise residential dummy" represents the low-rise exclusive residential districts of categories 1 or 2, a common designation in the so-called "upscale residential areas". Therefore, if the regression coefficient for the low-rise residential dummy variable is positive, the respective area is considered to have higher land prices than other use districts because of the importance of the living environment. The "Residential/quasi-residential dummy" is designated near stations and along large roads, representing residential districts of categories 1 or 2 and quasi-residential district. The building coverage and floor area ratio are set higher than those in other use districts, and the range of building use is wide. Therefore, if the regression coefficient for the Residential/quasi-residential dummy variable is positive, the land price is assumed to be higher than in other regions because of the preference for convenient transportation and commercial use. The "Use district dummy" was based on the medium-to-high rise exclusive residential districts of categories 1 or 2.

The "Quasi-fire prevention dummy" and "Fire prevention dummy" represent quasi-fire prevention districts and fire prevention districts, respectively, where building restrictions such as fireproof buildings and quasi-fireproof buildings are applied, which generally increase construction costs. However, at the same time, because it improves safety, the land price of the corresponding region is not necessarily low. If the regression coefficient for the "Quasi-fire prevention dummy" and "Fire prevention dummy" variables is positive, it is concluded that the land prices are high due to higher safety restrictions; if it is negative, the land

prices are considered low because of the decision to avoid higher construction costs.

3.3 Model and Spatial Variogram Estimations and their Transition

In this section, we investigate the land price model using the official land prices in Tokyo. Since official land prices of residential areas in Tokyo vary drastically according to the survey sites and the covariance structure of the residuals of the land price model has regional characteristics, the assumption of second-order spatial stationarity does not hold for all regions. For this reason, we apply a cluster analysis that divides the data into four clusters using a hierarchical clustering method with Ward's linkage. In this process, we calculate the distances between municipalities using the position of the center of gravity and the average and variance of the land prices of each municipality.

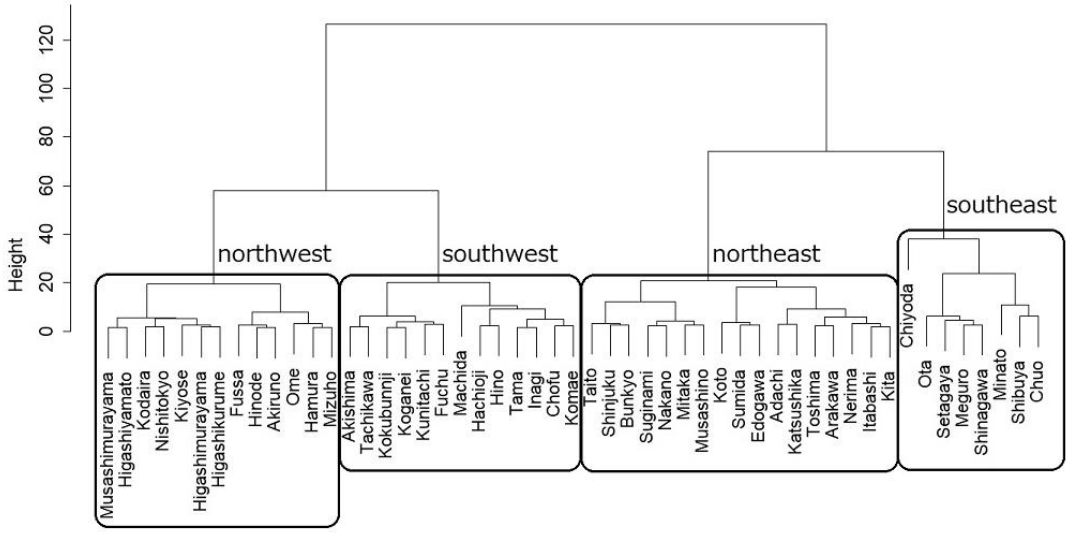


Figure 3.4: Dendrogram of the Tokyo municipalities.

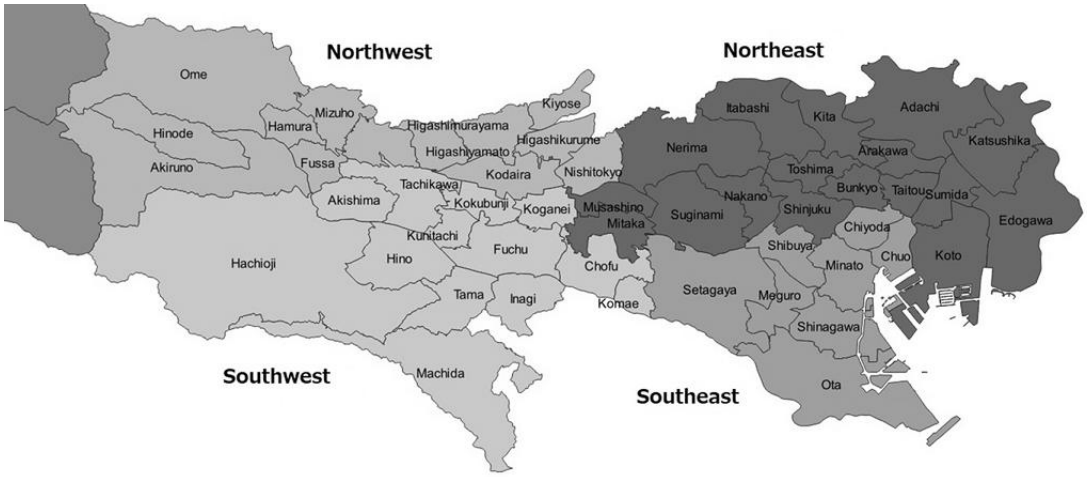


Figure 3.5: Spatial clustering of the Tokyo municipalities dividing into four regions.

Figures 3.4 and 3.5 show a dendrogram of the hierarchical clustering analysis and a map of the resulting four clusters in Tokyo, respectively. The town of Okutama and the village of Hinohara (the areas located west and municipality names are missing) were excluded because the official land prices of their residential areas were not evaluated during the analysis period. The southeast is a region that includes Chiyoda, Minato, and Chuo Wards, the so-called heart of the capital that concentrates government offices, commerce, and head office of large corporations, as well as the "Jonan district", which contains popular high-class residential areas such as Shinagawa, Meguro, and Ota Wards. The northeast is a region that includes the part located in the center of the 23 wards like Shinjuku, Toshima, and Bunkyo Wards, which still preserve the atmosphere of a "sub-center" and "downtown" aimed at decentralizing central functions, as well as the cities of Musashino and Mitaka, which, despite not being included in the 23 wards, are among the most desired cities to live in according to recent surveys. The regional transportation in the southwest, which includes the Minamitama (South Tama area) and the southern Kitatama (South part of North Tama area), became remarkably more convenient with the recent urban redevelopment. Hachioji has the largest population In the municipalities of the Tama area in Tokyo, and the city of Tama has a remarkable population increase rate. The northwest includes

the Nishitama area and northern Kitatama area (North and West parts of North Tama area), and the plains and hills are developing into Tokyo's commuter towns. It is an area that features rare landscapes in Tokyo, such as fields, hills, and mountains.

To detect a global trend in the four regions, we chose five explanatory variables– access index, distance from the closest station, front road width, floor area ratio, and land area –at five different times, that is, five and ten years before and after the 2008 financial crisis (1998, 2003, 2008, 2013 and 2018). We performed the Kruskal-Wallis H test on all these years [5]. As a result, for the access index and front road width, the null hypothesis that the median of all five-time points is the same cannot be rejected in any of the four regions with a significance level of 5%. In the distance to the closest station, a significantly negative trend was observed in the northeast area. Presumably, this is due to the extension of railroad lines and the construction of new ones, such as the Fukutoshin Line (No. 13, opened in 2008), as well as changes in official price points targeting standard land. For the floor area ratio, the null hypothesis could not be rejected in the southeast and northeast, but a positive trend was observed in the southwest and northwest. This may be due to changes in land regulations due to the development of the "Tama Site Development Basic Plan" and the replacements of the official price points that followed. In the land area, a negative trend was observed in the northeast and southwest regions. This is also attributed to changes in the land price points caused by changes in the definition of standard land over time.

[5]Even if a significant difference is found in the Kruskal-Wallis H test, it cannot be known which group has a significant difference. Therefore, the positive and negative of the trend are judged from the transition of the group's median.

Table 3.3: Estimated regression coefficient of non-spatial models in 2018.

Explanatory variable	Whole Tokyo	Southeast	Northeast	Southwest	Northwest
Intercept	16.9877 (0.1501)	15.3838 (0.3391)	15.6013 (0.1682)	20.5621 (0.3123)	18.0458 (0.3721)
Access index	−1.0021 (0.0209)	−0.7520 (0.0532)	−0.6107 (0.0275)	−1.5838 (0.0548)	−1.1237 (0.0581)
Distance to station	−0.2542 (0.0113)	−0.1852 (0.0282)	−0.1909 (0.0153)	−0.3069 (0.0139)	−0.2151 (0.0177)
Front road width	0.1082 (0.0222)	0.0575 (0.0437)	0.0691 (0.0255)	0.0523 (0.0317)	0.0633 (0.0422)
Floor area ratio	0.0205 (0.0217)	0.2177 (0.0409)	0.0356 (0.0187)	0.1076 (0.0706)	0.0683 (0.0786)
Land area	0.1337 (0.0149)	0.2211 (0.0230)	0.1003 (0.0157)	−0.0419 (0.0327)	−0.0275 (0.0341)
South-headed	0.0038 (0.0136)	0.0642 (0.0303)	0.0391 (0.0164)	−0.0135 (0.0198)	0.0113 (0.0242)
Driveway	−0.0172 (0.0248)	−0.1517 (0.0911)	0.0604 (0.0343)	−0.0238 (0.0337)	−0.0247 (0.0296)
Gus equipment	−0.2801 (0.0269)			−0.1878 (0.0329)	−0.1441 (0.0337)
Sewerage	−0.3217 (0.1343)			−0.3084 (0.1074)	
Low-rise residential	0.0819 (0.0279)	0.1453 (0.0521)	0.0840 (0.0253)	0.1015 (0.0756)	−0.0007 (0.0805)
Residential/quasi-Residential	−0.0280 (0.0233)	−0.2141 (0.0452)	−0.0358 (0.0224)	−0.0241 (0.0418)	−0.0759 (0.0563)
Quasi-fire prevention	0.1339 (0.0197)	−0.0331 (0.0619)	−0.1104 (0.0319)	0.1633 (0.0401)	0.0169 (0.0316)
Fire prevention	0.2908 (0.0574)	0.0035 (0.0954)	−0.1666 (0.0603)		
Adjusted R^2	0.8550	0.7849	0.7142	0.8013	0.8378

Table 3.3 shows the OLS estimates of 2018 land prices. Recall that the OLS method constitutes the global model that has non-spatial structures. The numbers in parentheses are the standard errors. Some regions do not have gas equipment dummy or sewerage dummy parameters because those facilities had already been built in the entire official target area before 2018, while others do not have the fire prevention dummy because no fire prevention district was selected as a target area for the official land prices that year. The characteristics of each region become clear when the respective estimated coefficients are compared to those of Tokyo as a whole. It is possible to identify that the larger regional differences in the parameters of access index, distance to station, floor-area ratio, land area, residential/quasi-residential dummy and quasi-fire prevention dummy exist. The access index and distance to station coefficients show that the official land prices in the west have a greater negative impact than those in the east, and the floor-area ratio coefficient shows that the positive impact of floor-area ratio on the official land pricesis largest in the southeast. Regarding the land area coefficient, the fact that the southeast and northeast have positive coefficients indicates no volume discount effect. Also, the residential/quasi-residential dummy coefficient is negative in all four regions, but a significant negative estimate was only obtained in the southeast. This implies that commercial elements negatively affect housing preferences in the southeast. The quasi-fire prevention dummy coefficient has a significant negative impact at the 5% level in the northeast and a significant positive impact at the 5% level in the southwest. From this, it can be inferred that in the northeast, the negative impact of the increase in construction costs decreases the official land prices; in the southwest, security and safety increases the official land prices. These facts show that the factors determining official land prices are not too different from our intuitive understandings.

Table 3.4: Estimated coefficients of the spatial process model by IRWGLS in 2018.

Explanatory variable	Whole Tokyo	Southeast	Northeast	Southwest	Northwest
Intercept	14.1047 (0.3123)	13.3544 (1.0087)	14.3610 (0.2664)	17.1242 (0.4202)	17.7954 (0.4450)
Access index	−0.3736 (0.0436)	−0.2132 (0.1583)	−0.2960 (0.0436)	−1.0353 (0.0737)	−1.1048 (0.0695)
Distance to station	−0.1535 (0.0234)	−0.0890 (0.0839)	−0.1587 (0.0243)	−0.1799 (0.0187)	−0.2163 (0.0211)
Front road width	0.0885 (0.0461)	0.1145 (0.1299)	0.1099 (0.0405)	0.0416 (0.0426)	0.0763 (0.0504)
Floor area ratio	0.0195 (0.0451)	−0.0282 (0.1217)	0.0104 (0.0297)	0.1326 (0.0950)	0.0894 (0.0940)
Land area	0.0728 (0.0310)	0.1411 (0.0683)	0.0430 (0.0249)	0.0101 (0.0440)	−0.0379 (0.0408)
South-Headed	0.0079 (0.0297)	−0.0114 (0.0900)	0.0197 (0.0260)	0.0082 (0.0266)	0.0145 (0.0289)
Driveway	−0.0145 (0.0515)	0.0129 (0.2711)	−0.0063 (0.0544)	−0.0304 (0.0454)	−0.0144 (0.0354)
Gus equipment	−0.0674 (0.0560)			−0.0594 (0.0442)	−0.0917 (0.0403)
Sewerage	−0.1527 (0.2793)			−0.1711 (0.1444)	
Low-Rise Residential	0.0289 (0.0579)	0.0174 (0.1551)	0.0336 (0.0401)	0.0919 (0.1018)	0.0793 (0.0963)
Residential/Quasi-Residential	−0.0043 (0.0484)	−0.0495 (0.1346)	0.0104 (0.0355)	0.0579 (0.0563)	0.0175 (0.0673)
Quasi-Fire Prevention	0.0675 (0.0411)	−0.0392 (0.1841)	−0.0205 (0.0506)	0.0596 (0.0539)	0.0164 (0.0378)
Fire Prevention	0.0950 (0.1194)	−0.0099 (0.2838)	−0.0598 (0.0956)		

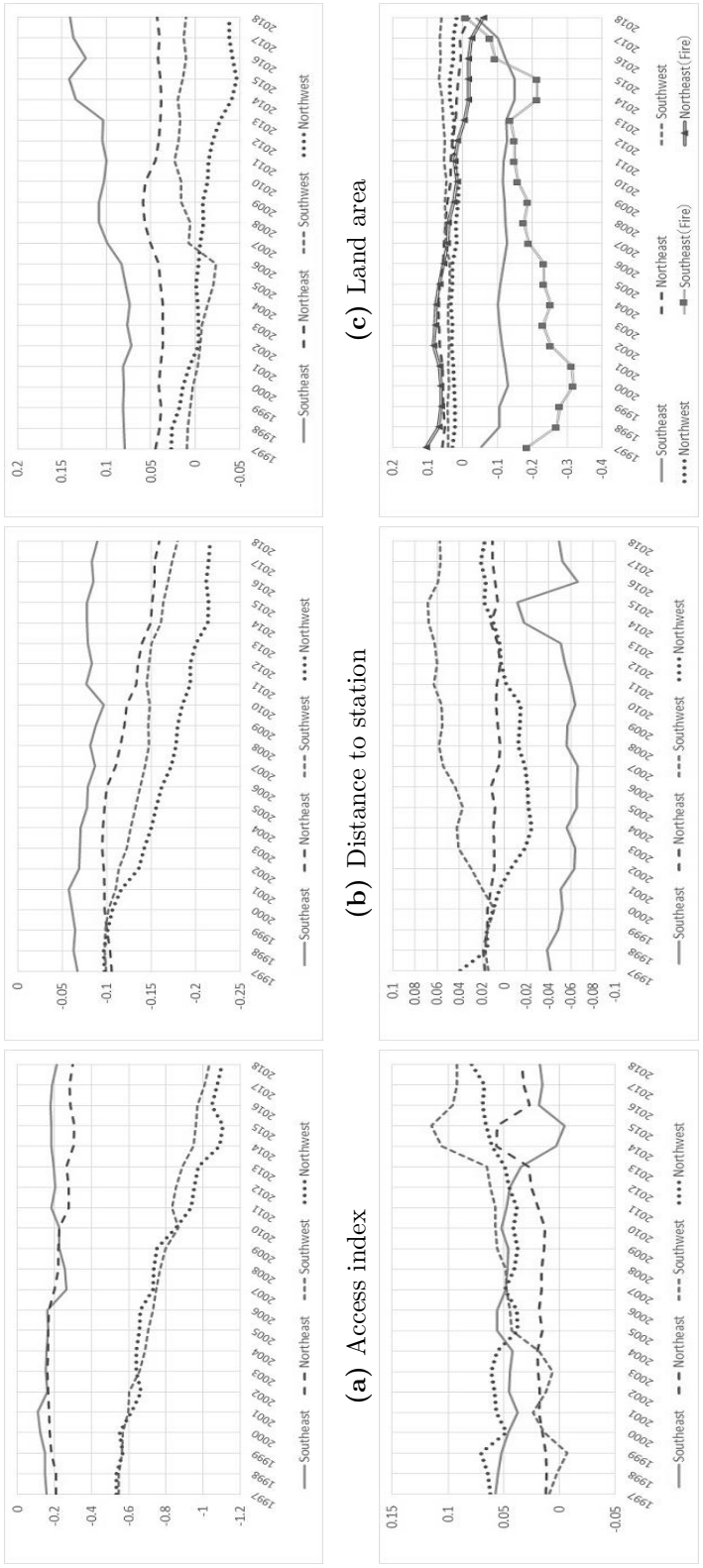
Table 3.4 summarizes the 2018 land prices estimated parameters using the IRWGLS method. Compared to the results of the non-spatial model, the absolute values of the spatial process model estimates are smaller, and the standard error is larger. Compared to Table 3.3, the estimated coefficients of the floor-area ratio and low-rise residential dummy in the southeast are no longer significantly positive at the 5% level, and the same applies to the low-rise residential dummy in the northeast. In the southwest, the residential/quasi-residential dummy coefficient becomes positive, and the difference with the official land prices of low-rise exclusive residential districts decreases. Also, the fact that the land area coefficient changes from negative to positive indicates that, despite not being statistically significant, there is no volume discount effect. In the northwest, the low-rise residential dummy coefficient changes to positive, differentiating itself from the other use districts. Most of the signs of the other coefficients match those of the OLS. Following previous studies, we adopted a spatial variogram with a spherical model and estimated the covariance structure with a valid distance[6] of 6.67 km in the southeast, 5.56 km in the northeast, and 8.89 km in the southwest and northwest.

Table 3.5: Mean square errors of the non-spatial and spatial process models in 2018.

	Southeast	Northeast	Southwest	Northwest
MSE				
Non–spatial model	0.0490	0.0315	0.0428	0.0279
Spatial process model	0.0167	0.0074	0.0130	0.0094

Table 3.5 shows the mean square errors (MSE) of the non-spatial and spatial process models for each of the four regions in 2018. The MSE of the land price models shows that the MSE of the spatial process model was reduced to about 1/5 to 1/3 of the MSE of the non-spatial model. From the above, it is judged that the spatial process model fits better than the non-spatial model.

[6]The distance between points required to estimate the spatial variogram. The behavior of the variogram becomes unstable because the number of points decreases when the distance between points is long. In this study, we visually determined the spatial variogram at the first iteration of IRWGLS method to determine its effective distance.



(a) Access index (b) Distance to station (c) Land area

(d) Low-residential dummy (e) Residential/quasi-residential dummy (f) Quasi-fire/fire prevention dummy

Figure 3.6: Transitions of regression coefficients.

3.3 Model and Spatial Variogram Estimations and their Transition

To see the time-series transitions of the estimated coefficients, we provide the following figures and tables. Figure 3.6 shows the annual changes in the regression coefficients of the access index, distance from the closest station, land area, low-rise residential dummy, residential/quasi-residential dummy, and quasi-fire prevention/fire prevention dummy of the estimated spatial process model.

Table 3.6: Test statistics and its p-values for the null hypothesis of no slope in the linear regression with time trend for access index, distance from station, residential dummy, residential/quasi-residential dummy and quasi fire/fire prevention dummy.

Explanatory variable	Tokyo	Coefficient	$p$value	90%lower	90%upper
Access index	Southeast	−0.0034	0.0091	−0.0052	−0.0013
	Northeast	−0.0068	0.0000	−0.0085	−0.0051
	Southwest	−0.0245	0.0000	−0.0256	−0.0235
	Northwest	−0.0308	0.0000	−0.0338	−0.0277
Distance to station	Southeast	−0.0012	0.0000	−0.0016	−0.0008
	Northeast	−0.0032	0.0000	−0.0038	−0.0027
	Southwest	−0.0038	0.0000	−0.0041	−0.0035
	Northwest	−0.0063	0.0000	−0.0068	−0.0059
Land area	Southeast	0.0032	0.0000	0.0026	0.0038
	Northeast	0.0001	0.5227	−0.0002	0.0005
	Southwest	0.0009	0.0258	0.0003	0.0016
	Northwest	−0.0032	0.0000	−0.0036	−0.0029
Low-residential	Southeast	−0.0020	0.0001	−0.0027	−0.0013
	Northeast	0.0013	0.0005	0.0008	0.0019
	Southwest	0.0052	0.0000	0.0045	0.0059
	Northwest	0.0002	0.6350	−0.0005	0.0009
Residential/quasi-residential	Southeast	0.0003	0.5503	−0.0005	0.0011
	Northeast	−0.0004	0.0028	−0.0006	−0.0002
	Southwest	0.0027	0.0000	0.0022	0.0032
	Northwest	0.0002	0.7212	−0.0008	0.0013
Fire prevention	Southeast	−0.0005	0.5405	−0.0020	0.0009
	Northeast	−0.0033	0.0000	−0.0040	−0.0025
	Southwest	0.0013	0.0000	0.0010	0.0015
	Northwest	0.0000	0.9736	−0.0004	0.0004
	Southeast (fire)	0.0093	0.0000	0.0066	0.0120
	Northeast (fire)	−0.0063	0.0000	−0.0071	−0.0054

Table 3.6 summarizes the test results of the null hypothesis of the zero slopes in the linear regression for these variables with the explanatory variable time t. In this table, the "90% lower limit" and the "90% upper limit" represent the lower and upper limits of the 90.

The transition of the access index coefficients can be divided into eastern and western Tokyo. That is, in the southeast and northeast, the coefficient remained mostly stable with a small negative value throughout the analysis period; however, in the southwest and northwest, it followed a downward trend. These results indicate that in the western part of Tokyo, the longer time required to access the city center has a stronger negative impact on official land prices and the gap with the eastern part of Tokyo is widening every year. It can be seen that the coefficient of distance from the closest station is negative in all regions throughout the analysis period, and the three regions other than the central part (southeast) have negatively smaller values than the southeast. In summary, the further from the central area in Tokyo and the closest station, the lower the land prices tend to be, and the impact of these coefficients increase yearly.

The estimated land area coefficient was positive in the southeast in Tokyo. This indicates that there is no volume discount effect, and the larger the land area, the more expensive it is. This trend is stronger in the southeast than in other regions, and it increases even further over time, which represents the increasing brand power of certain regions.

The low-rise residential dummy coefficient shows a negative trend in the southeast. The difference from the standard – the medium-to-high residential exclusive districts – has disappeared, which suggests that because of the high levels of concentration in the city center and saturation of land usage, the tendency to emphasize living environment elements has decreased. On the other hand, a positive trend can be seen in the southwest, a sign that more people tend to prioritize living environment elements.

As for the residential/quasi-residential dummy coefficient, the southeast has a negative value, as regions with strong commercial elements have lower official land prices than other use districts. In contrast, a positive trend can be seen in the southwest, where there is a growing tendency to emphasize convenience and

commercial elements. This result is likely attributed to recent land redevelopment and improvement plans in the Tama area.

Regarding the coefficient of the quasi-fire/fire prevention dummy, the southeast has a negative coefficient and lower prices than other restricted regions. This suggests that this area avoids increases in construction costs due to restrictions, but their impact has been decreasing. In addition, a positive trend can be seen in the coefficient of the quasi-fire prevention dummy of the southwest, where, contrary to the southeast, higher levels of safety and security due to restrictions tend to push the official land prices up. It can be seen that, in the first half of the analysis, the fire prevention districts in the southeast lower the official land prices sharply, but that impact lowers in the latter half, and it has been almost non-existent in recent years. Moreover, it can be seen that the effect of the other coefficients other than the quasi-fire prevention dummy in the southwest and fire prevention dummy in the northeast on official land prices lessens over the years.

In panel (d) of Figure 3.6, there is a discontinuity in the regression coefficient of the low-rise residential dummy (except in the northwest) of 2013 and 2014, and the same in the regression coefficients of the residential/quasi-residential dummy and fire prevention dummy of the southeast in panels (e) and (f). A discontinuity can also be seen between 2015 and 2016. Presumably, this is the effect of a decrease in the number of official target points in 2014 and an increase in 2016. Following the changes in the official land price target areas, there was a large change in the ratio of the number of points in low-rise exclusive residential districts and residential/quasi-residential districts in the southeast[7].

[7]The proportion of low-rise residential areas in the southeast changed to 48.3%, 46.7%, 46.7% and 48.3% from 2013 to 2016. The percentages of residential/quasi-residential areas changed to 29.2%, 30.7%, 30.7%, and 29.1%. No other year was observed when the proportion of such use areas fluctuated by 1.6%.

Table 3.7: Estimated variogram parameters of the spatial process models in 2018.

	Southeast	Northeast	Southwest	Northwest
Variogram parameter				
Nugget	0.0013	0.0054	0.0086	0.0051
Partialsill	95.6808	0.0842	0.0397	0.0256
Range(km)	9183	26.90	7.66	8.54

Table 3.7 shows the spatial variogram estimates for each of the four regions in the spatial process models for 2018. The range and partial sill values of the southeast are large because the spatial variogram approached a linear relationship within the valid distance, and the linear approximation part in the spherical model with a small distance difference was fitted to the empirical variogram. Considering that the larger the ratio of sill and nugget effect or range, the stronger the spatial correlation of the random fields, it becomes clear that the residuals of the spatial process model of eastern Tokyo have a stronger spatial correlation than the west side. The nugget effect of the southwest is larger than that of the southeast and northeast, mainly composed of 23 wards where the official land prices are relatively high.

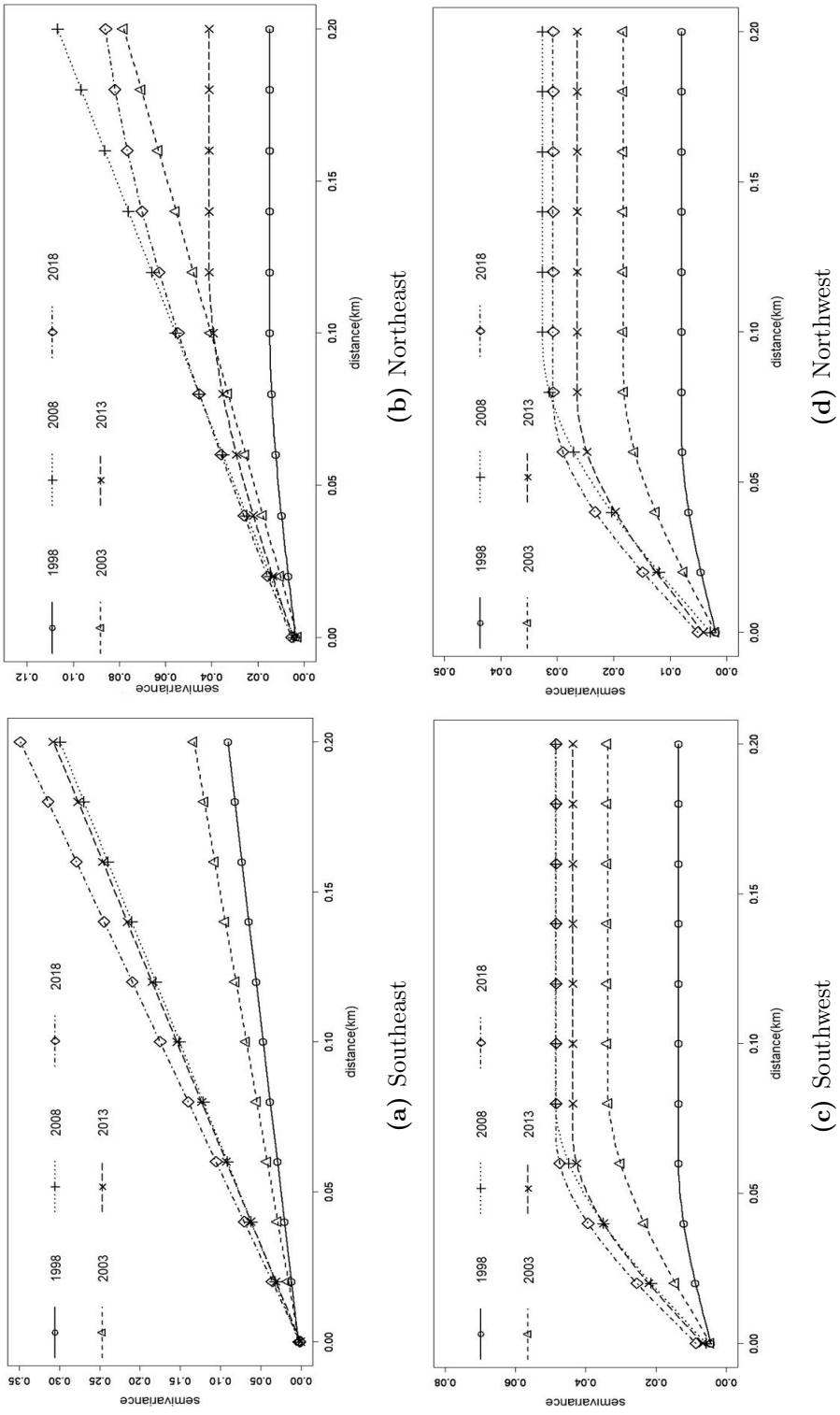


Figure 3.7: Changes in the variograms for four regions.

Figure 3.7 shows the variation of the estimated spatial variogram of each region for five different years. Like before, to provide better visualization of the annual changes, we chose the year 2008 (owing to the financial crisis) and 5 and 10 years before and after it (i.e., 1998, 2003, 2008, 2013, and 2018). The linear parts show that when the change rate increases, the degree of attenuation of the covariance function also increases. Furthermore, increases in the variogram sill represent increases in the variance of the spatial process of the residuals. The graphs show that in all four regions, the variogram sill of spatial process model residuals was increasing until the 2008 financial crisis; then, it temporarily decreased after the crisis except for the southeast and has been increasing again in recent years, reaching the same level before the 2008 financial crisis. In addition, the shape of the variograms in the four regions changes over time, indicating the difficulty of dividing the region and performing spatiotemporal analysis that assumes secondary stationarity not only in the spatial direction but also in the time series direction (Inoue et al., 2009).

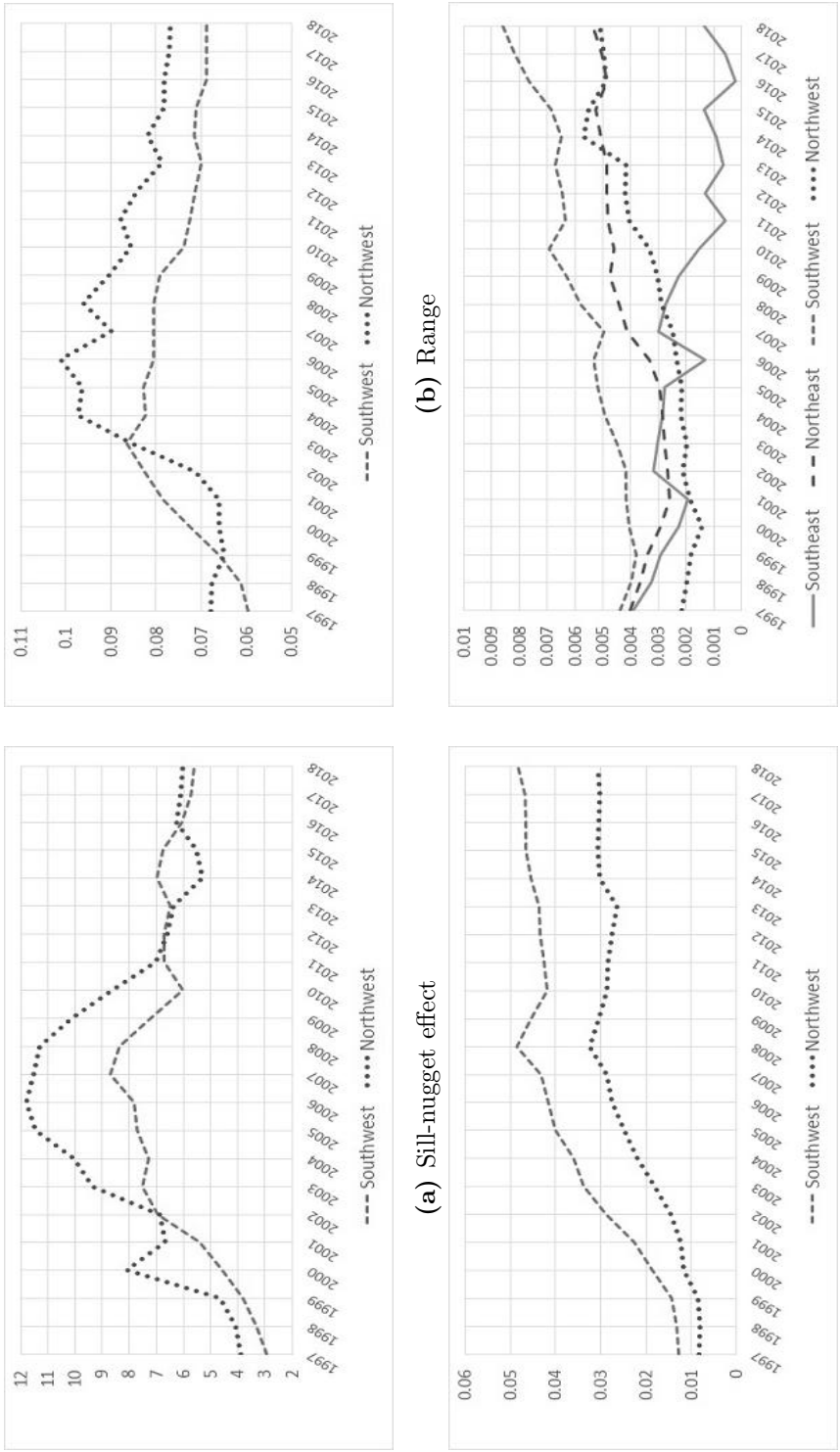


Figure 3.8: Transitions of the estimated variogram parameters in Tokyo

Table 3.8: Test Statistics and its p-values for the null hypothesis of no slope in the linear regression with time trend for variogram parameters.

Tokyo	Period	Coefficient	$p$value	90%lower	90%upper
Sill–nugget effect of western Tokyo					
Southwest	1997 –2008	0.5532	0.0000	0.4628	0.6436
	2008 –2018	−0.1750	0.0083	−0.2703	−0.0798
Northwest	1997 –2008	0.7973	0.0000	0.6446	0.9501
	2008 –2018	−0.4890	0.0015	−0.6892	−0.2888
Range of western Tokyo					
Southwest	1997–2008	0.0020	0.0021	0.0011	0.0029
	2008–2018	−0.0011	0.0003	−0.0014	−0.0007
Northwest	1997–2008	0.0036	0.0002	0.0025	0.0047
	2008–2018	−0.0018	0.0000	−0.0022	−0.0013
Sill of western Tokyo					
Southwest	1997–2008	0.0035	0.0000	0.0032	0.0038
	2008–2018	0.0003	0.2482	−0.0001	0.0006
Northwest	1997–2008	0.0024	0.0000	0.0021	0.0026
	2008–2018	0.0000	0.9151	−0.0003	0.0003
Nugget					
Southeast		−0.0001	0.0000	−0.0002	−0.0001
Northeast		0.0001	0.0000	0.0001	0.0002
Southwest		0.0002	0.0000	0.0002	0.0002
Northwest		0.0002	0.0000	0.0002	0.0002

Figure 3.8 shows the value of the sill-nugget effect ratio of the regression residuals of the spatial process models in western Tokyo, the range and sill of the same region and the annual changes in the nuggets of the four regions in Tokyo. Also, Table 3.8 shows the test result of the rate of change of the linear regression of these parameters. In all tests of the rate of change, except for the nugget, it is possible to see a trend inflection point; for this reason, we performed the test on the period before the 2008 financial crisis and another after it. From the annual changes in the sill-nugget effect ratio and range, it is possible to see that the spatial correlation of the spatial process model residuals in western Tokyo was becoming stronger up until the 2008 financial crisis but became weak after it. A similar change in spatial correlation can be seen in the southwest and northwest, but it fluctuates more intensely in the northwest.

The sill represents the variance of a spatial process, according to its definition. A larger variance indicated a higher degree of variation in the regression residuals of the spatial process model. The fact that western Tokyo takes a statistically significant positive value, even at the 1% level, until the 2008 financial crisis demonstrates that the variation increased every year. After the 2008 financial crisis, no trend was identified, and the variation in the residuals remained stable.

The nuggets of the four regions in Tokyo show a downward trend in the southeast and an overall upward trend in the other regions. The nugget indicates a local variation of the residuals between two land price model observation points. It is empirically known that real estate has strong individuality, and the rise in the nugget indicates that the land price dispersion is relatively becoming stronger, which cannot be explained by explanatory variables of the land price model in this analysis[8]. During the analysis period, the land price dispersion became stronger in the three regions, except for the southeast, where they became weaker. Regarding the trends in each region, in the early downturn phase of official land prices of the analysis period, the nugget is also in a downward trend in all four regions; however, it switched to an uptrend in all regions except the southeast around 2002. In addition, the southeast has been on a downward trend regardless of the 2008 financial crisis, but the three other regions do not show a downward trend, not even in the downturn phase of prices that followed.

In summary, in the southeast, the regression coefficient of the distance from the closest station had the smallest negative impact on official land prices among the four regions. Moreover, the factors that increase prices in low-rise exclusive residential districts are becoming weaker every year, and the regulations of quasi-fire/fire prevention districts have also been affecting prices less and less every year. On the other hand, the preference for brands has caused the official land prices in the southeast to soar, a tendency that grows stronger every year. Furthermore, the effect of land price dispersion, one factor that determines land prices, has also decreased. In the northeast, the negative impact of the regression coefficients of the distance to the closest station on the land prices has been increasing since

[8]The land price dispersion include alert areas in hazard maps, the presence of crime, local noises, daylight conditions, and the location of garbage collection points; however, the effect of these variables on land prices is a topic for future research.

around 2006, as well as the impact of the land price dispersion. In the southwest, the negative impact of the regression coefficients of the access index and distance to the closest station on land prices has been increasing. In addition, the official land prices of use districts other than medium-to-high-rise exclusive residential districts tend to be higher. This characteristic is not observed in other regions and is also becoming stronger yearly. Moreover, the fact that quasi-fire prevention regulations elevate official land prices more strongly yearly is also characteristic. The strength of the spatial correlation of residuals showed a remarkable uptrend until the 2008 financial crisis, but after that, it switched to stable movement as the individual factors of the neighboring land increased. In the northwest, the regression coefficients of the access index and distance to the closest station have a stronger negative impact on the official land prices than in the southwest. The impact of the coefficient of the distance to the closest station, in particular, was the largest among the four regions. It is the only region where the volume discount effect can be identified after the 2008 financial crisis. The strength of spatial correlation and the impact of the land price dispersion have shown a trend similar to that of the southwest.

4

Study of Land Price Model in Spatial Econometrics

4.1 Models and Methodology

We denote $\boldsymbol{y}_t(\boldsymbol{s})$ as the logarithmic public land price vector of site $\boldsymbol{s} \in D$ in region D at time t. The global or nonspatial model can then be expressed as follows:

$$\boldsymbol{y}_t(\boldsymbol{s}) = \boldsymbol{X}_t^{'}(\boldsymbol{s})\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t(\boldsymbol{s}), \tag{4.1.1}$$

where $\boldsymbol{X}_t(\boldsymbol{s})$ is the matrix of the explanatory variables described in the previous section, $\boldsymbol{\beta}_t$ is the vector of the regression coefficients including the constant term, and $\boldsymbol{\varepsilon}_t(\boldsymbol{s})$ is the error term, which is assumed to be independent at time t and for site $\boldsymbol{s}$. $\boldsymbol{x}'$ denotes the transpose of the matrix or vector $\boldsymbol{x}$. The regression coefficients of the non-spatial models were common for all sites.

GWR uses location-wise estimates to model spatially variable relationships. Let $y_t(\boldsymbol{s}_i)$ be the logarithmically transformed official land price at each site $\boldsymbol{s}_i$ with n observation sites $(\boldsymbol{s}_i, i = 1, \ldots, n)$ at time t. Furthermore, let the k-dimensional vector of the explanatory variable be $\boldsymbol{X}_t(\boldsymbol{s}_i) = [1, x_{1,t}(\boldsymbol{s}_i), \cdots, x_{k-1,t}(\boldsymbol{s}_i)]^{'}$. We denote the vector of local regression coefficients by $\boldsymbol{\beta}_{t,i}(k \times 1)$ and the error term by $\varepsilon_t(\boldsymbol{s}_i)$. Then, the GWR model can be expressed as

$$y_t(\boldsymbol{s}_i) = \boldsymbol{X}_t^{'}(\boldsymbol{s}_i)\boldsymbol{\beta}_{t,i} + \varepsilon_t(\boldsymbol{s}_i). \tag{4.1.2}$$

Under GWR, to estimate $\boldsymbol{\beta}_{t,i} = [\beta_{0,t,i}, \beta_{1,t,i}, \cdots, \beta_{k-1,t,i}]^{'}$, all observations are weighted by the distance to i. We use the generalized least squares (GLS) method with the following weight matrix:

$$\boldsymbol{V}_{t,i}^{\frac{1}{2}} \boldsymbol{y}_t(\boldsymbol{s}) = \boldsymbol{V}_{t,i}^{\frac{1}{2}} \boldsymbol{X}_t^{'}(\boldsymbol{s}) \boldsymbol{\beta}_{t,i} + \boldsymbol{V}_{t,i}^{\frac{1}{2}} \boldsymbol{\varepsilon}_t(\boldsymbol{s}). \tag{4.1.3}$$

Here, matrix $\boldsymbol{V}_{t,i}$ is a diagonal matrix, and its j-th component $v_{t,i,j}$ is the weight given to site j:

$$\boldsymbol{V}_{t,i} = \operatorname{diag}\left(v_{t,i,1}, \ldots, v_{t,i,n}\right). \tag{4.1.4}$$

The estimator of the local regression coefficient at site i and time t is given by

$$\hat{\boldsymbol{\beta}}_{t,i} = \left[\boldsymbol{X}_t(\boldsymbol{s})\, \boldsymbol{V}_{t,i} \boldsymbol{X}_t^{'}(\boldsymbol{s})\right]^{-1} \boldsymbol{X}_t(\boldsymbol{s})\, \boldsymbol{V}_{t,i} \boldsymbol{y}_t(\boldsymbol{s}). \tag{4.1.5}$$

In the GWR model, it is important to define weight matrix $\boldsymbol{V}_{t,i}$. To this end, we used the Gaussian distance-decay function[9]:

$$v_{t,i,j} = \exp\left(-\frac{d_{i,j}^2}{\delta_t^2}\right), \tag{4.1.6}$$

where $d_{i,j}$ is the Euclidean distance between i and j. δ_t is the bandwidth of a common spatial kernel at time t. The value of δ_t has a trade-off relationship between bias and variance. If δ_t is too small, it can be estimated without bias because it uses only local data, but since there is little data that can be used, the variance of the estimation becomes large. On the other hand, if δ_t is too large, it can be estimated stably because it uses a wide range of data, but it may contain data with different structures, which causes a bias in the estimation. Bandwidth

[9]The other decay functions are shown below.

$$\text{Bisquare model}, v_{t,i,j} = \begin{cases} \left[1 - \left(\frac{d_{i,j}}{\delta_t}\right)^2\right]^2 & (if\ d_{i,j} < \delta), \\ 0 & (otherwise). \end{cases}$$

$$\text{Tricube model}, \ v_{t,i,j} = \begin{cases} \left[1 - \left(\frac{d_{i,j}}{\delta_t}\right)^3\right]^3 & (if\ d_{i,j} < \delta), \\ 0 & (otherwise). \end{cases}$$

In this study, we adopt the Gaussian type, which has the best root mean square error.

δ_t is determined by minimizing the Cross-Validation error (CV) of the following equation:

$$\hat{\delta}_t = \underset{\delta}{\operatorname{argmin}} \operatorname{CV}(\delta_t), \quad \operatorname{CV}(\delta_t) \quad = \sum_{i=1}^{n} \left[y_{t,i} - \hat{y}_{t,\neq i}(\delta_t) \right]^2, \tag{4.1.7}$$

where $\hat{y}_{t,\neq i}(\delta_t)$ is the predicted value for the site i by a point near site i, without using site i. Other methods for determining δ_t include the modified Akaike Information Criterion (Fotheringham et al., 2002). If the spatial distribution of the observed points is not constant, an adaptive kernel that adjusts the bandwidth according to the number of samples, not the distance, may be used; see, for example, Lu et al. (2015).

GWR can also be used for spatial prediction at points where the observed values are unknown. Let the explanatory variable for site $\boldsymbol{s}_0$ be $\boldsymbol{X}_t(\boldsymbol{s}_0)$. Then, the predicted value of the log official land price becomes:

$$\hat{y}_t(\boldsymbol{s}_0) = \boldsymbol{X}_t^{'}(\boldsymbol{s}_0)\hat{\boldsymbol{\beta}}_{t,s_0}, \tag{4.1.8}$$

$$\hat{\boldsymbol{\beta}}_{t,s_0} = [\boldsymbol{X}_t(\boldsymbol{s}_0)\boldsymbol{V}_{t,s_0}\boldsymbol{X}_t^{'}(\boldsymbol{s}_0)]^{-1}\boldsymbol{X}_t(\boldsymbol{s}_0)\boldsymbol{V}_{t,s_0}\boldsymbol{y}_t(\boldsymbol{s}). \tag{4.1.9}$$

For example, see Leung et al. (2000) and Harris et al. (2011). The corresponding variance of the prediction error becomes:

$$\begin{aligned} Var[y_t(\boldsymbol{s}_0) - \hat{y}_t(\boldsymbol{s}_0)] = [1 + \boldsymbol{X}_t(\boldsymbol{s}_0)[\boldsymbol{X}_t(\boldsymbol{s}_0)\boldsymbol{V}_{t,s_0}\boldsymbol{X}_t^{'}(\boldsymbol{s}_0)]^{-1} \\ \times \boldsymbol{X}_t(\boldsymbol{s}_0)\boldsymbol{V}_{t,s_0}^2\boldsymbol{X}_t^{'}(\boldsymbol{s}_0)[\boldsymbol{X}_t(\boldsymbol{s}_0)\boldsymbol{V}_{t,s_0}\boldsymbol{X}_t^{'}(\boldsymbol{s}_0)]^{-1}\boldsymbol{X}_t^{'}(\boldsymbol{s}_0)]\sigma_{\varepsilon,t}^2, \end{aligned} \tag{4.1.10}$$

where $\sigma_{\varepsilon,t}^2$ denotes the variance of the error term $\varepsilon_t(\boldsymbol{s}_0)$ for site $\boldsymbol{s}_0$.

Brunsdon et al. (1999) pointed out that, for the GWR and mixed GWR models, the bandwidths of the common spatial kernels are sometimes restrictive, and the resulting GWR estimates tend to be inflexible. Additionally, Wheeler and Tiefelsdorf (2005) explained that, under the GWR model, there exists instability that creates multicollinearity due to the similarities of local explanatory variables. Hence, Yang (2014) proposed the MGWR model, which applies a different bandwidth for each explanatory variable for the spatial kernels. The MGWR model can provide more location-specific regression surfaces, which makes it possible to avoid multicollinearity between variables. In this study, we use the following extended algorithm, as proposed by Lu et al. (2017).

Extended back-fitting algorithm[10].

Step 0 Data formatting: We denote the log land price and data matrix by $\boldsymbol{y}_{i,t}$ and $\boldsymbol{X}_t$, respectively, for time $t(1 \leq t \leq T)$ and site $i(1 \leq i \leq p)$. Let $\boldsymbol{V}_{k,t,i}^{(0)}$ be the initial weight matrix for t, i, and the kth regression coefficients in the GWR model. The initial kernel bandwidth is set to $bw_{k,t}^{(0)}$. The required precision is denoted by $\tau > 0$ and the maximum number of iterations is set as $\mathbb{N}$.

Step 1 Initialization: Initial estimates $\hat{\boldsymbol{\beta}}_t^{(0)} = [\hat{\boldsymbol{\beta}}_{0,t}^{(0)}, \hat{\boldsymbol{\beta}}_{1,t}^{(0)}, \cdots, \hat{\boldsymbol{\beta}}_{k-1,t}^{(0)}]'$ are obtained by the GWR model. Then, we calculate $\hat{\boldsymbol{y}}_{0,t}^{(0)} = \boldsymbol{X}_{0,t}^{'} \circ \hat{\boldsymbol{\beta}}_{0,t}^{(0)}$, $\hat{\boldsymbol{y}}_{1,t}^{(0)} = \boldsymbol{X}_{1,t}^{'} \circ \hat{\boldsymbol{\beta}}_{1,t}^{(0)}, \cdots,$ $\hat{\boldsymbol{y}}_{k-1,t}^{(0)} = \boldsymbol{X}_{k-1,t}^{'} \circ \hat{\boldsymbol{\beta}}_{k-1,t}^{(0)}$. Here, $\boldsymbol{X}_{h-1,t}$ denotes the h-th row of matrix $\boldsymbol{X}_t$ and $\circ$ is the Hadamard product. We obtain the residual sum of squares $\mathrm{RSS}^{(0)}$=$\sum(\boldsymbol{y}_t - \sum_{i=0}^{k-1} \hat{\boldsymbol{y}}_{i,t}^{(0)})^2$.

Step 2 Update the (n)-th estimates using the estimates of the $(n-1)$-th iteration as follows: we re-define the explanatory variable as $\boldsymbol{X}_{l,t}(0 \leq l \leq m)$.

[a] Calculate $\xi_{l,t}^{(n)} = \boldsymbol{y} - \sum_{j \neq l}^{m} Latestyhat\left[\hat{\boldsymbol{y}}_{j,t}^{(n-1)}, \hat{\boldsymbol{y}}_{j,t}^{(n)}\right]$, where $\sum_{j \neq l}^{m}$ denotes the sum of numbers other than l and

$$Latestyhat\left[\hat{\boldsymbol{y}}_{j,t}^{(n-1)}, \hat{\boldsymbol{y}}_{j,t}^{(n)}\right] = \begin{cases} \hat{\boldsymbol{y}}_{j,t}^{(n)}, if\ \hat{\boldsymbol{y}}_{j,t}^{(n)}\ exists \\ \hat{\boldsymbol{y}}_{j,t}^{(n-1)}, otherwise. \end{cases} \tag{4.1.11}$$

[b] We calculate bandwidth $bw_{k,t}^{(n)}$ using criteria such as the CV scoring method and obtain weight matrix $\boldsymbol{V}_{k,t,i}^{(n)}$.

Finally, we calculate $\hat{\boldsymbol{\beta}}_{l,t}^{(n)}$ using $\xi_{l,t}^{(n)}$ and $\boldsymbol{X}_{l,t}$.

[c] We update $\hat{\boldsymbol{y}}_{l,t}^{(n)} = \boldsymbol{X}_{l,t}^{'} \circ \hat{\boldsymbol{\beta}}_{l,t}^{(n)}$.

Step 3 Using $\hat{\boldsymbol{\beta}}_t^{(n)} = [\hat{\boldsymbol{\beta}}_{0,t}^{(n)}, \hat{\boldsymbol{\beta}}_{1,t}^{(n)}, \cdots, \hat{\boldsymbol{\beta}}_{k-1,t}^{(n)}]'$, we calculate $\hat{\boldsymbol{y}}_t^{(n)}$ and $\mathrm{RSS}^{(n)}$ and obtain rate of change $\mathrm{CVR}^{(n)}$:

$$\mathrm{CVR}^{(n)} = \frac{\mathrm{RSS}^{(n)} - \mathrm{RSS}^{(n-1)}}{\mathrm{RSS}^{(n-1)}}. \tag{4.1.12}$$

[10] A back-fitting algorithm is a simple iterative procedure used in statistics to fit a generalized additive model (Breiman and Friedman, 1985).

If $\mathrm{CVR}^{(n)} < \tau$ or $n \geq \mathbb{N}$, the calculation ends. Otherwise, $n = n + 1$ and the process is repeated.

4.2 The Data

Table 4.1: VIF for the explanatory variables in 2018 used in Chapter 3.

	Access	Dis	Width	Floor	Area	southD	dwD	gD	sD
VIF	2.021	1.350	1.513	9.007	1.298	1.009	1.067	1.187	1.018
	lowrD	rD	qfD	fD					
VIF	5.215	1.722	2.481	1.903					

Table 4.2: VIF for the explanatory variables in 2018 used in this chapter.

	Access	Dis	Width	Area	gD	lowrD	rD
VIF	1.468	1.285	1.372	1.258	1.180	1.671	1.494

As mentioned above, the GWR model has an inherent problem of local multicollinearity. Table 4.1 shows the 2018 explanatory variables' VIF used in Chapter 3. It can be seen that the VIF of the floor area ratio coefficient and the low-rise residential area dummy coefficient are high. This is because the floor area ratio is low in the low-rise residential areas. Furthermore, 86.3% of residential/quasi-residential areas are in quasi-fireproof areas. With the above in mind, this chapter uses the seven explanatory variables shown in Table 4.2[11]. Table 4.3[12] shows the explanatory variables and dependent variables used in this chapter.

[11] As a guide, the explanatory variables were selected so that the correlation coefficient between the explanatory variables would not be 0.8 and the VIF of 2.5 or more.

[12] For more information, see Section 3.2.

Table 4.3: List of variables and their overviews.

Dependent variable	Summary
Official land prices (yen/m^2)	Official land price data for residential areas in Tokyo from 1997 to 2018. Logarithmic transformation.
Explanatory variables	
1, Access index (minutes)	Access index of the target points. A weighted average of the commuting time from the closest station to the 6 terminal stations by the number of passengers. Logarithmic transformation.
2, Distance to station (m)	Distance to the closest major station. Logarithmic transformation.
3, Front road width (m)	Front road width. For two more roads, we selected the maximum width road. Logarithmic transformation.
4, Land area (m^2)	Acreage of the target area. Logarithmic transformation.
5, Low-rise residencial	Dummy variable. Low-rise residencial area: 1, others: 0.
6, Residential/quasi-residential	Dummy variable. Residential or quasi-residencial area: 1, others: 0.
7, Gas equipment	Dummy variable. Presence: 0 or absence: 1 of gas facility.

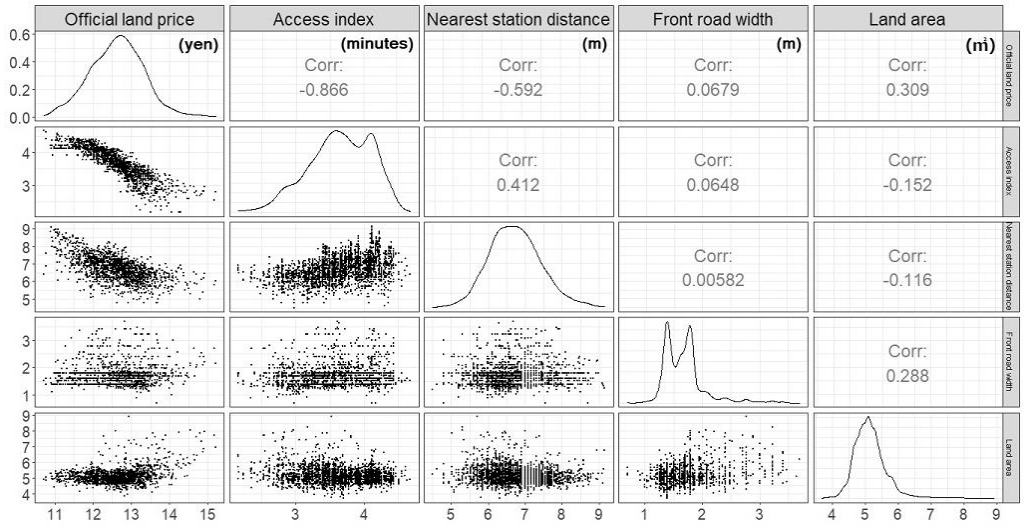


Figure 4.1: Scatterplots of the logarithmic transformations of Tokyo land price, access index, nearest station distance, front road width and land area. Kernel density plots are shown diagonally, while the graphs above the diagonal indicate the correlation coefficients.

Figure 4.1 shows a scatterplot matrix of public land prices, access index, nearest station distance, front road width, and land area. The diagonal figures show the kernel density plots of the variables after logarithmic transformation, and the upper ones show the correlation coefficients on the two corresponding variables. This figure shows that official land prices are negatively correlated with the access index and distance to the nearest station. The degree of correlation is stronger for the access index than for the nearest station distance. Additionally, a weak positive correlation exists between the access index and nearest station distance, land area and official land price, and front road width.

4.3 Model Estimation and its Transition

Here, we perform parameter estimation using the models presented in Section 4.1: the non-spatial model, GWR model, and MGWR model. A non-spatial model is a global regression model whose coefficients are common for all sites and is estimated using OLS. For GWR and MGWR models, the local regression parameters that show the spatial patterns and heterogeneity are estimated.

Table 4.4: Regression coefficients for the non-spatial and GWR models for 2018.

	Non–spatial model		GWR model(Kernel bandwidth 1.41 km)				
	$\hat{\beta}_i$	SE	min	first quantile	median	third quantile	max
Intercept	17.3990	0.1181	2.3637	14.1987	15.4629	18.1178	38.9184
Access	−1.0633	0.0182	−5.4377	−1.1618	−0.5631	−0.2583	1.5504
Distance	−0.2770	0.0112	−0.4836	−0.2392	−0.1772	−0.1208	0.1633
Width	0.1430	0.0218	−0.2991	0.0368	0.0916	0.1664	0.7921
Area	0.1378	0.0150	−0.7251	−0.0106	0.0488	0.1125	1.3120
lowrD	0.0098	0.0181	−0.6611	−0.1087	−0.0275	0.0456	0.5553
rD	−0.0099	0.0217	−0.6033	−0.0683	0.0153	0.0993	0.4892
gD	−0.2836	0.0273	−2.2197	−0.7126	−0.3128	−0.1007	1.1098

Table 4.5: Regression coefficients for the GWR and MGWR models for 2018.

	GWR model		MGWR model					
	mean	SD	min	first quantile	median	third quantile	max	KBwidth(km)
Intercept	16.4557	2.6373	12.9792	13.8973	15.1192	19.2707	21.2709	0.58
Access	−0.7830	0.6063	−1.6806	−1.4486	−0.4205	−0.1955	−0.1586	2.52
Distance	0.1695	0.0443	−0.2592	−0.1972	−0.1720	−0.1112	−0.0664	3.76
Width	0.0972	0.0143	0.0749	0.0833	0.0987	0.1102	0.1172	14.41
Area	0.0414	0.0758	−0.1907	−0.0063	0.0423	0.0891	0.2231	1.90
lowrD	−0.0065	0.0902	0.3621	−0.0373	0.0047	0.0463	0.1654	2.37
rD	0.0244	0.0302	−0.0277	−0.0027	0.0165	0.0478	0.0958	6.81
gD	−0.4706	0.3869	−1.1953	−0.8197	−0.4709	−0.0785	0.0500	3.38

Table 4.4 compares the regression coefficients of the non-spatial and GWR models. Under the non-spatial model, the low-rise residential and residential area dummies are insignificant at the 5% significance level. The local regression coefficient on the GWR model is estimated for each site, and there is a range in the distribution of the regression coefficients. If we compare the median values of the regression coefficients estimated by the GWR model, the absolute value of the estimates, significant under the non-spatial model, generally becomes smaller. Table 4.5 shows a comparison with the MGWR model. If we compare the regression coefficients on the median values, the estimates for the GWR and MGWR models take similar values but smaller absolute values under the non-spatial model. The range of each regression coefficient, which was large under the GWR model, is smaller under the MGWR model, probably because of the common bandwidth for spatial kernels for all explanatory variables under the GWR model. Specifically, this bandwidth might be too large or too small for each variable in the GWR model and was estimated adequately under a variable-specific bandwidth for spatial kernels in the MGWR model.

According to the signs of the first and third quartiles of the estimates under the GWR and MGWR models in Tables 4.4 and 4.5, the signs are reversed for the land area, low-rise residential area dummy, and residential/quasi-residential area dummy variables. Thus, the effects of these variables on land prices are locally different, indicating the complex heterogeneity of land prices. On the other hand, using quantile regression, Zietz et al. (2008) estimated the effects of land conditions to differ between the high- and low-price housing price ranges in Utah, the United States, and found that the signs of the lower and upper quantile regression coefficients are the same. Similarly, the sign inversion for the quantile regression coefficients was not confirmed by Shimizu et al. (2016), who analyzed prices in the real estate market of the Tokyo metropolitan area. Therefore, the MGWR model seems to somewhat overfit the land prices.

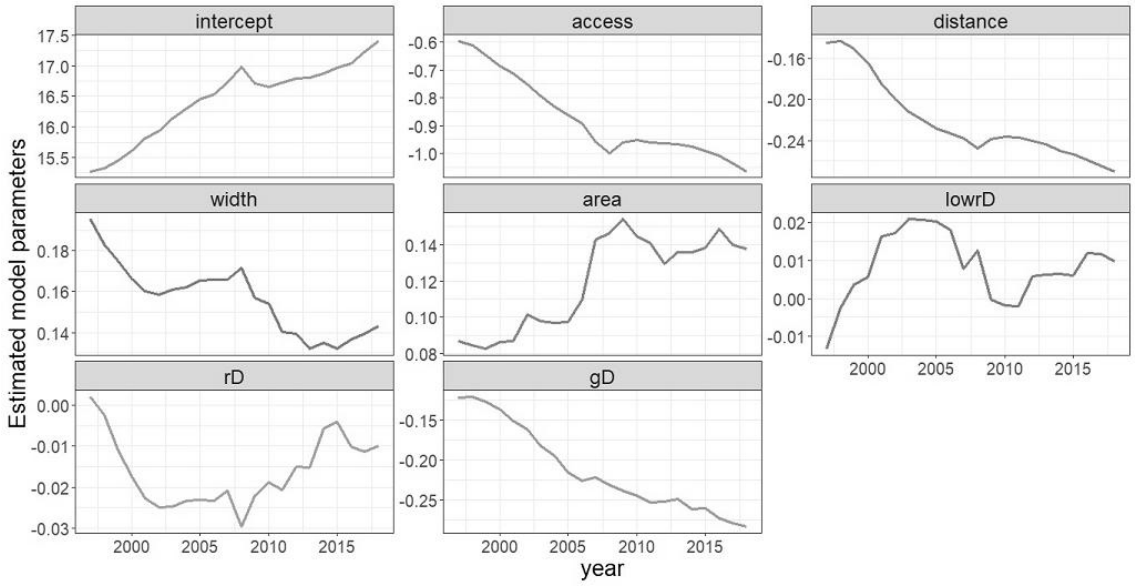


Figure 4.2: Transition of the estimated parameters for the non-spatial model.

Figure 4.2 shows the time-series transition of the estimated regression coefficients under the non-spatial model. A positive trend exists for the coefficients on the intercept and land area and a negative trend for those on the access index, distance to the nearest station, front road width, and gas equipment dummy. Additionally, the estimated values for the access index, distance from the nearest station, and regression coefficient on the gas equipment dummy became negative, affecting official land prices. The transitions of each explanatory variable under non-spatial models can be interpreted intuitively using our understanding of the determinants of land prices in general. Regarding the downward trend of the access index estimate, if we consider that the access index has almost no trend, the impact of price declines increases throughout the observation period for sites with poor access to the city center. Moreover, although the average distance to the nearest station is also on a downward trend, the negative impact on land prices is increasing, indicating that the negative impact on prices is increasing for sites far from the nearest station.

In the non-spatial model, the coefficient of land area becomes positive but has positive and negative values in both the GWR and MGWR models. Regarding the scale of the land area and the fact that negative corrections have been found

for the inheritance tax land price and property tax land price evaluations, Tabuchi (1996) indicate that small land areas have a premium when the land is difficult to divide or when developments that will expand land are expected. Thus, the effect of land area on land prices varies significantly locally to a great extent.

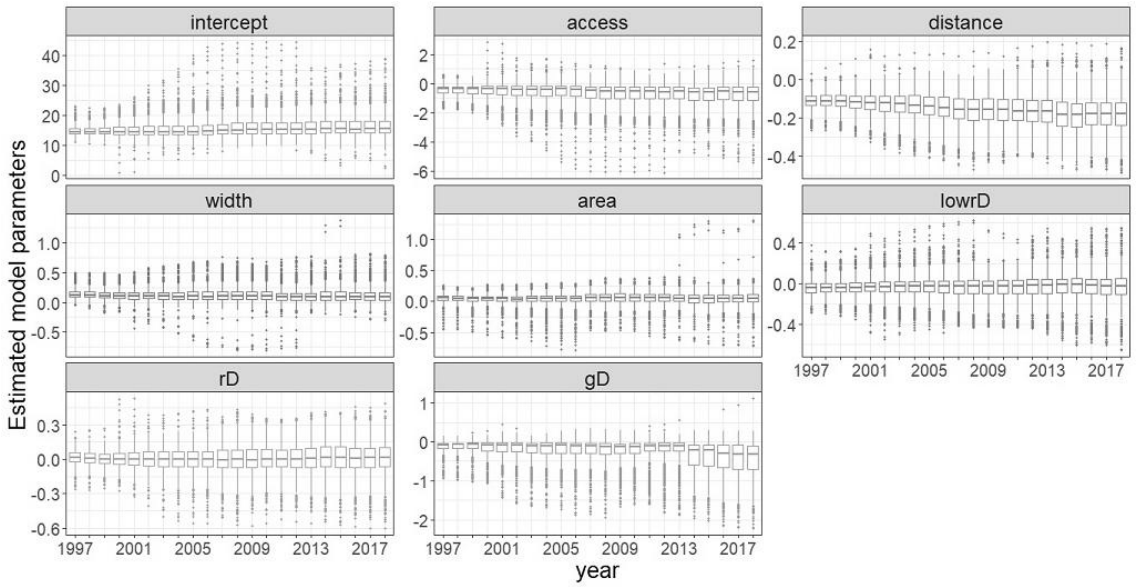


Figure 4.3: Boxplots of the transition for the estimated GWR model parameters.

Figure 4.3 shows the time-series transition of the estimated regression coefficients in the GWR model. The Gaussian distance-decay function was adopted, and the common bandwidth for the spatial kernels was determined using the CV scoring method, according to Equation 4.1.7. Except for the intercept, the range of the regression coefficients for the access index and gas equipment dummy is larger than that for the other regression coefficients. Outliers were also present for all regression coefficients. The regression coefficients on the access index, nearest station distance, and gas equipment dummy showed a negative trend in recent years, similar to the non-spatial model. This indicates that if both explanatory variables at two different points in time are at the same level, the effect of reducing land prices becomes stronger over time. No visual trend was observed in the coefficients of the other explanatory variables.

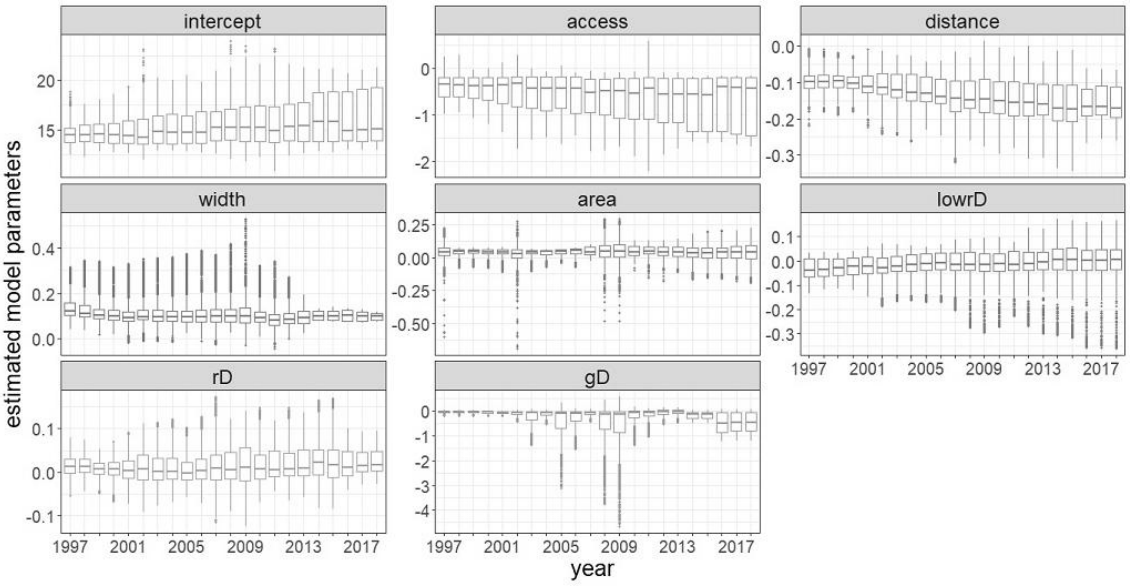
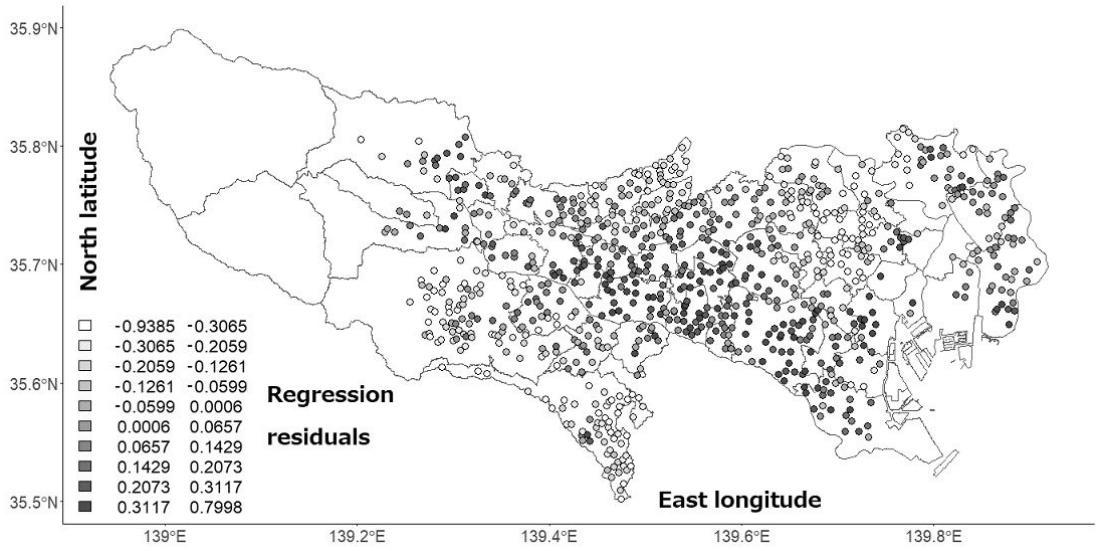


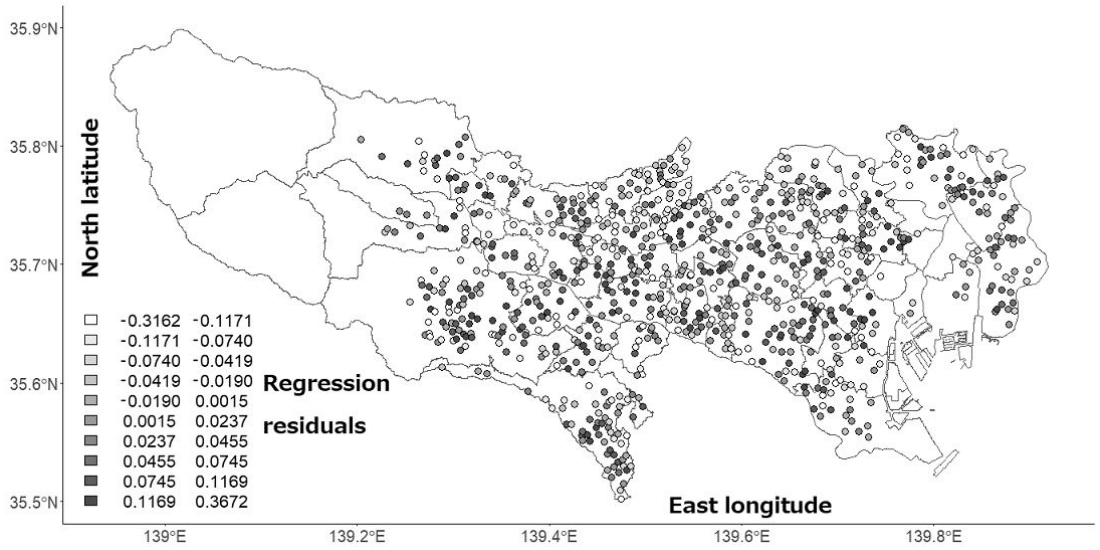
Figure 4.4: Boxplots of the transitions for the estimated parameters under the MGWR model.

Figure 4.4 shows the time-series transition of the estimated regression coefficients under the MGWR model using box plots. We used the algorithm in Lu et al. (2017) for parameter estimation. The bandwidth for the variable-specific spatial kernels was determined by converging CVR in Equation 4.1.12. The regression coefficients have a smaller range and fewer outliers than the GWR model. In addition to the access index and nearest station distance, a negative trend can be confirmed for the front road width and a positive trend for the low-rise residential area dummy. The ranges of the boxplots for the access index, nearest station distance, low-rise residential area dummy and residential area dummy become larger, indicating that the explanatory variable's dispersion becomes stronger with respect to land price. Additionally, the increase in the range of the constant terms means that the land price dispersion for the explanatory variables not used in this analysis is likely to increase [13].

[13]The dispersion of land prices, the so-called individuality of land, includes areas of caution on hazard maps, the existence of crime, local sunshine, noise conditions, and location of garbage collection sites.



(a) Non-spatial model



(b) GWR model

Figure 4.5: Spatial distribution of 2018 regression residuals for the non-spatial model and the GWR model.

Figure 4.5 shows the spatial distribution of the regression residuals of the non-spatial and GWR models for 2018. Equation 4.3.1 was used to calculate the

regression residuals.

$$\text{Regression residual} = y_t(\boldsymbol{s}_i) - \hat{y}_t(\boldsymbol{s}_i). \tag{4.3.1}$$

A significant spatial correlation was found in the residuals of the non-spatial model at the 5% level (Moran's I=0.5004); however, no significant spatial correlation was found in the residuals of the GWR model (Moran's I=0.0063). From the spatial distribution of the residuals in the non-spatial model, in addition to Chuo, Chiyoda, Minato, and Setagaya Ward, Chofu, Musashino, Mitaka, and Fuchu City have a large residual, which is less than the observed value. Kita, Itabashi, Toshima, Shinjuku Ward, Kiyose, Higashikurume, Higashimurayama, and Hachioji City had small residuals, and the results were overestimated from the observed values. Spatial correlation can be confirmed visually. From the spatial distribution of the residuals by the GWR model, the range of the regression residuals is smaller than the non-spatial model, and it seems that there is no relationship between the magnitude of the absolute value of the residuals and the region. No large spatial correlation is visually observed.

Table 4.6: Fitting performances of the models for 2018.

	Non–spatial model	GWR model	MGWR model
Kernel bandwidth(km)	—	1.41	0.58 – 14.41
Adjusted R^2	0.8492	0.9746	0.9821
AICc	365.18	−1632.58	−2024.02
MSE	0.0734	0.0067	0.0040
Root mean square error(%)	21.9019	5.9864	4.5662
Moran's I of residuals	0.5004 (p-value 0.0000)	0.0063 (p-value 0.1585)	−0.0336 (p-value 1.0000)

Table 4.6 summarizes the fitting performances of the non-spatial, GWR and MGWR models by the land price function in 2018. The MSE indicates the mean squared error and the root mean square error are defined by the following equation

Equation 4.3.2[14]:

$$\text{Root mean square error} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\exp(y_t(\boldsymbol{s}_i)) - \exp(\hat{y}_t(\boldsymbol{s}_i))}{\exp(y_t(\boldsymbol{s}_i))}\right]^2} \times 100(\%). \tag{4.3.2}$$

Regarding the goodness of fit, it is desirable that the AICc and root mean square error (%) are small and the adjusted R^2 close to 1. As for the spatial correlation of residuals, it is desirable that Moran's I be close to zero because the spatial correlation cannot be confirmed for the error term if the spatial regression model is fitted properly. The MGWR model outperformed the non-spatial and GWR models in 2018.

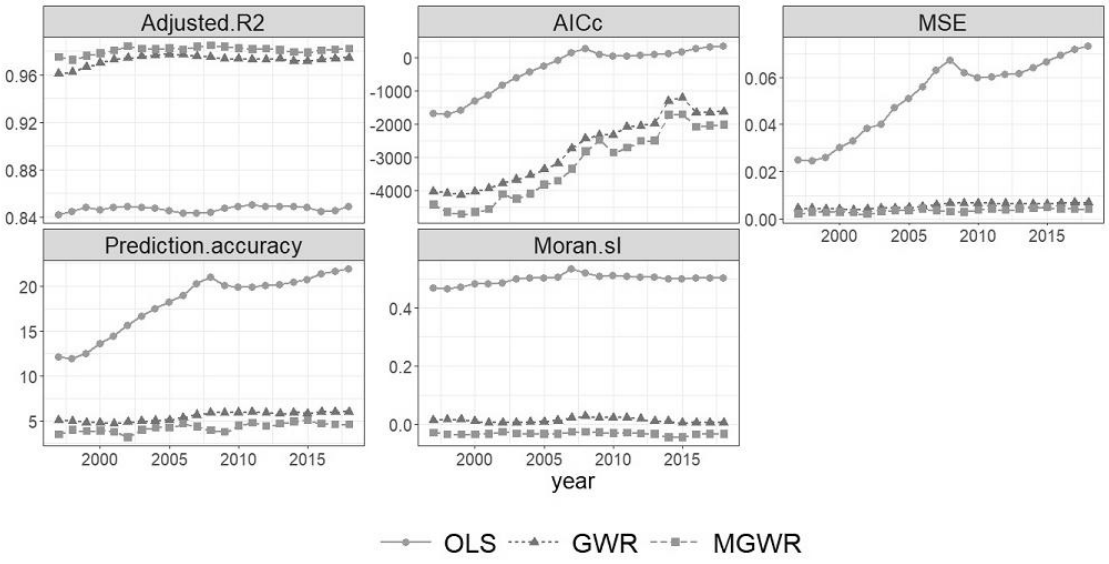


Figure 4.6: Transitions of the performance measures.

Figure 4.6 shows the time-series transition of the fitting performance of each model. Prediction accuracy indicates the root-mean-square error. The MGWR model had the best fit for all three models every year. Because the adjusted R^2

[14]It has been pointed out that the inverse transformation of the natural logarithm transformation calculates the bias prediction amount (Cressie, 1993). However, in this study, we use the inverse exponential transformation that was adopted by Chiles and Delfiner (2012) and Tolosana-Delgado and Pawlowsky-Glahn (2007).

of the non-spatial model changes to approximately 0.84, the non-spatial model can explain a large proportion of the official land price, but the residual Moran's I is approximately 0.50. If there is a spatial correlation, the adjusted R^2 is overestimated (Tsutsumi and Seya, 2012). The fit of the GWR and MGWR models was significantly better than that of the non-spatial model. The adjusted R^2 value of the GWR model was approximately 0.97, and the AICc ranged from $-4,000$ to $-1,500$. Because the transition of the residual Moran's I is approximately 0.03, no significant spatial correlation is observed. In the MGWR model, the adjusted R^2 is approximately 0.98, and the AICc ranges from $-5,000$ to $-2,000$, which is better than the GWR model, and the MSE and root mean square error are also improved. The change in residual Moran's I ranged from -0.04 to -0.03, and no significant spatial correlation was observed even at the 1% level. In addition, the MSE, root mean square error and residual Moran's I of the MGWR model were stable.

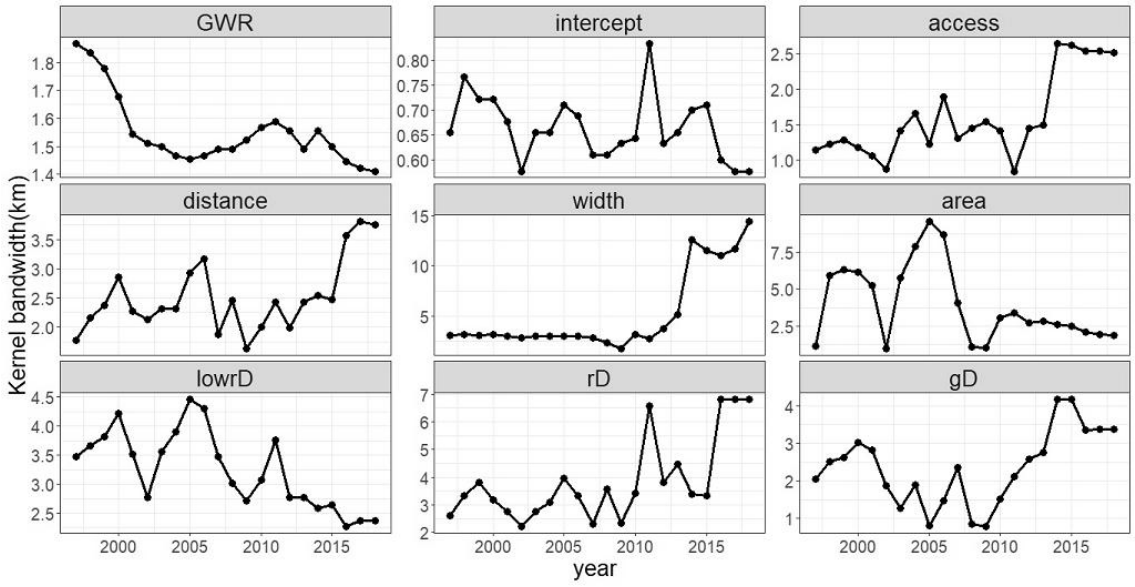
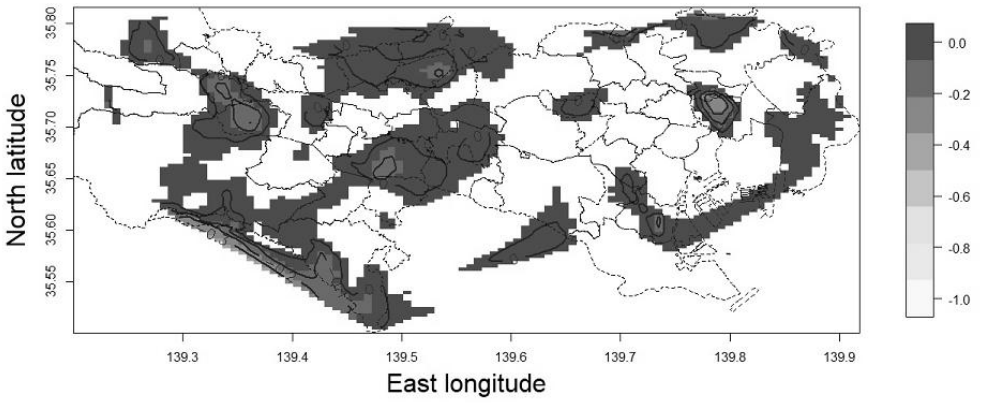


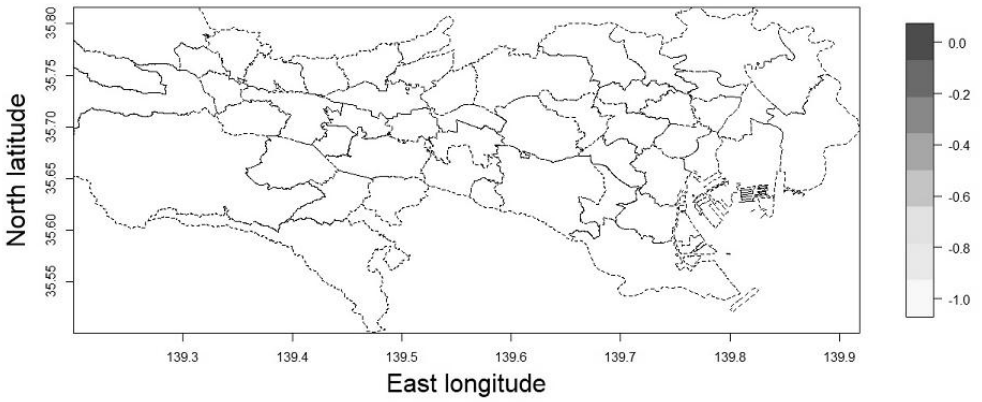
Figure 4.7: Transitions of the bandwidths for variable-specific kernels under the MGWR model and the common bandwidth for the GWR model.

Figure 4.7 shows the time-series transitions of the bandwidth for the variable-specific spatial kernels in the MGWR model. The "GWR" label in the top-left indicates the common bandwidth for spatial kernels under the GWR model.

Under this model, the kernel bandwidth changed from 1.4km to 1.9km. After the burst of the bubble economy, the range of the kernel bandwidth narrowed during the downward trend of official land prices and has remained stable since then. Under the MGWR model, the trend in kernel bandwidth for each explanatory variable was confirmed, regardless of the trend in official land prices. The kernel bandwidth of the constant term is smaller than those of the other explanatory variables. Moreover, a downward-sloping trend is observed over time because of the increase in land price dispersion, which cannot be explained by the explanatory variables included in this study. The kernel bandwidths of the access index, nearest station distance, front road width, residential area dummy and gas equipment dummy increased from 2013 to 2016. The cause is thought to be the large change in the number of publicly announced points. The fact that the kernel bandwidth of the front road width is larger than those of the other explanatory variables, similar predicted values continue in the study area, and an upward trend can be seen in recent years indicates it may become a global explanatory variable. We would like to consider these cases in future research.



(a) GWR model



(b) MGWR model

Figure 4.8: Regression surfaces of front road width in 2018 for the GWR and MGWR models.

Figure 4.8 shows the spatial regression surfaces for front road width in 2018 under the GWR and MGWR models. A cubic spline is applied for spatial interpolation. For the GWR model shown in Figure 4.8, negative values are

observed for the central and peripheral areas of the 23 wards and some parts of the Tama area (West part of Tokyo). As pointed out by Morioka and Fujita (1995) and Tokuda (2009), the negative impact on the environment, such as the increase in noise as road width increases, causes a decrease in value; however, the pattern of the changes in land prices is not constant even between the areas along major roads. Since the estimated bandwidth for front road width is 1.41 km under the GWR model and 14.41 km under the MGWR model, the local negative effect of front road width cannot be confirmed under the MGWR model and has weakened in recent years. In other words, the state of the spatial patterns of the regression coefficients changed significantly because of the use of variable-specific kernels for each explanatory variable.

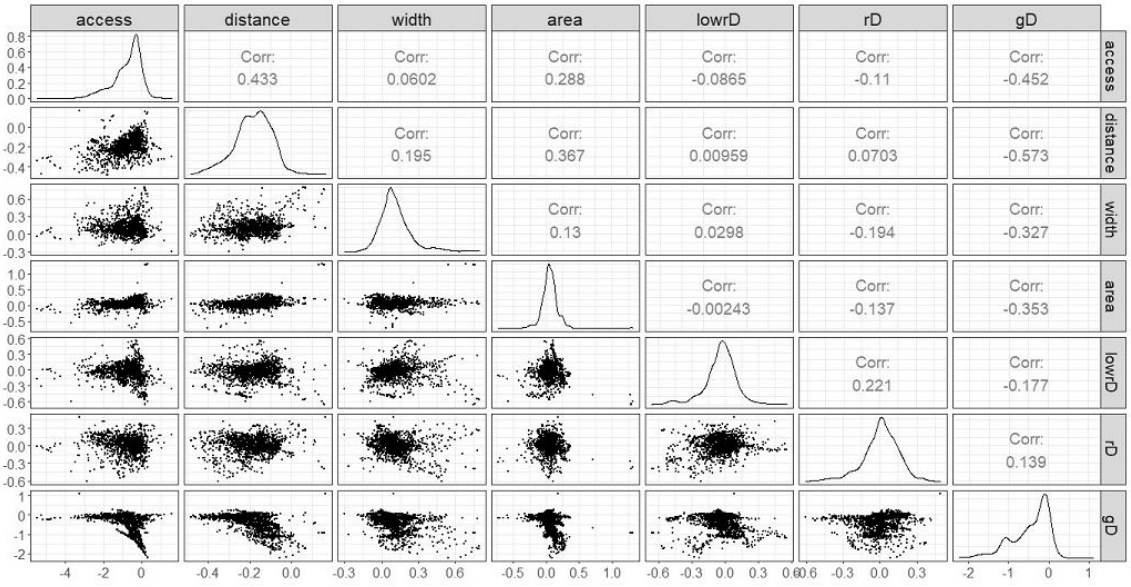


Figure 4.9: Scatterplots with kernel density estimates for each local regression coefficient in the GWR model for 2018. The elements above the diagonal show correlation coefficients of the corresponding pairs.

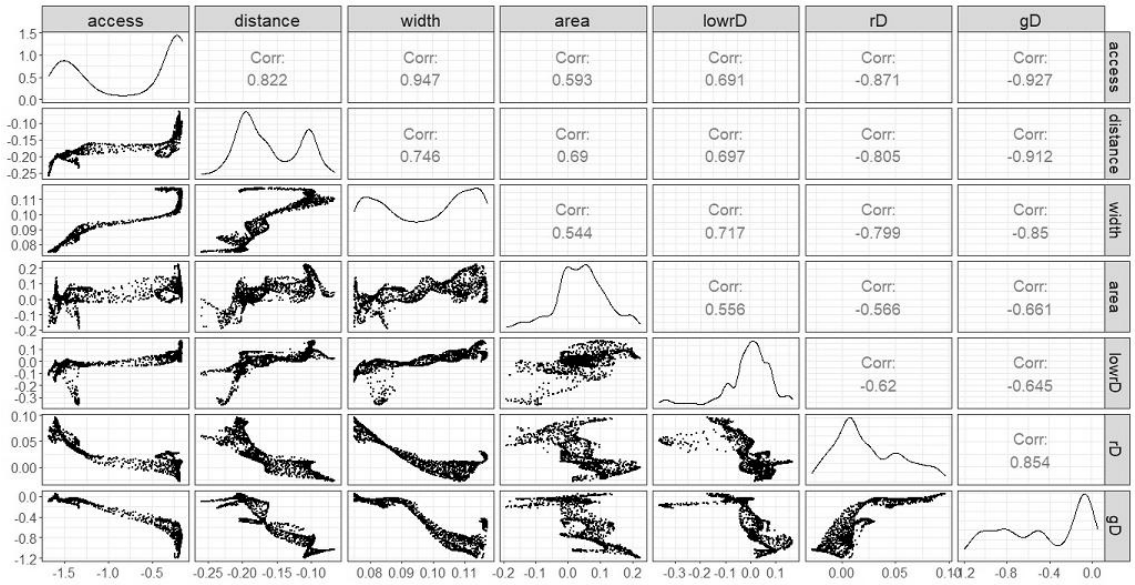


Figure 4.10: Scatterplots with kernel density estimates for each local regression coefficient in the MGWR model for 2018. The elements above the diagonal show correlation coefficients of the corresponding pairs.

Figures 4.9 and 4.10 show the scatterplot matrix of the estimated local regression coefficients for the GWR and MGWR models, respectively, in 2018. The kernel density estimations of the estimated local regression coefficients are plotted on the diagonal component of the figure, and the correlation coefficients of the two corresponding variables are shown above the diagonal. From the two figures, the estimated values are distributed with clusters, and for some variables, a local linear relationship in the distribution of the estimated values can be confirmed. Comparing Figures 4.9 and 4.10, under the MGWR model, the absolute value of the correlation coefficients among all variables is larger than under the GWR model, and the non-linear relationship of the estimated coefficients is clearer. In addition, the shape of the kernel density function of the estimated value is that of a unimodal distribution for the GWR model and a multimodal distribution for the MGWR model. The local regression coefficient of the MGWR model was estimated such that the range was smaller than that of the GWR model, and its coefficient was dispersed within the range.

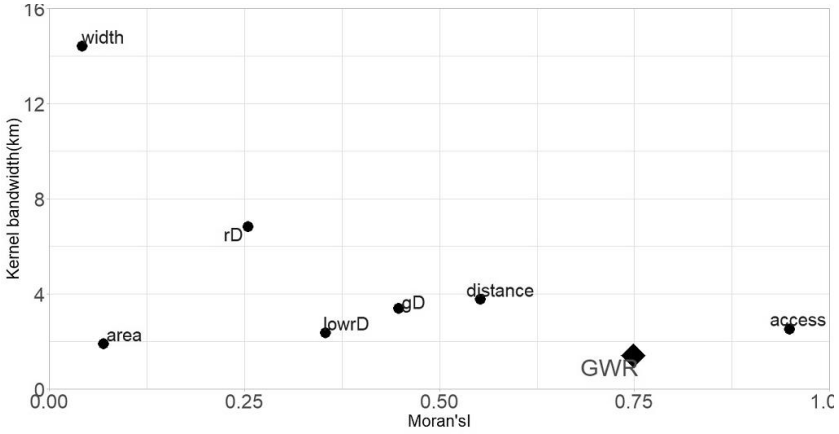
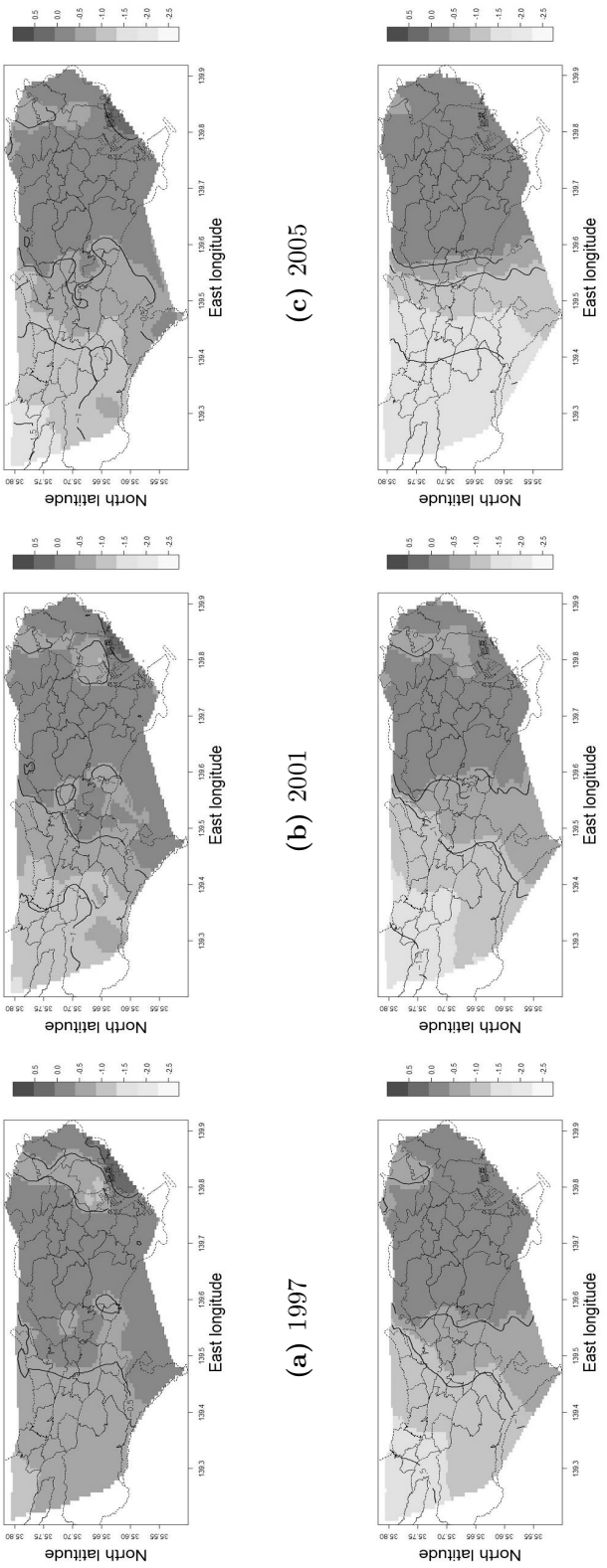


Figure 4.11: Scatterplot of the spatial correlation and bandwidth estimate of each explanatory variable for 2018.

Figure 4.11 shows the relationship between the spatial correlation of the explanatory variables in 2018 and the kernel bandwidth estimated using the MGWR model. In this figure, we also add the spatial correlation of official land prices and the kernel bandwidth estimated by the GWR model. We thus confirm that the stronger the spatial correlation between the explanatory variables, the smaller the kernel bandwidth, and the weaker the spatial correlation between the explanatory variables, the larger the kernel bandwidth. Because the kernel bandwidth is determined by the balance between the standard error and the bias of the explanatory variable estimates for a certain site, it can be inferred that if the spatial correlation is strong, highly accurate estimates can be made with values that are closer to each other. Additionally, the small kernel bandwidth of the GWR model is attributed to the smaller kernel bandwidth of the constant term in the MGWR model (0.58 km).

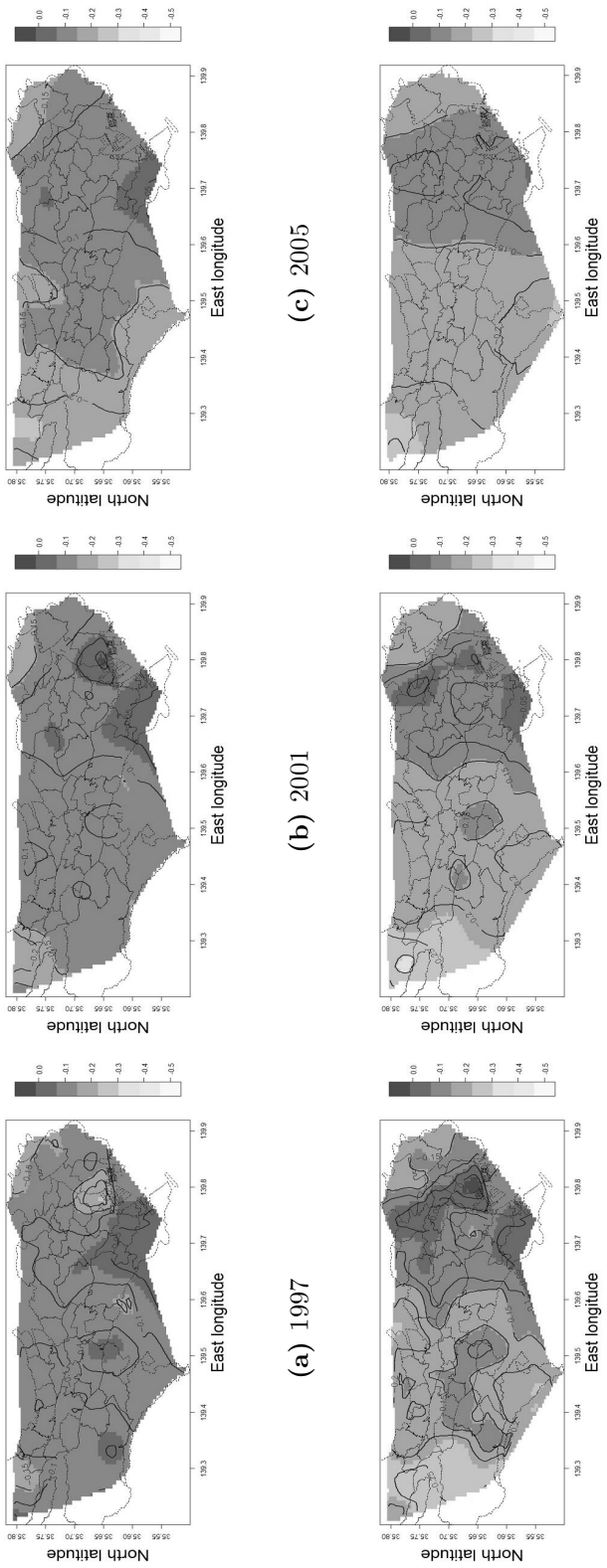


(a) 1997 (b) 2001 (c) 2005

(d) 2009 (e) 2013 (f) 2017

Figure 4.12: Transition of the spatial patterns of the estimated local regression surfaces for the access index from 1997 to 2017.

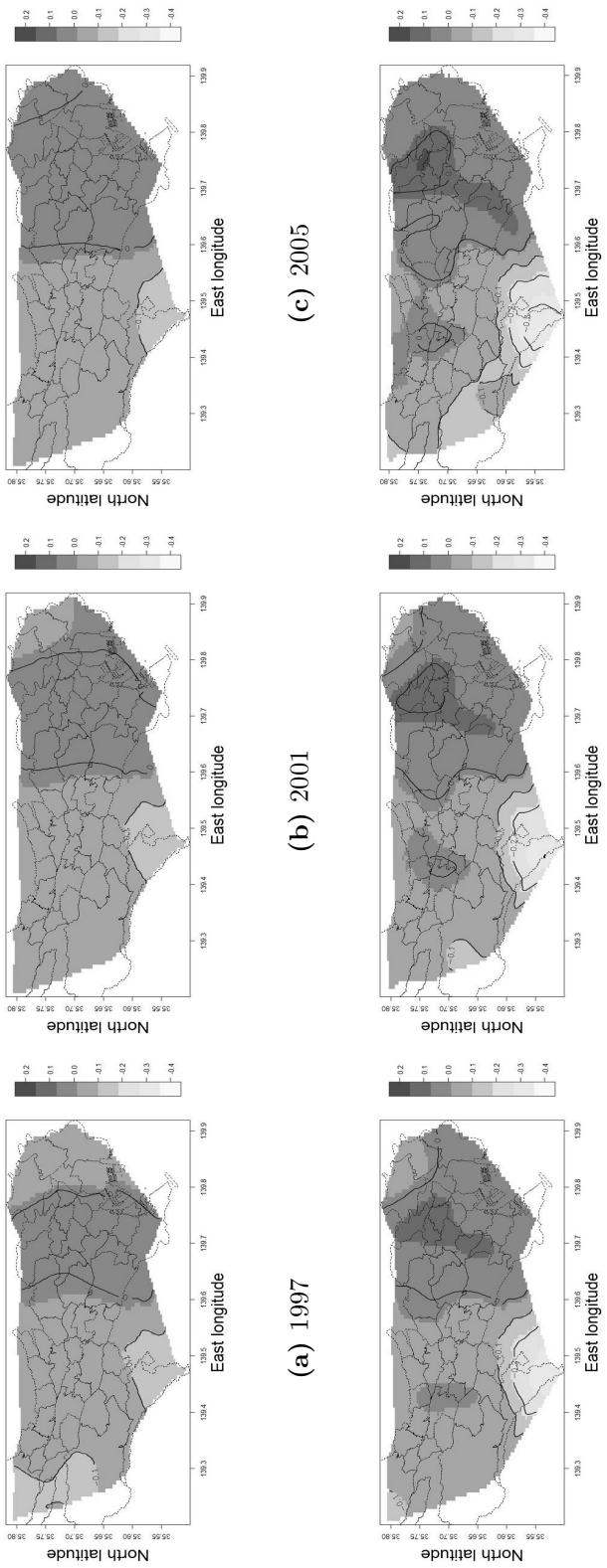
Based on the above results, the MGWR model was used to estimate the local regression surfaces for each explanatory variable. Specifically, we visualize the secular changes in the spatial patterns of the coefficients. Figure 4.12 shows the secular changes in the spatial patterns of the estimated local regression surfaces of the access index in the MGWR model. To make these changes easier to visualize, the spatial patterns for 1997, 2001, 2005, 2009, 2013 and 2017 are shown in 4-year intervals before and after 2009, when the effects of the 2008 financial crisis are evident in the official land prices. Overall, the coefficient is high in the east and low in the west. Initially, it was relatively high in the western part of the 23 wards and part of the North Tama area of Tokyo, but after the Lehman shock, it became high in 23 wards and low in the Tama area. Due to the nature of the access index, the value increases toward the west, meaning that the negative impact on land prices in the west becomes stronger. Furthermore, as time progressed, the difference increased. In other words, over time, the negative impact of the access index on land prices becomes stronger towards the west.



(a) 1997 (b) 2001 (c) 2005

(d) 2009 (e) 2013 (f) 2017

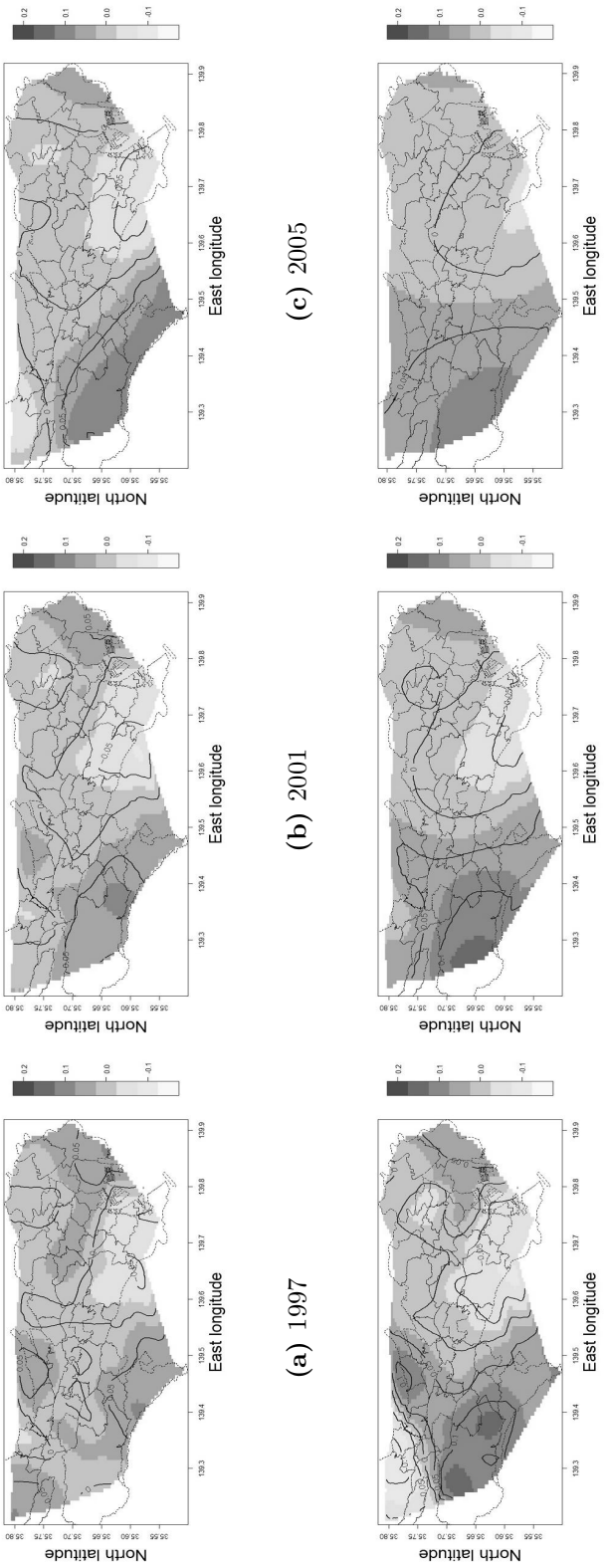
Figure 4.13: Transition of the spatial patterns of the estimated local regression surfaces of the distance from nearest station from 1997 to 2017.



(a) 1997 (b) 2001 (c) 2005

(d) 2009 (e) 2013 (f) 2017

Figure 4.14: Transition of the spatial patterns of the estimated local regression surfaces of the low-rise residential area dummy from 1997 to 2017.



(a) 1997 (b) 2001 (c) 2005

(d) 2009 (e) 2013 (f) 2017

Figure 4.15: Transition of the spatial patterns of the estimated local regression surfaces of the residential area dummy from 1997 to 2017.

Figure 4.13 shows the secular changes in the spatial patterns of the estimated local regression surfaces for the distance from the nearest station in the MGWR model. Figure 4.14 shows secular changes in the spatial patterns of the estimated local regression surfaces for the low-rise residential area dummy. Figure 4.15 shows secular changes in the spatial patterns of the estimated local regression surfaces for the residential area dummy.

Regarding the surfaces of the regression coefficient on the distance from the nearest station, there was no significant difference among areas in 1997, except for the relatively small values in the center of 23 wards. Over time, the coefficient in the eastern part of the 23 wards and the Tama area decreased, while that in the western part of the 23 wards increased. In areas where the coefficient was low, the negative impact of the distance from the nearest station on land prices became stronger, and the difference in land prices compared to areas with a high coefficient became larger.

Regarding the low-rise residential area dummy, the coefficient was initially high in the western part of the 23 wards; however, over time, areas with high coefficients appeared in the western part, mainly in Tachikawa City, Kunitachi City, and Kokubunji City, which are major cities in the Tama area. In addition, areas with low coefficients appeared, mainly in Machida City and Hachioji City. The transition after the 2008 financial crisis is also continuous, as the Kita, Toshima, Bunkyo, Shinjuku, Shibuya, and Meguro wards, among others, have high coefficients. Land prices in low-rise residential areas are higher in high-class residential areas, such as the “Ichigaya, Shinjuku Ward,” “Koishikawa, Bunkyo Ward,” “Shoto, Shibuya Ward.” Also Kita ward which has been regaining attention in recent years as being part of a “livable city” has high coefficients. In addition, there are areas with a positive coefficients in the western part of Tokyo centering on Tachikawa City. In contrast, there are negative values with low coefficients in the area centered on Machida City and areas where low-rise residential areas have lower land prices. It can be inferred that these facts in the Tama area are the result of redevelopment.

The residential area dummy is roughly divided into areas with high coefficients, such as Hachioji City and Machida, and areas with low coefficients, such as areas from the north side to the southwest side of the 23 wards, and

Chofu City, Komae City, regardless of the period. These effects were somewhat mitigated after the Lehman shock, but the coefficient is high in the western part of Tokyo and low in the eastern part. Areas with positive coefficients indicate land prices are higher in residential and quasi-residential areas, where convenience and commerciality are high, and building uses are relaxed, such as near stations and large-scale commercial facilities. It can be inferred that these facts in the Tama area are also the result of redevelopment.

Table 4.7: Time difference correlation coefficients between kernel bandwidth and average published land prices.

Time difference from average official land price	−3	−2	−1	0	+1	+2	+3
Time difference correlation coefficient	−0.2133	0.0203	0.2670	0.4559	0.6022	0.7095	0.6930

Finally, Table 4.7 shows the time difference correlation coefficients between the common bandwidth (km) of the spatial kernels under the GWR model and the average published land prices (logarithmic transformation). The time difference correlation coefficient was calculated by shifting the kernel bandwidth for each year based on the average land price. The kernel bandwidth and average land prices behave similarly, but when the time difference correlation is calculated, the kernel bandwidth lags the average land price by approximately two years. This fact is expected to help confirm trends in the real estate market.

5

Transitions of Spacial Variogram Estimations: Revisited

5.1 Consideration of Explanatory Variables

The analysis of temporal changes in the previous two chapters does not include economic factors as explanatory variables. Many land price regression analyses use the individuality of land prices as an explanatory variable, and apart from spatiotemporal analysis, few regression analyses use typical economic indicators in spatial analysis. According to Kitasaka (2012), land prices in Tokyo lag behind economic trends by about a year. Therefore, considering the causal relationship, the land price model should include economic factors. Hence, the municipality's average annual income (Income) was added to the explanatory variables. In addition, the official prices up to 2020 cannot take into account the effects of Covid-19. It is now possible to estimate the impact of the Corona disaster on land prices using the published prices in 2021. In Section 4.1, we remove the explanatory variables of the floor area ratio, add the average annual income per municipality, and use the published prices from 1997 to 2021 to perform a reanalysis. Table 5.1 presents the VIFs of the explanatory variables used in this study.

Table 5.1: VIF for the explanatory variables in 2021 used in this chapter.

	Access	Distance	Width	Acre	Income	southD	dwD	gD	sD
VIF	2.761	1.355	1.511	1.353	2.221	1.010	1.063	1.185	1.015
	lowrD	rD	qfD	fD					
VIF	2.069	1.575	2.102	1.660					

5.2 The Data

Table 5.2: List of variables and their overviews.

	Overview
Dependent variable	
Official land price (yen/m^2)	Official land price data for residential areas between 1997 and 2021
Explanatory variable	
1, Access index (min)	Average commute time from the closest station to six major stations of Tokyo by the average number of users
2, Distance from station (m)	Distance from the closest station
3, Front road width (m)	In case of two or more roads, the widest ones were adopted
4, Land area (m^2)	
5, Average annual income (yen)	Average annual income by municipality, divide by 1000
6, South-headed	Dummy variable. Front road is located southeast, south or southwest 1, other directions 0
7, Driveway	Dummy variable. Front road is a driveway 1, if not 0
8, Gus equipment	Dummy variable. Present 0, Absent 1
9, Sewerage	Dummy variable. Present 0, Absent 1
10, Low-rise residetial	Dummy variable. Low-rise exclusive residential districts of categories 1 or 2 are set at 1, others 0
11, Residential/quasi-residential	Dummy variable. Residential/quasi-residential districts of categories 1 or 2 are set at 1, others 0
12, Quasi-fire prevention	Dummy variable. Quasi-fire prevention districts 1, others 0
13, Fire prevention	Dummy variable. Fire prevention districts 1, others 0

Table 5.3: Summary statistics for access index, distance from the closest station, front road width and land area in 2021.

	Access index	Distance from station	Front road width	Land area
Sample size	1,542	1,542	1,542	1,542
Average	42.556	1063.184	6.038	211.427
Standard error	0.4575	24.081	0.09200	8.533
Median	39.833	800	5.4	154
Mode	61.067	1200	6	165
Standard deviation	17.964	945.612	3.613	335.089
Kurtosis	-0.5088	14.596	13.051	167.101
Skewness	0.4179	3.208	22.697	10.707
Max	106.966	8400	40	7320
Min	8.526	100	2	47

Table 5.2 presents the explanatory and dependent variables used in this study. Table 5.3 shows the summary statistics of the access index (min), distance from the closest station (m), front road width (m), and land area (m^2), as of 2021. The access index represents the convenience of accessing to central Tokyo. For this variable, we used the average travel time from the station closest to the six major stations: Shinjuku, Ikebukuro, Tokyo, Shibuya, Ueno, and Shinagawa. Data on distance from the station, front road width, and land area were obtained from official land price announcements published by the MLIT[15]. As shown in Table 2, the average time from Tokyo to each terminal station was 43min, the average distance from the closest station was 1,060m, the average front road width was 6.0m and the average land area was 211m^2. In addition, the skewed distributions of the distance from the closest station, front road width and land area were observed with several large outliers.

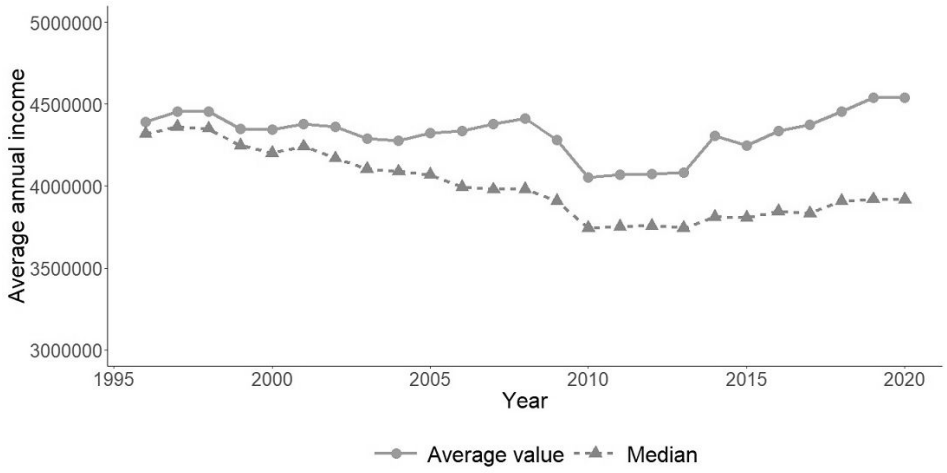


Figure 5.1: Tokyo's average annual income from 1996 to 2020.

[15]For more information, see Section 3.2.

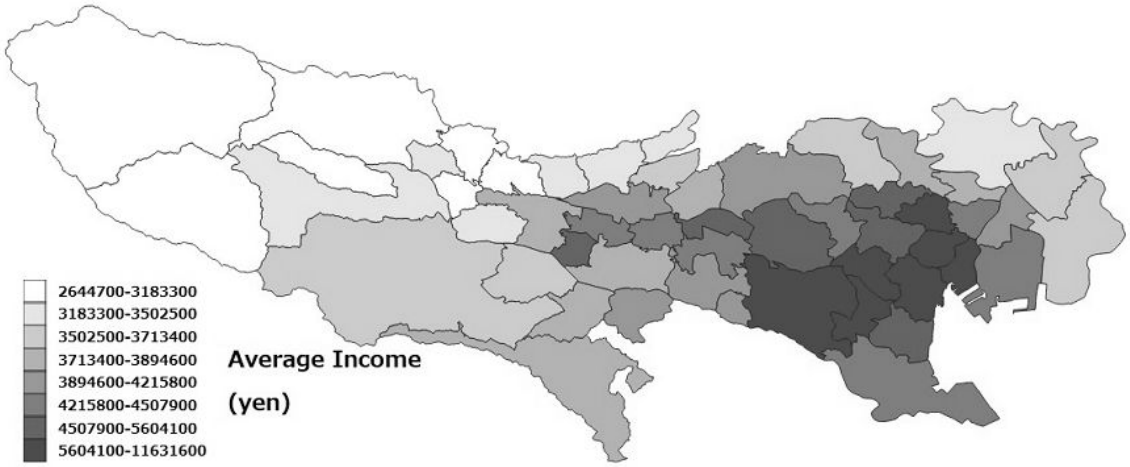


Figure 5.2: Spatial distribution of average annual income by Tokyo municipality in 2020.

Figure 5.1 shows the changes in Tokyo's average annual income (tax included) from 1996 to 2020 per municipality. Using the "Basic Survey on Wage Structure" released by the Ministry of Health, Labor and Welfare, the annual income for each municipality in Tokyo was calculated[16]. Assuming that the annual income of one year affects the land price of the next year, the annual income per municipality one year before the analysis period was used as an explanatory variable. During the "lost 20 years" after the bubble burst, the average annual income showed a negative trend. During the 2008 financial crisis, the average annual income fell further; however, it has gradually gained a positive trend since then. In recent years, the divergence between the mean and the median has increased. It can be inferred that very high annual incomes affect the average value. The Gini coefficient for the municipality's average annual income was 0.072 in 1996 and 0.141 in 2020. Figure 5.2 shows the spatial distribution of average annual income of the Tokyo municipality. Land price and average annual income are highly correlated when comparing the two distributions. These findings are consistent with various studies, including Tiwari (2000), Sutton (2002) and Chen et al. (2007). The high average annual income in the eastern part of Tama, which is

[16]The average annual income in Tokyo was calculated independently by taking the weighted average of the published annual income of each municipality by population.

adjacent to the central part of the 23 wards, is also reflected in the corresponding high land prices.

5.3 Results of the Estimated Parameters and their Temporal Variation

In this section, we investigate the land price model using official land prices from 1997 to 2021 in Tokyo. As we have indicated thus far, the assumption of second-order spatial stationarity does not apply to all regions. As in Chapter 3, we used Ward's linkage-based hierarchical clustering technique to apply cluster analysis, which divides the data into four regions. For this, we calculate the distances between municipalities using the average and variance of the land prices of each municipality[17]. In this study, only the mean and variance of land prices were used for clustering. There is no relationship between clustering and the economic trends of the municipalities. As a result, the four regions of Tokyo were divided into the same regions as those in Chapter 3.

To detect a global trend in the four regions, we chose five explanatory variables – access index, distance from the closest station, front road width, land area, and average annual income – at five different times, that is, six and twelve years before and after the 2008 financial crisis (1997, 2003, 2009, 2015, and 2021). We performed the Kruskal-Wallis H test for all years. Consequently, for the access index and front road width, the null hypothesis that the median of all five time points is the same cannot be rejected in any of the four regions at a significance level of 5%. A significantly negative trend was observed in the northeast for the distance to the closest station. Presumably, this is due to the extension of railroad lines and the construction of new ones, such as the Fukutoshin Line (No. 13, opened in 2008), as well as changes in official price points targeting standard land. In the land area, a negative trend was observed in the northeast and southwest. This may be due to changes in land regulations occasioned by the development of the "Tama Site Development Basic Plan" and the replacement of the standard point that followed. In terms of the average annual income, the

[17]For more information, see Section 3.3.

null hypothesis was rejected for all four regions. However, the trends in the four regions were broadly divided into two categories. In the southeast, positive trends were confirmed, except for a few years after the 2008 financial crisis. In other regions, except in recent years, negative trends have also been found.

Table 5.4: Estimated regression coefficient of non-spatial models in 2021.

Explanatory variable	Whole Tokyo	Southeast	Northeast	Southwest	Northwest
Constant	8.4384 (0.3416)	7.9105 (0.6647)	6.4151 (0.5433)	10.3150 (2.0010)	7.3465 (2.2242)
Access index	-0.7330 (0.0206)	−0.5143 (0.0493)	−0.3936 (0.0260)	−1.3602 (0.0786)	−0.9366 (0.0781)
Distance from station	-0.2401 (0.0095)	−0.2365 (0.0240)	−0.1959 (0.0121)	−0.3069 (0.0144)	−0.2013 (0.0182)
Front road width	0.1052 (0.0190)	0.2100 (0.0392)	0.0810 (0.0204)	0.0317 (0.0323)	0.0707 (0.0416)
Land area	0.0600 (0.0125)	0.1666 (0.0193)	0.0772 (0.0122)	−0.0563 (0.0330)	−0.0566 (0.0340)
Average annual income	0.9496 (0.0340)	0.8603 (0.0587)	1.0400 (0.0564)	1.1716 (0.2112)	1.2309 (0.2406)
South-headed	−0.0022 (0.0121)	0.0543 (0.0265)	0.0286 (0.0131)	−0.0194 (0.0201)	0.0284 (0.0237)
Driveway	−0.0200 (0.0210)	−0.0930 (0.0797)	0.0064 (0.0272)	−0.0442 (0.0343)	−0.0475 (0.0293)
Gus equipment	−0.2848 (0.0234)			−0.1608 (0.0355)	−0.1228 (0.0333)
Sewerage	−0.3928 (0.1135)			−0.3475 (0.1085)	
Low-rise residential	0.0188 (0.0167)	0.0347 (0.0326)	−0.0481 (0.0158)	−0.0236 (0.0440)	−0.0294 (0.0380)
Residential/quasi-residential	0.0159 (0.0184)	−0.0266 (0.0379)	0.0172 (0.0161)	0.0163 (0.0416)	−0.0734 (0.0548)
Quasi-fire prevention	0.1761 (0.0174)	0.0131 (0.0731)	−0.0013 (0.0317)	0.1756 (0.0385)	0.0891 (0.0283)
Fire prevention	0.2113 (0.0439)	−0.0101 (0.0909)	−0.0310 (0.0483)		
Adjusted R^2	0.9091	0.8625	0.8389	0.8138	0.8602

Table 5.4 shows the OLS estimates for 2021 land prices. Recall that the OLS method constitutes a global model with non-spatial structures. The numbers in parentheses are standard errors. Some regions do not have gas equipment dummy or sewerage dummy parameters because these facilities had already been built at the entire official target point before 2021, while others do not have the fire prevention dummy because no fire prevention district was selected as a target point for the official land prices of that year. The characteristics of each region become clear when the respective estimated coefficients are compared with those of Tokyo as a whole. Larger regional differences in the parameters of access index, land area, low-rise residential dummy, residential/quasi-residential dummy and quasi-fire prevention dummy exist. The access index coefficient shows that official land prices in the west have a greater negative impact than those in the east. Regarding the land area coefficient, the fact that the southeast and northeast have positive coefficients indicates no volume discount effect. Although the regression coefficient has not reached the 5% significance level, the Low-Rise Residential dummy coefficient is positive in the southeast, implying the existence of regional brands. Conversely, the coefficient is negative at the 5% significance level in the northeast. Similarly, although the 5% significance level is not met, the residential/quasi-residensial dummy coefficient is positive in the northeast and southwest. This implies that commercial elements do not negatively affect housing preferences in the area. The quasi-fire prevention dummy coefficient has a significant positive impact at the 5% level in the southwest and the northwest. Generally, areas with fire protection regulations have stricter building restrictions; therefore, construction costs are higher than in unregulated areas. However, it can be inferred that in the southwest and northwest, the positive impact of the increase in security and safety by regulations caused official land prices to increase[18]. These facts show that the factors determining official land prices are not significantly different from our intuitive understanding.

[18]There are different approaches on the regression coefficient of the land price model, especially on the sign of those with positive and negative. For details, see Section 4.3.

Table 5.5: Estimated coefficients of the spatial process model by IRWGLS in 2021.

Explanatory variable	Whole Tokyo	Southeast	Northeast	Southwest	Northwest
Constant	10.0284 (0.5018)	9.1962 (0.9687)	10.3823 (0.8533)	14.2126 (5.1538)	11.8526 (3.4423)
Access index	-0.7220 (0.0304)	−0.2454 (0.1328)	−0.3599 (0.0673)	−1.1829 (0.0802)	−1.1753 (0.0811)
Distance from station	−0.1977 (0.0140)	−0.1621 (0.0698)	−0.1735 (0.0371)	−0.1983 (0.0201)	−0.2184 (0.0231)
Front road width	0.0902 (0.0279)	0.2062 (0.1114)	0.0926 (0.0631)	0.0424 (0.0443)	0.0729 (0.0533)
Land area	0.0756 (0.0184)	0.1522 (0.1048)	0.0534 (0.0451)	−0.0048 (0.0899)	−0.0457 (0.0927)
Average annual income	0.7086 (0.0660)	0.5470 (0.0555)	0.5262 (0.0369)	0.4790 (0.0534)	0.7925 (0.0588)
South-headed	0.0104 (0.0178)	0.0470 (0.0749)	0.0185 (0.0390)	0.0016 (0.0258)	0.0210 (0.0310)
Driveway	−0.0187 (0.0309)	0.0095 (0.2183)	−0.0030 (0.0798)	−0.0471 (0.0460)	−0.0258 (0.0368)
Gas equipment	−0.1479 (0.0344)			−0.0680 (0.0497)	−0.0780 (0.0424)
Sewerage	−0.1747 (0.1670)			−0.1708 (0.1472)	
Low-rise residential	0.0188 (0.0245)	0.0758 (0.1417)	0.0204 (0.0600)	−0.0262 (0.9970)	0.0083 (0.9861)
Residential/quasi residential	0.01085 (0.0270)	−0.0012 (0.1163)	0.0132 (0.0553)	0.0704 (0.0566)	0.0307 (0.0701)
Quasi-fire prevention	0.1190 (0.0256)	−0.0710 (0.2092)	0.0404 (0.0889)	0.0850 (0.0555)	0.0569 (0.0411)
Fire prevention	0.1671 (0.0545)	−0.1501 (0.2797)	0.0598 (0.1509)		

Table 5.5 summarizes the 2021 estimated land price parameters using the IRWGLS method. Compared to the results of the non-spatial model, the absolute values of the spatial process model estimates were smaller, and the standard error was larger. Compared with Table 5.4, the estimated coefficient of the low-rise residence dummy in the northeast is no longer significantly negative at the 5% level. In the northwest, the low-rise residential and residential/quasi-residential dummies become positive, and the difference in the official land prices between the two residential districts decreases. In addition, the estimated coefficient for the land area is smaller than in other areas. Most of the signs of the other coefficients match those of the OLS. Following previous studies, we adopted a spatial variogram with a spherical model and estimated the covariance structure with a valid distance of 6.67km in the southeast, 5.56km in the northeast, and 8.89km in the southwest and northwest[19].

[19]See Section 3.3 for the valid distance.

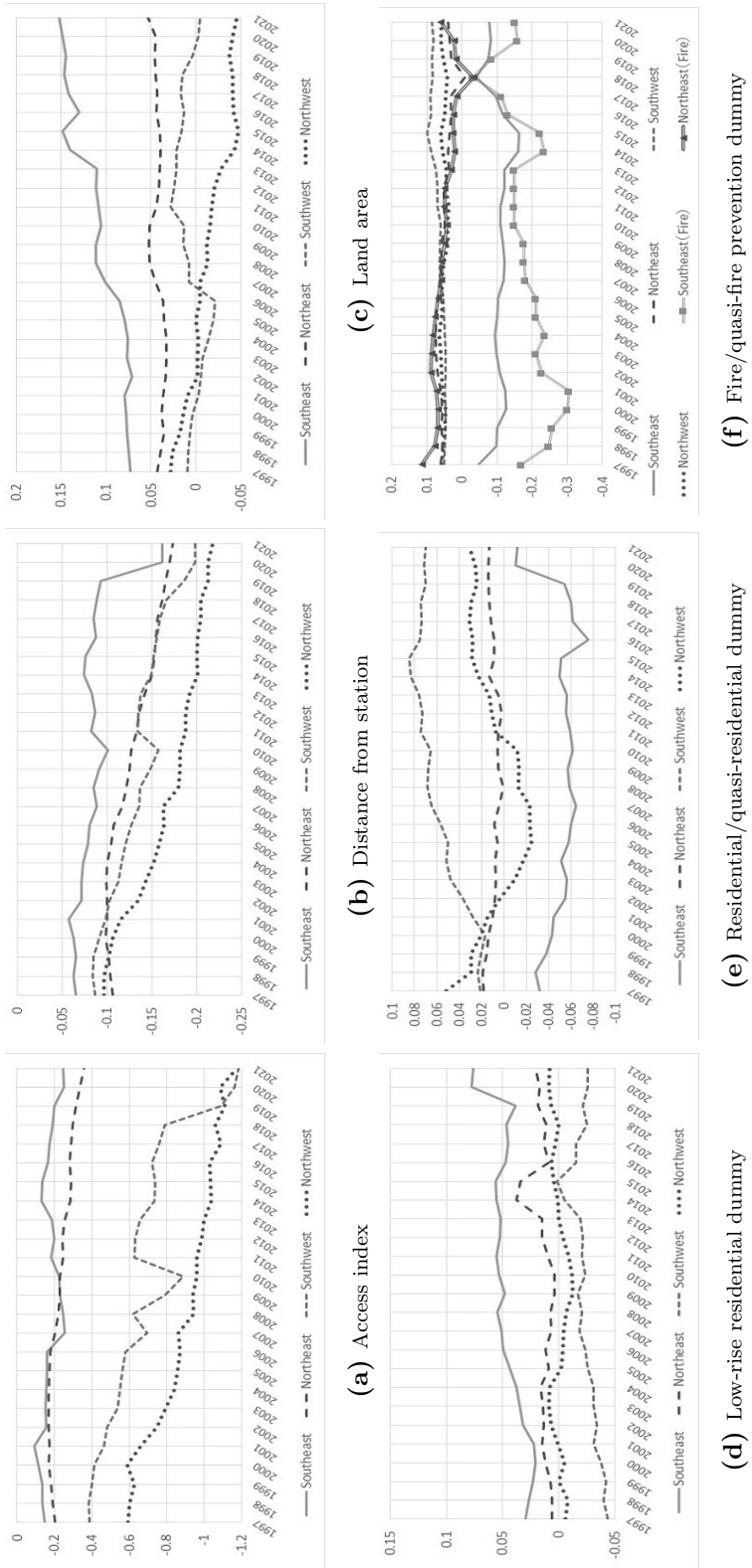


Figure 5.3: Transitions of the regression coefficients.

To observe the time-series transitions of the estimated coefficients, we provide the following figures and tables: Figure 5.3 shows the annual changes in the regression coefficients of the access index, distance from the closest station, land area, low-rise residence dummy, Residential/quasi-residential dummy, and quasi-fire prevention/fire prevention dummy of the estimated spatial process model. Table 5.6 summarizes the test results of the null hypothesis of the zero slopes in the linear regression for these variables with the explanatory variable time t. In this table, the "90% lower bound" and the "90% upper bound" represent the lower and upper bounds of the 90% confidence interval.

Table 5.6: Test statistics and its p-values for the null hypothesis of no slope in the linear regression with time trend for access index, distance from station, low-rise residential dummy, residential/quasi-residential dummy and quasi fire/fire prevention dummy

Explanatory Variable	Region	Estimates	p-value	90% L. B.	90% U. B.
Access index	Southeast	−0.0038	0.0003	−0.0053	−0.0025
	Northeast	−0.0071	0.0000	−0.0086	−0.0062
	Southwest	−0.0265	0.0000	−0.0266	−0.0245
	Northwest	−0.0305	0.0000	−0.0325	−0.0281
Distance from station	Southeast	−0.0029	0.0000	−0.0036	−0.0019
	Northeast	−0.0033	0.0000	−0.0038	−0.0030
	Southwest	−0.0039	0.0000	−0.0041	−0.0039
	Northwest	−0.0058	0.0000	−0.0062	−0.0053
Land area	Southeast	0.0035	0.0000	0.0029	0.0040
	Northeast	0.0002	0.4378	−0.0002	0.0005
	Southwest	0.0005	0.2209	−0.0002	0.0011
	Northwest	−0.0029	0.0000	−0.0034	−0.0030
Low-rise residential dummy	Southeast	0.0021	0.0010	0.0016	0.0027
	Northeast	0.0011	0.0108	0.0003	0.0018
	Southwest	0.0027	0.0021	0.0020	0.0032
	Northwest	0.0000	0.9764	−0.0006	0.0006
Residential/ quasi-residential dummy	Southeast	0.0008	0.0588	−0.0001	0.0016
	Northeast	−0.0002	0.0263	−0.0004	-0.0001
	Southwest	0.0024	0.0000	0.0017	0.0027
	Northwest	0.0005	0.2456	−0.0002	0.0014
Quasi-fire/ prevention dummy	Southeast	0.0003	0.5540	−0.0009	0.0015
	Northeast	−0.0028	0.0000	−0.0035	−0.0025
	Southwest	0.0013	0.0000	0.0011	0.0014
	Northwest	0.0004	0.1837	−0.0001	0.0008
Fire prevention dummy	Southwest	0.0080	0.0000	0.0057	0.0105
	Northeast	−0.0046	0.0000	−0.0059	−0.0040

The transition of the access index coefficients can be divided into eastern and western Tokyo. The coefficient mostly remained stable in the southeast and northeast, with a small negative value throughout the analysis period; however, in the southwest and northwest, it followed a downward trend. These results indicate that in the western part of Tokyo, the longer time required to access

the city center has a stronger negative impact on official land prices every year, and the gap with the eastern part of Tokyo is widening. It can be seen that the coefficient of distance from the closest station is negative in all regions throughout the analysis period, and the three regions other than the central part (southeast) have negatively smaller values than the southeast. In these areas, the farther an area is from the the closest station, the lower the land prices tend to be, and the impact of this coefficient increases yearly.

The estimated land area coefficient was positive in eastern Tokyo. This indicates that the larger the land area is, the more expensive it is. This trend is stronger in the southeast than in other regions and increases even further over time, representing the increasing brand power of certain regions. By contrast, a volume discount effect can be seen in the northwest in the latter half of the analysis period, increasing gradually. The low-rise residential dummy coefficient was positive in the southeast. There was a strong tendency to emphasize the living environment throughout the study period.

On the other hand, as for the residential/quasi-residential dummy coefficient, the southeast has a negative value, as regions with strong commercial elements have lower official land prices than other districts. In contrast, a positive trend can be seen in the southwest, where there is a growing tendency to emphasize convenience and commercial elements. This result can likely be attributed to the recent land redevelopment and improvement plans in the Tama area.

Regarding the coefficient of the quasi-fire/fire prevention dummy, the southeast has a negative coefficient and lower prices than other restricted regions. This suggests that this area avoids increases in construction costs due to restrictions, but their impact has recently been decreasing. In addition, a positive trend can be seen in the coefficient of the quasi-fire prevention dummy of the southwest, where, contrary to the southeast, higher levels of safety and security due to restrictions tend to increase official land prices. It can be seen that, in the first half of the analysis, the fire prevention districts in the southeast lowered the official land prices sharply, but this impact was lower in the latter half. Moreover, it is possible to see that the coefficients for the quasi-fire prevention dummy other than in the southwest also affect the official land prices less and less over the years.

In panel (d) of Figure 5.3, there is a discontinuity in the regression coefficient of the low-rise residential dummy (in the northeast and southwest) in 2014 and 2016, and the same in the regression coefficients of the land area and fire prevention dummy of the southeast in panels (c) and (f). Presumably, this was the effect of a decrease in the number of official target points in 2014 and an increase in 2016. Following the changes in the official land price target points, there was a significant change in the ratios of the number of points in low-rise exclusive residential districts and in residential/quasi-residential districts[20]. We can see sudden jumps in the regression coefficients of the distance from the closest station, the low-rise residential dummy, the residential/quasi-residential dummy and fire prevention dummy in the southwest in 2020. This is because of the addition of an extremely high price site of 12,800,000 yen per $1m^2$ for land price announcements in 2020 and 2021.

Table 5.7: Estimated variogram parameters with sum of squared errors of land price models in 2021.

	Southeast	Northeast	Southwest	Northwest
Estimates for spatial variogram				
Nugget	0.0206	0.0063	0.0077	0.0047
Partial sill	198.2442	0.0360	0.0377	0.0165
Range (km)	51,190	16.47	5.39	4.88
Mean of squared residuals				
Non-spatial model: MSE	0.0376	0.0197	0.0438	0.0268
Spatial model: MSE	0.0272	0.0084	0.0149	0.0095

Table 5.7 shows the spatial variogram estimates for each of the four regions in 2021, together with the mean squared error (MSE) of the non-spatial and spatial process models. The southeast's range and partial sill values are large because the spatial variogram approached a linear relationship within the valid distance[21], and the linear part in the spherical model with a small distance difference was fitted to the empirical variogram. Considering that the larger the ratio of the sill and nugget effect or the distance of the range, the stronger the spatial correlation

[20]For more information, see Section 3.3.
[21]For more information, see Section 3.2.

of the random fields, it is clear that the residuals of the spatial process model of eastern Tokyo have a stronger spatial correlation than the west. The nugget of the southwest is larger than that of the northeast, which has more than half of the 23 wards where the official land prices are relatively high. Moreover, the MSE of the land price models showed that the MSE of the spatial process model was reduced approximately from 1/3 to 2/3 of the MSE of the non-spatial model.

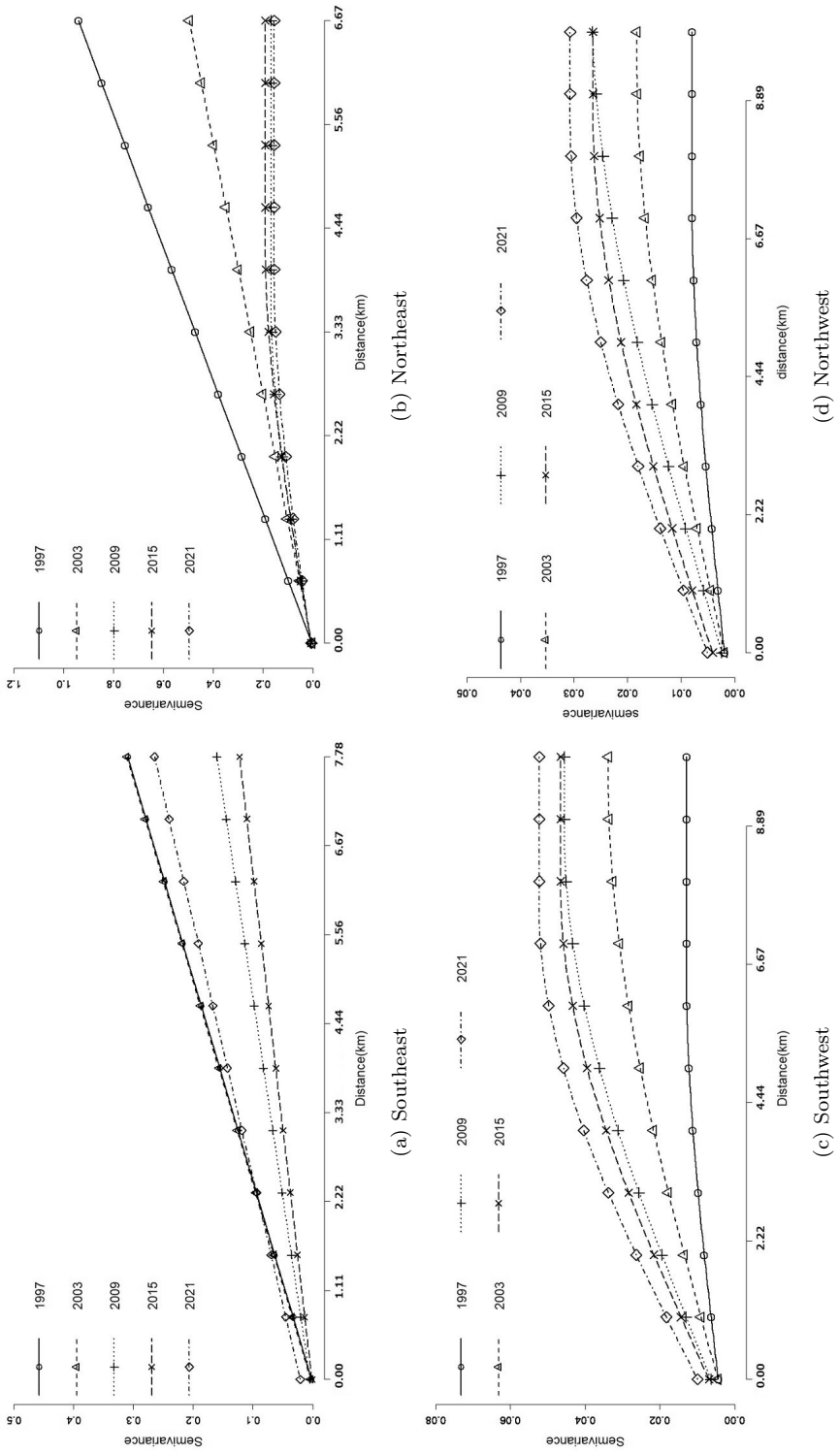


Figure 5.4: Changes in the variograms for four regions.

Figure 5.4 shows the variation in the estimated spatial variogram for each region over five different years. As before, to provide better visualization of the annual changes, we chose the year 2009 (owing to the impact of the financial crisis of 2008) and 6 and 12 years before and after it (i.e., 1997, 2003, 2009, 2015, and 2021). The linear parts show that when the rate of change increases, the degree of attenuation of the covariance function also increases. Furthermore, an increase in the sill of the variogram represents an increase in the variance of the spatial process of the residuals. In the southeast, the decrease in sills after the financial crisis was remarkable, and it can be seen that it has recovered to the level of 1997 in recent years. In the northeast, before the 2008 financial crisis, the value of sills decreased, and the shape of the variogram has remained unchanged since the financial crisis. In contrast, the value of sills continues to increase in western Tokyo.

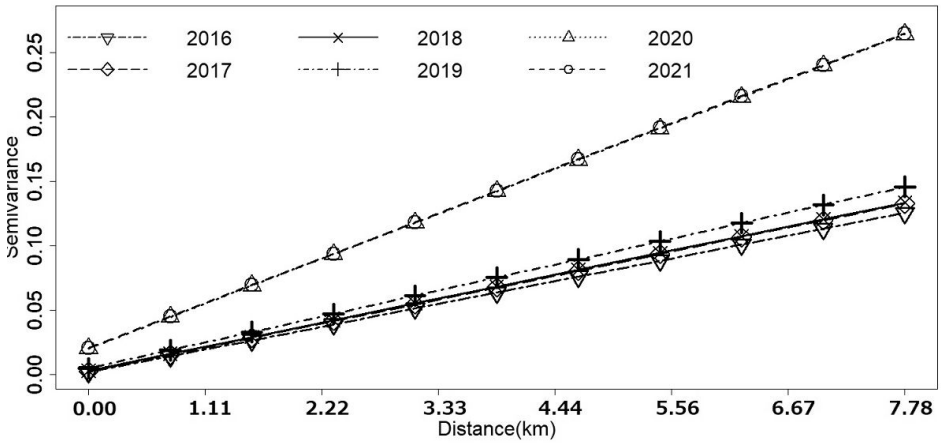


Figure 5.5: Change in the variograms in the southeast from 2016 to 2021.

Figure 5.5 shows the change of the variogram in the southeast over the last six years. From 2016 to 2019, variogram sills show a moderate increasing trend. Some very high land price points were added in 2020, and a jump of shape of the variogram was observed in 2020. In 2020 and 2021, the shape of the variogram was almost the same. It can be inferred that the impact of the coronavirus outbreak on land prices has stopped the increase in variogram sills. However, the impact is so far limited to Central Tokyo because there is no change in the

residual structure of the southeast, where the impact of the change in the official land price point is greater, and the trend of variogram change in other regions has not changed.

It should be noted that the spatial structure and transition of the residuals are slightly different in Chapter 3, especially in the eastern part of Tokyo, due to the change in the explanatory variables. The major change was the incorporation of the economic trends of each municipality into the model. In Chapter 3, in the southeast, the sill has stagnated in the 2008 financial crisis but has increased again in recent years. In Chapter 5, it is declining without stagnation. In the northeast, the sill has been on a recovery trend since the 2008 financial crisis. In Chapter 5, it is stagnant, without increasing.

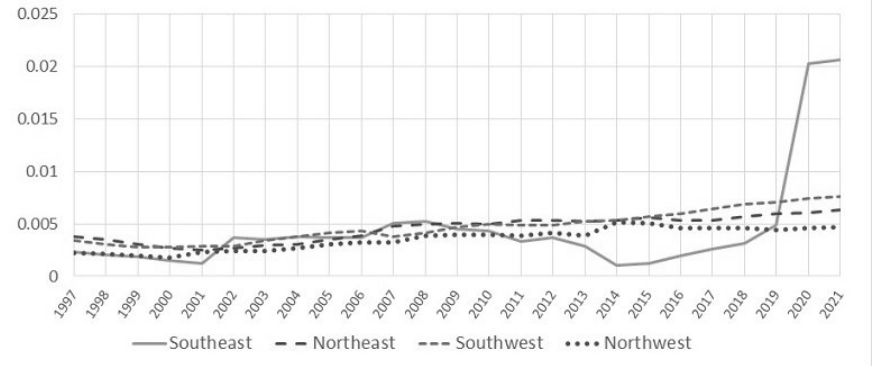


Figure 5.6: Nugget effect.

Table 5.8: Test statistics and its p-values for the null hypothesis of no slope in the linear regression with time trend for nugget.

Tokyo region	estimates	p-value	90% L. B.	90%U. B.
Whole Tokyo area nugget				
Southeast	0.0003	0.0431	0.0001	0.0006
Northeast	0.0001	0.0000	0.0001	0.0001
Southwest	0.0002	0.0000	0.0002	0.0003
Northwest	0.0002	0.0000	0.0002	0.0002

Figure 5.6 shows the nugget effects of the regression residuals of the spatial process models in Tokyo. In addition, Table 5.8 shows the test results for the

rate of change of the linear regressions of the nuggets. The nuggets of the four regions in Tokyo showed a downward trend in the southeast about until 2014 and an overall upward trend in the other regions. It is empirically known that real estate has strong individuality. The nugget indicates a local variation of the residuals between two observation points of the land price model, and the rise in the nugget indicates that the dispersion against the neighboring land relative to the land prices becomes stronger, which cannot be explained by the land price model of this analysis. During the analysis period, land price dispersion[22] became stronger in the three regions, except for the southeast, where they became weaker. Regarding the trends in each region, in the early downturn phase of official land prices, the nuggets also showed a downward trend in all four regions; however, it switched to an uptrend in all regions except the southeast. In addition, the southeast has been on a downward trend regardless of the 2008 financial crisis, but the three other regions do not show a downward trend, even in the downturn phase of prices that followed. As stated earlier, we can see a jump in the nugget of the southeast from 2018 to 2020 due to the additions of extremely high land prices of 32,300,000 yen per $1m^2$ in 2019 and 12,800,000 yen per $1m^2$ in 2020.

In summary, in the southeast, including Chuo, Chiyoda, and Shibuya wards, land prices in areas with strong brand power and areas that emphasize the living environment are high, and this tendency is becoming stronger year by year. The regression coefficients of the access index and the distance from the closest station in the southeast had the smallest negative impact on official land prices among the four regions. Moreover, quasi-fire/fire prevention districts' regulations also have negative impact prices yearly, but the impact has become smaller in recent years. However, the preference for brands has caused official land prices in the southeast to soar, a tendency that grows stronger every year. Furthermore, the effect of land price dispersion, one of the factors that determine land prices, has also decreased excluding the effects of outliers. In the sub-central area of Shinjuku and Ikebukuro and the cities included in the northeast such as Musashino and

[22]Once again, the dispersion of land prices here, the so-called individuality of land, includes land price variability that cannot be explained by the explanatory variables used in the land price model. For example, areas of caution on hazard maps, the existence of crime, local sunshine, noise conditions, and location of garbage collection sites

Mitaka, the negative impact of the regression coefficients of the distance from the closest station is becoming stronger. In the northeast, the negative impact of the regression coefficients of the distance to the closest station on land prices has been increasing since the 2008 financial crisis, as has the land price dispersion against neighboring lands. In the southwest, which includes cities like Fuchu, Kokubunji, Tachikawa and Hachioji, the regression coefficients of the access Index and distance from the closest station have a strong negative impact on official land prices, and it is possible to see that this impact becomes even stronger as time progresses. In addition, official land prices in residential/quasi-residential areas, which emphasize commerciality rather than the living environment, tend to be higher in the southwest. This characteristic is not observed in other regions and is becoming stronger yearly. The fact that quasi-fire prevention regulations increase official land prices yearly is also characteristic. The strength of the spatial correlation of residuals showed a remarkable uptrend until the 2008 financial crisis, but after that, the degree of uptrend has decreased as the land price dispersion increased. In the northwest, which includes west Tokyo and the city of Ohme, the annual changes in the regression coefficients of the access index and distance from the closest station are similar to those of the southwest, which have a stronger negative impact on official land prices than in the southwest. In particular, the impact of the coefficient of the distance to the closest station was the largest among the four regions. The strength of spatial correlation of residuals and the impact of land price dispersion on the neighboring land showed a trend similar to that of the southwest.

Lastly, regarding the impact of the coronavirus outbreak on the 2021 Tokyo land prices, we can see a decline in the land price, which was similar to the case of the 2008 financial crisis, whereas there were few changes in the determinants of land functions. The structural change of the spatial variogram could not be confirmed, except for stagnation in the expansion of spatial dispersion in southeast Tokyo. In the west, an increase in spatial dispersion in the regression residuals was observed, maintaining the previous movement; however, no significant structural changes were observed. However, since it takes a long time to take measures against corona, it is necessary to pay close attention to future land price determinants and structural changes.

6

Summary and Conclusion

As stated in Furuya (2004), the estimated regression coefficient may differ greatly depending on the estimation method, and the estimated land price model may differ depending on the explanatory variables. However, the common content of this study, which integrates the three studies, shows that there is a direction for the secular change in land price formation so far. We itemize it below. 1; The polarization and widening gap in land prices is recognized to be caused by changes in transportation convenience, and the impact is increasing year by year. 2; Except for the central part of the city, the variation in the residual of the land price model is increasing year by year. 3; There are areas where certain use areas are expensive or low due to resident preferences that differ from other areas. 4; It was confirmed that the spatial structure of the model residuals changed slightly depending on the selected explanatory variables.

In recent times, every time the official/standard land prices have been published, the media releases articles about the growing problem of polarization of land prices between metropolitan areas, such as Tokyo and Osaka, and other regions in Japan. However, from a global perspective, this study reveals that the polarization and widening gap in land prices is progressing even within Tokyo. Comparing the area from the central to the western parts of the 23 wards of Tokyo and the area from the western part of Kitatama to the eastern part of Nishitama or other areas, polarization is progressing, especially because the parameters on the explanatory variables that represent transportation convenience changed, so that the land price differences increased over time.

The variation of a local population according to the level of convenience in the area is considered one of the reasons for land price polarization and widening price gap. We confirmed that between land prices between areas where transportation is convenient and areas where transportation is not convenient polarized and dispersed due to changes in the parameters of the access index and the distance from the nearest station over time, and this tendency is becoming stronger. According to the "Reference Material on Official Land Prices" published by the MLIT in 2012, there is a very strong positive correlation between population density in the 47 prefectures in the country and the average official land prices of residential areas, with a correlation coefficient of 0.8950. In addition, Saita et al. (2016) states that since the collapse of the bubble economy, there is a positive correlation between the working-age population ratio and land prices, and a negative correlation between the elderly population ratio and land prices. In this study, we calculated the daytime population density of municipalities that contain official target sites in Tokyo and found that there is a very strong positive correlation, with a coefficient of 0.7824, between the official land prices after logarithmic conversion and population density. The first time that the land policies in post-bubble Japan faced population decline (which started in 2008) was in the "Mid-to-Long Term Vision of Land Policies" in 2009, but it was limited to a broad confirmation of the current situation in the entire of Japan, and it did not get to the point of addressing the polarization of land prices. From a different perspective, the concentration of population in urban areas and the depopulation of rural areas have been serious issues in Japan since the late 1960s. Since the Great East Japan Earthquake of 2011, a population policy that proposes shifting governmental functions away from Tokyo once again gained momentum following the reawakening of the disadvantages of concentrating all of them in the capital. However, nowadays, the decentralization of urban functions is not progressing slowly, probably because the residents avoid the disadvantage of the convenience of urban decentralization. It is also noted that the suburban areas began to face challenges caused by the aging of their residents, and an increase in vacant housing may cause the rise in spatial polarization (Kubo, 2020; Kubo and Yui, 2020). Since 2008, Japan's total population has been declining. In the future,

even in Tokyo, land prices are expected to become more polarized in the 23 wards and the Tama area, where the population is declining significantly.

Next, the impact of land price dispersion of local area is increasing. The dispersion of land price includes areas of caution on hazard maps, the existence of crime, local sunshine and noise conditions, and the location of garbage collection sites. Increasing the nugget effect of the spatial variogram and the increase in the range of constant terms, the land price dispersion other than the explanatory variables used in this study strengthened. Increasing the range of each local regression coefficient of the MGWR model indicates that price dispersion of the land, which is a factor in land price formation, is gradually strengthened.

From the above results, after the burst of the economic bubble, the land prices in Tokyo show two different developments: the progress of polarization and widening gap in land prices in global area and the increase in land price dispersion in local area. In the future, against the backdrop of population decline, the trends in each region will become clearer and the local differences in the living environment will be reflected more in land prices. As the 23 wards are expected to be affected by the population decline, so the possibility of a fall in land prices is undeniable due to the fact that land prices are originally high. Also, from the transition of the local regression surface of the low-rise residential area dummy, a low-rise residential area set in a high-class residential area is more expensive from the central part to the northern part of the 23 wards and the area around Kokubunji city and Tachikawa city, and cheaper in the area centered on Hachioji City and Chofu City. From the transition of the local regression surface of the residential/quasi-residential area dummy, the residential and semi-residential areas that focus on convenience and commerciality are more expensive in western Tokyo, centered on Hachioji City, and cheaper in eastern Tokyo, centered on the 23 wards. They are areas with no significant difference between daytime and nighttime population densities, there are segregations within use districts. These two tendencies were established before the 2008 financial crisis, and have been continuous and stable after the crisis and may be one of the causes of the polarization of land prices. From these results, it seems that the basic plan for base development, among others, has had a specific effect in the Tama area.

In the wake of the coronavirus outbreak, the official land prices in Tokyo in 2021 have declined whereas other factors such as the determinants of the land prices and spatial structures of the regression residuals have been constant. However, it is still possible to confirm that the spatial correlation of the regression residuals in the 23 wards including the city center where the average land price is the highest is stagnated. The impact of corona is likely to be limited and temporary in central Tokyo. It is quite possible to see that the impact is in transition. Future land price trends need to be monitored. In addition, the momentum for decentralization, which once went down, is increasing again in the long-term measures against corona.

As a future research topic, it will be prudent to study how to assess the difference between official land prices and actual prices. In addition to the official land prices, the public evaluation of land prices involves the benchmark land price, roadside land price, and property tax valuation, but assessing the differences among these four is also a topic for future research. Occasionally, the MLIT or other related authorities publish the difference between the actual price and official land price as the rate of divergence, but the point where the transaction took place and the official target point are seldom the same; thus far, no quantitative relationship between the two prices has been found. Considering how strong the individual factors of land are, solving this problem will represent a significant contribution to the land price information system. One of the other future research topics is to build an MGWR model that includes global explanatory variables. We will consider constructing a mixed MGWR model that quantitatively measures the characteristics of the explanatory variables and enhances the consistency of the model in future studies. Additionally, we extend the model to the space-time dimension. For that matter, the same can be said for the spatial process model. Huang et al. (2010) and Fotheringham et al. (2015) proposed a geographically and temporally weighted regression model (GTWR) with weights for both the space and time dimensions for Calgary, Canada, and London, respectively, and reported that the GTWR model improved the forecasting accuracy. Further, Wu et al. (2019) proposed an MGTWR model, which is a multiscale version of the GTWR, and estimated the land price model

in Shenzhen, Guangdong Province, China. Analyzing using these models is also a future research topic.

References

Bagan, H. and Yamagata, Y. (2012). Landsat analysis of urban growth: How Tokyo became the world's largest megacity during the last 40 years. *Remote Sensing of Environment*, 127:210–222.

Basu, S. and Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. *Journalof Real Estate Finance and Economics*, 17(1):61–85.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlations (with discussion). *Journal of the American Statistical Association*, 80(391):580–619. doi: 10.2307/2288473.

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.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39(3):497–524.

Chen, M.-C., Tsai, I.-C., and Chang, C.-O. (2007). House prices and household income: Do they move apart? evidence from Taiwan. *Habitat International*, 31(2):243–256.

Chica-Olmo, J. (2007). Prediction of housing location price by a multivariate spatial method:cokriging. *Journal of Real Estate Research*, 29(1):91–114.

Chica-Olmo, J., Cano-Guervos, R., and Chica-Olmo, M. (2013). A coregionalized model to predicthousing prices. *Urban Geography*, 34(3):395–412.

Chica-Olmo, J., Cano-Guervos, R., and Chica-Rivas, M. (2019). Estimation of housing price variations using spatio-temporal data. *Sustainability*, 11(6):1551.

Chiles, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty, Second Edition*. Wiley, New York.

Cho, S. H., Bowker, J. M., and Park, W. M. (2006). Measuring the contribution of water and green space amenities to housing values: An application and comparison of spatially weighted hedonic models. *Journal of Agricultural and Resource Economics*, 31(3):485–507.

Cressie, N. A. C. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17(5):563–586.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Cressie, N. A. C. and Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Mathematical Geology*, 12(2):115–125.

Dubin, R. (1998). Predicting house prices using multiple listings data. *Journal of Real Estate Finance and Economics*, 17(1):35–59.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, Chichester.

Fotheringham, A. S., Charlton, M. E., and Brunsdon, C. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30(11):1905–1927.

Fotheringham, A. S., Crespo, R., and Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, 47(4):431–452.

Furuya, T. (2004). Bayesian geographically weighted regression model and its application for land price model estimation (In Japanese). *Journal of the City Planning Institute of Japan*, 39(3):787–792.

Goodman, A. C. (1978). Hedonic prices, price indices and housing markets. *Journal of Urban Economics*, 5(4):471–484.

Harris, P., Brunsdon, C., and Fotheringham, A. S. (2011). Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stochastic Environmental Research and Risk Assessment*, 25(2):123–138.

Hasegawa, K., Tahishita, M., and Shimizu, C. (2006). Consideration on hedonic analysis and its application—using residential land sales data in metropolitan areas— (In Japanese). *2006 Basic Survey Business Report on Land Policy*. 341–388.

Hatta, T. and Ohkawara, T. (1993). Population, employment, and land price distributions in the Tokyo metropolitan area. *The Journal of Real Estate Finance and Economics*, 6(1):103–128.

Hein, C. and Pelletier, P. (2006). *Cities, Autonomy, and Decentralization in Japan*. Taylor & Francis.

Helbich, M., Brunauer, W., Vaz, E., and Nijkamp, P. (2014). Spatial heterogeneity in hedonic house hrice models: The case of Austria. *Urban Studies*, 51(2):390–411.

Hengl, T., Heuvelink, G. B., and Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33(10):1301–1315.

Huang, B., Wu, B., and Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3):383–401.

Inoue, R., Shimizu, E., Yoshida, Y., and Li, Y. (2009). Visualization of spatial distribution and temporal transition of land price in Tokyo 23 wards based on spatio-temporal kriging interpolation (In Japanese). *Theory and Applications of GIS*, 17(1):13–24.

Inoue, R., Sugiura, A., Yoneyama, S., and Nakanishi, W. (2016). A proposal for land price information service based on comparison between appraised prices and transaction prices (In Japanese). *JSCE Proceedings D3 (Civil Engineering Planning)*, 72(1):1–13.

Kanno, Y. and Shiohama, T. (2020). About the secular change of the spatial variogram in the universal kriging of the official land price of Tokyo (In Japanese). *Japanese Journal of Applied Statistics*, 49(2):47–69.

Kanno, Y. and Shiohama, T. (2021). Geographically weighted regression for Tokyo official land prices and their temporal variation (In Japanese). *Research Memoir of Official Statistics*, 78:55–74.

Kanno, Y. and Shiohama, T. (2022). Land price polarization and dispersion in Tokyo: A spatial model approach. *Asia-Pacifc Journal of Regional Science*, 6(2):807–835.

Kato, N. (2005). On estimation and use of real estate price functions (In Japanese). *Journal of the Japan Statistical Society*, 34(2):131–161.

Kikuchi, T. (2008). Recent progress in Japanese geographical studies on sustainable rural system : Focusing on recreating rurality in the urban fringe of the Tokyo metropolitan area. *Geographical Review of Japan 81*, 81(5):336–348.

Kitasaka, S. (2012). Land prices and macroeconomic variables in Japan (In Japanese). *The Doshisha University Economic Review Keizaigaku-ronso*, 64(2):81–104.

Kubo, T. (2020). *Divided Tokyo*. Springer.

Kubo, T. and Yui, Y. (2020). *The Rise in Vacant Housing in Post-growth Japan*. Springer.

Kuntz, M. and Helbich, M. (2014). Geostatistical mapping of real estate prices: An empirical comparison of kriging and cokriging. *International Journal of Geographical Information Science*, 28(9):1904–1921.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, Boca Raton.

Leung, Y., Mei, C. L., and Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, 32(1):9–32.

Li, J. and Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, 6(3-4):228–241.

Li, Y. and Monzur, T. (2018). The spatial structure of employment in the metropolitan region of Tokyo: A scale-view. *Urban Geography*, 39(2):236–262.

Lu, B., Brunsdon, C., Charlton, M., and Harris, P. (2017). Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 31(5):982–998.

Lu, B., Harris, P., Charlton, M., and Brunsdon, C. (2015). Calibrating a geographically weighted regression model with parameter-specific distance metrics. *Procedia Environmental Sciences*, 26:109–114.

Masunari, K. (2007). Analysis of official land price by kriging (In Japanese). *Bulletin of the Japanese Society of Computational Statistics 2007*, 18(2):107–122.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8):1246–1266.

Ministry of Land, Infrastructure, Transport and Tourism (2020). Examination work of appraisal evaluation method corresponding to individualization and multipolarization of land prices (In Japanese). Techreport, Ministry of Land, Infrastructure, Transport and Tourism, Land/ Construction Industry Bureau,
`https://www.mlit.go.jp/totikensangyo/content/001402950.pdf`.

Mochida, N. (2008). *Fiscal Decentralization and Local Public Finance in Japan*. Routledge.

Montero, J. and Larraz, B. (2011). Interpolation methods for geographical data: housing and commercial establishment markets. *Journal of Real Estate Research*, 33(2):233–244.

Morioka, T. and Fujita, T. (1995). Economic value of environmental quality and environmental services and their reflection in the market—comparison of the principles of the three evaluation methods and their application examples— (In Japanese). *The Japanese Journal of Real Estate Sciences*, 9(4):10–19.

Nozawa, Y. and Higuchi, Y. (2001). A study of derivation and changes of excess commuting in the Tokyo metropolitan area (In Japanese). *Studies in Regional Science*, 31(1):149–161.

Omura, K. (2010). Structure and problems of the Tokyo metropolitan area: The future of peripheries. *disP-The Planning Review*, 46(181):36–50.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Saita, Y., Shimizu, C., and Watanabe, T. (2016). Aging and real estate prices: Evidence from Japanese and US regional data. *International Journal of Housing Markets and Analysis*, 9(1):66–87.

Sakai, T., Kawamura, K., and Hyodo, T. (2016). Logistics facility distribution in Tokyo metropolitan area: Experiences and policy lessons. *Transportation Research Procedia*, 12:263–277.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. CRC press.

Seya, H. and Tsutsumi, M. (2014). *Spatial Statistics—From Natural Sciences to Humanities and Social Sciences—* (In Japanese). Asakura Shoten.

Shimizu, C., Nishimura, K., and Watanabe, T. (2016). House prices at different stages of the buying/selling process. *Regional Science and Urban Economics*, 59:37–53.

Sutton, G. D. (2002). Explaining changes in house prices. *BIS Quarterly Review*, 32(1):46–60.

Tabuchi, T. (1996). Quantity premia in real property markets. *Land Economics*, 72(2):206–217.

Tiwari, P. (2000). Housing demand in Tokyo. *International Real Estate Review*, 3(1):65–92.

Tokuda, M. (2009). The analysis of the land value by the hedonic approach: The residential quarter in shiga prefecture as a subject (In Japanese). *Hikone Ronso*, 381:183–205.

Tolosana-Delgado, R. and Pawlowsky-Glahn, V. (2007). Kriging regionalized positive variables revised: Sample space and scale considerations. *Mathematical Geology*, 39(6):529–558.

Tsutsumi, M. and Seya, H. (2012). Basics of applied spatial statistics: Spatial statistics and spatial econometrics (In Japanese). *Proceedings of the Institute of Statistical Mathematics*, 60(1):3–25.

Uesugi, D. (2012). Local model estimation of land price level in each small area: A case study in Saitama city (In Japanese). *Setsunan Economic Review*, 2(1–2):1–20.

Wang, R., Derdouri, A., and Murayama, Y. (2018). Spatiotemporal simulation of future land use/cover change scenarios in the Tokyo metropolitan area. *Sustainability*, 10(6):2056.

Wheeler, D. C. and Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2):161–187.

Wu, C., Ren, F., Hu, W., and Du, Q. (2019). Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, 33(3):489–511.

Yang, W. (2014). *An Extension of Geographically Weighted Regression with Flexible Bandwidth*. PhD thesis, St Andrews.

Yokohari, M., Brown, R. D., and Takeuchi, K. (1994). A framework for the conservation of rural ecological landscapes in the urban fringe area in Japan. *Landscape and Urban Planning*, 29(2-3):103–116.

Zheng, X.-P. (2007). Economies of network, urban agglomeration, and regional development: A theoretical model and empirical evidence. *Regional Studies*, 41(5):559–569.

Zhou, H. and Gao, H. (2020). The impact of urban morphology on urban transportation mode: A case study of Tokyo. *Case Studies on Transport Policy*, 8(1):197–205.

Zietz, J., Zietz, E. N., and Sirmans, G. S. (2008). Determinants of house prices: A quantile regression approach. *The Journal of Real Estate Finance and Economics*, 37(4):317–333.