SPATIAL CLUSTERING OF HOUSING CONSTRUCTION IN THE TOKYO METROPOLITAN AREA: AN APPLICATION OF SPATIALLY CLUSTERED FIXED-EFFECTS AND SPATIALLY CORRELATED RANDOM-EFFECTS MODELS

Atsushi Yoshida *
University of Tsukuba

Tatsuhiro Shichijo
Osaka Prefecture University

Abstract

We propose two types of econometric model, a *spatially clustered fixed-effects model* and a *spatially correlated random-effects model*, to examine area-based panel data. We investigate what factors influence housing construction in the Tokyo Metropolitan Area, incorporating unobservable factors, local regulatory differences in housing development and spillovers of local public or private goods, which may cause spatial clustering or correlation of housing construction. The unobservable area-effects are large in the east, west and north areas of the TMA but small in the south, where regulations against development are more severe than in the other areas. Clusters are found in huge cities.

Key words: cluster-effects model, housing construction, area-based panel data, spatial correlation, spillover effects

JEL classification: C23, R12, R15, R21

---

*CORRESPONDENCE: Atsushi Yoshida, Institute of Policy and Planning Sciences, University of Tsukuba, 1-1-1, Ten-noudai, Tsukuba, Ibaraki 305-8573, Japan. TEL&FAX: +81-29-853-5095. Email: ayoshida@sk.tsukuba.ac.jp

1 INTRODUCTION

In this paper, we will examine the factors that affect housing construction in the Tokyo Metropolitan Area (TMA) using area-based panel data and also consider relevant econometric issues related to the use of such data.

Housing construction is one of the main topics of urban economics. Many researchers have empirically investigated the factors influencing housing construction and their magnitudes - for example, Topel and Rosen (1988), DiPasquale and Wheaton (1994), and Mayer and Somerville (2000), among others. Unlike the authors of the current paper, these authors were concerned with total housing construction in a country, not with spatial disparity within a metropolitan region. Although Japan suffered from a serious economic depression after the burst of the bubble economy in the 1990s, a period sometimes referred to as "the lost decade", housing construction in the TMA increased drastically in the mid-1990s. This increase was largely attributable to a combination of factors, including a sharp decline in land prices, low interest rates and tax policies favorable to housing construction. Although housing construction has increased, at the same time we have observed that this increase is not spatially uniform over the TMA, but is uneven.

The purpose of this paper is, first, to investigate the factors that cause changes in the number of houses being constructed and the spatial disparity. The TMA consists of jurisdictional areas called municipalities. Municipalities can determine regional public spending policies or make regulations on economic activities that relate to the living standards of residents or to the amenities of the area. For example, via financial support, municipalities strongly influence the quality and quantity of healthcare services for residents and of facilities like elementary and junior high schools. In addition, urban planning or land use regulations, like zoning, sometimes restrict housing construction and the development of unused land for residential houses.

People who are concerned with the availability of healthcare and schooling facilities seriously consider which municipality is the best to live in before purchasing or constructing new houses. Assuming people are heterogeneous, those who have the same preferences or the same socio-economic characteristics tend to live together in the same area, as the Tiebout model predicts (see Tiebout 1956, Gramlich and Rubinfeld 1982 and Aaronson 2001). People who concerned with the healthcare and education of their children prefer to live in a municipality where the quality of education or of the public health care services is higher than in other municipalities. These demand-side factors partly cause the spatial disparity of housing construction within the TMA.

The regulations on land use make large housing developments costly or difficult. In general, most of the houses supplied in the TMA are condominiums. Although large housing developments enable developers to cut costs by enjoying scale economies, they may be costly for municipalities. This is because the municipalities have to provide additional facilities for new residents, which may involve constructing parks and elementary schools, hiring teachers and so on. Strict regulations, such as height restrictions on condominiums, or proportion restrictions, regulating the area of public open space in relation to the total developing area, increase the cost of development. Thus, differences in the severity of such regulations among municipalities, which are supply-side factors, cause differences in the number of houses being constructed and the observed spatial disparity.

As some of the factors that affect housing construction and the spatial disparity are usually unobservable or difficult to evaluate numerically, we cannot incorporate them as explanatory variables of an econometric model. For example, some municipalities impose regulations that force developers to make an agreement with the residents who live near the construction area of a planned condominium. It is very difficult to evaluate the severity of the regulations because it depends upon the characteristics of the residents or the area. Therefore, we have to take these as unobservable factors. Moreover, the regulations on housing construction themselves are sometimes unobservable because municipalities regulate developers through implicit codes of development, which are not necessarily publicly available. Often these regulations are similar to those in adjacent municipalities because each area faces similar problems in relation to housing construction. Therefore, the regulations can be regarded as unobservable factors that are spatially clustered.

In addition, it is often difficult to evaluate how the quality or quantity of local public or private goods/services impacts on housing construction. Hence, these factors also have to be taken as unobservable. An attractive facility in a municipality, like a new, well-designed museum, library or shopping mall built in a redeveloped inner area, may increase housing construction not only in that municipality but also in adjacent municipalities. Such a situation involves spillover effects from local public or private goods and services. Network infrastructure lying across the borders of municipalities, such as railways, subways, highways, roads and so on, will enlarge the spillover effects. Spillover effects are one of the sources of spatial clusters due to unobservable factors.

The second purpose of this paper is to propose econometric models that enable us to examine the clustering structure of these unobservable factors. We are concerned not only with how large the effects of these unobservable factors are within municipalities, but also with the question of which adjacent areas constitute a spatial cluster of the unobservable factors. These unobservable effects can be expressed as area-specific effects in panel data models. If we were not concerned with the clustering structure of the effects, but only with the parameters of observable explanatory variables, we could adopt a standard fixed-effects model that has area-specific effect parameters for each area. Then, we could obtain the within estimates for parameters of concern (see Hsiao 1986 for details). However, as we are concerned with the clustering structure, we require models that enable us to capture this structure. We will propose two types of model that correspond to the fixed- or random-effects models, namely, a *spatially clustered fixed-effects model* (SCFEM) and a *spatially correlated random-effects model* (SCREM).

In section 2, we will explain both the SCFEM and the SCREM. In relation to the SCFEM, we have to consider a statistical method that will detect which areas belong to which cluster. From a statistical viewpoint, this issue is regarded as a type of model selection problem. We adopt a resampling method, namely a *leave-one-out cross-validation*, since the method is robust to distributional assumptions and the calculation is easily implemented for a linear model. The method was introduced by Stone (1974) and Geisser (1975) and its applications for broad model selection problems are discussed in Davison and Hinkley (1997). However, it has not yet been applied to detect clusters with area-based data. The selection procedures, the forward-stepwise (FS) and backward-stepwise (BS) methods, are also discussed. Section 3 shows the results of the simulation studies, detailing how well the *leave-one-out cross-validation* works to detect the clusters. In addition, we compare the SCFEM estimates with the within estimates and find that the former are more efficient than the latter.

In the SCREM, spatial correlation of the area-specific errors between adjacent areas is modeled. Spatial correlation is recently focused and is often discussed in relation to the spatial interaction of agents' behaviors. Netz and Taylor (2002) empirically test implications from location theory using the location of Los Angeles-area gasoline stations in physical space and in the space of product attributes. They use the model proposed by Anselin (1988), where the dependent variable is spatially correlated. On the other hand, Pinkse and Slade (1998) employ the Probit model with the errors of the underlying linear model spatially correlated. We use the conditional Gaussian model for the area-specific effects errors. Cressie (1993) discussed the differences between the simultaneous Gaussian model, which is like Anselin's model, and the conditional Gaussian model. In panel data models, if we follow Anselin's model, we have to deal with the correlation between the explanatory variable, namely, the adjacent area's dependent variable, and the area-specific errors. This is also the case with panel data models that use a lagged dependent variable as a regressor. This will make the estimation more complex and difficult.

In section 3, the two models, the SCFEM and the SCREM, are employed to estimate the reduced form of housing-construction functions in the Tokyo Metropolitan Area using area-based panel data for three years. The estimated function includes both demand- and supply-side factors. As demand-side factors, we mainly consider the economic attributes of an area, for example, income per household and time-distance to Tokyo Station, and amenities or dis-amenities, for example, public spending for education or population density. The supply-side factors, like construction costs influenced by land use regulations or spillovers from local public or private goods, are treated as unobservable factors. Section 4 concludes the paper.

2 ECONOMETRIC MODELS WITH AREA-EFFECTS

Let us consider a model with area-based panel data. Assume that we have observations of m areas for T periods. The area-effects model is expressed as follows:

$$y_{it} = x_{it}\beta + u_i + v_{it}, \quad i = 1, \ldots, m; \; t = 1, \ldots, T \tag{1}$$

The $\{y_{it}\}$ and $\{x_{it}\}$ (a $1 \times K$ vector) are dependent and independent variables that represent the socio-economic properties of the ith area. The β ($K \times 1$) represents the relationship between them which is of concern for researchers. The u_i, an area-effect, represents unobservable socio-economic characteristics in the ith area. The v_{it} is an error that is independent and identically distributed for all i and t. Note that $\{x_{it}\}$ does not include a constant term in the SCFEM but does in the SCREM, as is the case for standard panel models.

Before explaining the econometric models, we have to clarify the meanings of the terms that we use, namely, region, area and cluster. The region is the geographical domain or space where economic activities are observed. In the empirical study in section 3, the Tokyo metropolitan area is adopted to represent the region. An area is defined as a minimum areal unit of the region. It is often a jurisdictional unit or a statistical unit. We adopt the municipality as an area in the empirical study. A cluster is a collection of areas, in which an area should be adjacent to at least one of the other areas that constitute the cluster.

2.1 THE SPATIALLY CLUSTERED FIXED-EFFEXTS MODEL

Let us assume there are q ($q \ll m$) clusters. As they are unobservable we have to determine the number of clusters, q, and their structure statistically. Then, the area-effects, $u_i, i = 1, \ldots, m$, should be classified into q classes, say $u_1, \ldots, u_q$. The SCFEM is expressed as follows:

$$y_{it} = \mathbf{x}_{it}\beta + u_q + v_{it}, \quad i = 1, \ldots, m; \; t = 1, \ldots, T; \quad \text{if } i \in q\text{th cluster} \tag{2}$$

The vector form of eq.(2) is:

$$\mathbf{y}_t = X_t\beta + D_0\mathbf{u}_0 + \mathbf{v}_t, \quad t = 1, \ldots, T. \tag{3}$$

The $\mathbf{y}_t$ $(m\times 1)$, X_t $(m\times K)$ and $\mathbf{v}_t$ $(m\times 1)$ are defined as $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$, $X_t = (\mathbf{x}'_{1t}, \ldots, \mathbf{x}'_{mt})'$ and $\mathbf{v}_t = (v_{1t}, \ldots, v_{mt})'$, respectively. The $\mathbf{u}_0 = (u_1, \ldots, u_q)'$ is a parameter vector to be estimated, which we will call cluster-effects. D_0 is an $m \times q$ matrix of dummies that indicates which area belongs to which cluster. For example, if sth area and lth area belong to the same cth cluster, then the cth element of the sth and the lth rows of D_0 are the same, namely 1, and the other elements of the rows are 0s. We will call the matrix a cluster-dummy matrix.

Now, we will consider how we can decide the rank of D_0 (namely q), identify the structure of D_0 based on the model eq.(3) statistically and estimate $\mathbf{u}_0$ and β. That is, we have to find how many clusters are there, which area belongs to which cluster and estimate the parameters of concern at the same time. Without the classification of area-effects into cluster-effects, we cannot obtain consistent estimates of $\mathbf{u}_0$ or a more efficient estimate of β than the within estimates.

From the statistical point of view, detecting the rank and structure of D_0 is regarded as a model selection problem. In this case, the largest model is the case where $u_1, \ldots, u_m$ have different values. That is, they are not classified into fewer classes. This is the standard fixed-effects model. On the other hand, the smallest model is the case where $u_1, \ldots, u_m$ have the same value, that is, they all are classified into one class, which is called the pooled model. There are a lot of possibilities of classification between the largest and the smallest models.

In a statistical model selection context, there are two major methods, one using information-based selection criteria, namely AIC, BIC and SBIC (see Lütkepohl (1991), for example), and the other using a resampling method-based selection criterion. The formulas of the former criteria are easily obtained as they are simple functions of the likelihood and the number of parameters. On the other hand, they depend heavily upon assumptions about the distributions of the random variables. Although the latter criterion is demanding computationally, it does not require any assumptions about exact distributions. Stone (1977) has proved the asymptotic equivalence of one form of the latter method, the *cross-validation*, to AIC.

One of the model selection criteria with the resampling methods we employed here is the *aggregate prediction error* (*APE*). In a regression model, it is defined as:

$$\Delta = \frac{1}{n}\sum_{j=1}^{m} E((Y_{+j} - \eta(X_j, \hat{F}))^2 | \hat{F})$$

where Y_{+j} is one of possible realizations at X_j, $\eta(X_j, \hat{F})$ being an estimate of the mean response function and $\hat{F}$ is an empirical joint distribution of Y and X that represents data. One of the estimates of the *APE* is obtained by using *leave-one-out cross-validation*, which is defined as:

$$\hat{\Delta}_{CV} = \frac{1}{n}\sum_{j=1}^{m}(y_j - \eta(x_j, \hat{F}_{-j}))^2$$

where $\hat{F}_{-j}$ represents the $n-1$ observations $\{(\mathbf{x}_k, y_k), k \neq j\}$. In a linear regression model, we have $\eta(\mathbf{x}_j, \hat{F}_{-j}) = \mathbf{x}_j\hat{\beta}_{-j}$ where $\hat{\beta}_{-j}$ is the estimate using only the data of Y and X, excluding the jth sample. See chapter 6 of Davison and Hinkley (1997) for more detail of the criterion and its estimation methods.

Let us explain the model selection procedure. The matrix form of eq.(3) is:

$$\mathbf{y} = X\beta + (\mathbf{1}_T \otimes D_0)\mathbf{u}_0 + \mathbf{v}, \tag{4}$$

where $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$, $X = (X_1', \ldots, X_T')'$ and $\mathbf{v} = (\mathbf{v}_1', \ldots, \mathbf{v}_T')'$. The purposes are to find the structure of clusters which is represented by D_0 and to estimate $\mathbf{u}_0$ and β. They are regarded as a sort of problems that select a set of explanatory variables to determine the rank and structure of D_0. In principle, to select these from all the possible combinations of explanatory variables, we have to calculate $\hat{\Delta}_{CV}$ for all combinations and select the combination that attains the minimum value. However, this is almost impossible because there are so many combinations.

In general, forward-stepwise (FS) or backward-stepwise (BS) methods are often used for selecting combinations of explanatory variables when trying all combinations is impossible. We employ both of these methods. The FS starts with the smallest model, dividing clusters or combining adjacent areas/clusters, and selects the combination that attains the smallest *APE*. The BS starts with the largest model, doing the same as the FS to attain the smallest *APE*.

Note that our setup of the variable-selection problem is different from the usual setup in the following two ways. First, a new explanatory variable is created by splitting a column vector of an existing cluster-dummy matrix or an explanatory variable is eliminated by integrating a column vector to an existing column vector of the matrix. For example, when the cth cluster composed of sth and lth adjacent areas is split into two areas, then the cth column of the cluster-dummy matrix is split into two columns. When the hth area is combined with the sth cluster, then the column vector corresponding to the hth area is eliminated and the sth element of the hth row vector turns from 0 to 1. Second, the combining process is subject to the adjacency restriction. Thus, two areas that share no border cannot be combined into a cluster. When the structure of clusters, which is represented by the cluster-dummy matrix, is changed in the searching process, then the adjacency of areas/clusters is also altered. Thus, we have to ensure that the adjacency information corresponds to the structure of the areas/clusters at every step of the procedure.

Let $A(m)$ be an adjacent matrix of areas in the largest model, which is an $m \times m$ symmetric dummy matrix indicating which areas are neighbors of an area. For example, if the (i, j) element of the matrix is 1, then the ith area is adjacent to the jth area. All of the diagonal elements are 1s by definition. The argument of the adjacent matrix, m, means the number of areas/clusters of the region. We call it the dimension of the adjacent matrix. The term dimension is also used for the cluster-dummy matrix, where it is equivalent to the number of column vectors of the matrix. When the model is at its largest, the dimension is m. On the other hand, if all areas belong to just one cluster, then the dimension is 1.

We will explain the process of the FS method first. In the initial condition, $D(1) = \mathbf{1}_m$ and $A(1) = \{1\}$. Let us redefine y_j and $\mathbf{x}_j$ as the jth element and jth row vector of $\mathbf{y}$ and X for all $j = 1, \ldots, mT$, respectively. Then *APE(k)* is calculated as:

$$APE(k) = \frac{1}{mT} \sum_{j=1}^{mT} (y_j - \mathbf{x}_j \hat{\beta}_{-j} - d_j(k) \hat{u}_{-j})^2$$

where $d_j(k)$ is the jth column of $\mathbf{1}_T \otimes D(k)$, $\hat{u}_{-j}$ and $\hat{\beta}_{-j}$ are the OLS estimates using the data, excluding $y_j, d_j(k)$ and $\mathbf{x}_j$. We use two steps, a dividing step and a combining step, to search for the cluster-dummy matrix that minimizes the *APE*. Both the FS and BS procedures consist of the following two steps:

STEP1 (Dividing Step): Select an area from the region and allocate it a different area-effect parameter. Note that the selection is conducted from all of the areas in the region, even if the area is already a component of a cluster. We select an area from a cluster even in cases where, due to the selection, the remaining areas no longer constitute a cluster. This is often the case where a selected area connects two areas or clusters that share no border. After conducting all possible selections and calculating their APEs, choose a division that attains the minimum of the APE. This step expands the dimension of $D(\cdot)$ and $A(\cdot)$.

STEP2 (Combining Step): Select two adjacent areas/clusters and combine them. This implies that we allocate them the same new area-effect parameter. After conducting all possible selections and calculating their APEs, choose the combining that attains the minimum of the APE. This step shrinks the dimension of $D(\cdot)$ and $A(\cdot)$.

The FS procedure first repeatedly conducts step 1 only to find a cluster-dummy matrix that minimizes APE. For example, if the jth area is selected in the first step from the initial condition, $D(2)$ becomes an $m \times 2$ matrix composed of two vectors: a vector where all the elements are 1 except the jth element, which is 0, and a vector where all the elements are 0, except the jth element, which is 1. $A(2)$ becomes a 2×2 matrix with all elements being 1. After the repetition of the procedure, we obtain $D(l)$ and $A(l)$ which corresponds to the temporally optimized division.

Then, we search for other possible combinations of areas around this temporally optimized division using the stepwise method that will make APE smaller. One cycle of the stepwise procedure consists of steps 1 and 2. We conduct both steps once each, compare the APEs and choose either the division or the combining. Of course, the combining always dominates the division in the first stepwise process, since we start at a state where further divisions never gain in APE. However, note that the division will gain in APE after at least one stepwise procedure going ahead. If one of the stepwise procedures does not give a smaller APE than the previous procedure, then we quit the procedure and take the structure of areas/clusters and estimates of the parameters in the previous cycle as the optimized model.

In the BS procedure, we start with $D(m) = I_m$ and $A(m)$. First, the combining step involves finding the cluster-dummy matrix that minimizes APE. Then, the stepwise procedure is repeated as per the FS method. A simple example of how to obtain the cluster-dummy matrix and the adjacent matrix is shown in the appendix.

2.2 THE SPATIALLY CORRELATED RANDOM-EFFECTS MODEL

In the SCREM, we specify the density function of the ith random area-effect, u_i, conditional on its adjacent areas as follows:

$$f(u_i|\{u_j, j \in \mathcal{N}_i\}) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left[-(u_i - m(u_j; j \in \mathcal{N}_i))^2/\sigma_u^2\right], \tag{5}$$

where $\mathcal{N}_i$ is a set of adjacent areas of the ith area and the conditional mean is assumed to be a linear combination of these areas as:

$$m(u_j; j \in \mathcal{N}_i) = \sum_{j \in \mathcal{N}_i} \lambda u_j = \sum_{j=1}^{m} \lambda c_{ij} u_j$$

where c_{ij} is a dummy of adjacency that satisfies $c_{ij} = c_{ji}$, $c_{ii} = 0$ and $c_{ij} = 1$ if area i and area j are adjacent to each other, namely share a border between them. Otherwise, $c_{ij} = 0$. The parameter λ shows how large the influence of the adjacent areas is. Note that an area is influenced by the adjacent areas regardless of their directions, the length of the border they share or socio-economic relationships between each pair of adjacent areas. This may be rather restrictive in expressing spatial correlation with a statistical model, as we know only the average effects of the neighbors with this model.

Then, the joint distribution of $\mathbf{u} = \{u_1, \ldots, u_m\}$ is obtained as follows:

$$\mathbf{u} \sim N(\mathbf{0}_m, \sigma_u^2 \Sigma), \quad \Sigma \equiv (I - \lambda C)^{-1} \tag{6}$$

where C is an $m \times m$ matrix with its i, jth element being c_{ij} and thus, $C = A(m)$. A detailed explanation of how to derive the joint distribution is found in Cressie (1993).

The $I - \lambda C$ should be non-singular for the well-definition of Σ. The λ should be restricted for the non-singularity. Let $e_1 < \cdots < e_m$ be eigenvalues of C. There are three cases for the restriction of λ: first, if $0 < e_1$, then $\lambda < e_m^{-1}$; second, if $e_1 < 0 < e_m$, then $e_1^{-1} < \lambda < e_m^{-1}$;

and finally, if $e_m < 0$, then $e_1^{-1} < \lambda$. (See chapter 6 of Cressie (1993) for more detail.) We will discuss the restriction again in estimating housing-construction functions in section 4.

Now, we will briefly sketch the estimation of the SCREM using the maximum likelihood method. We define the $mT \times 1$ vector $\mathbf{w}$ containing two types of errors as:

$$\mathbf{w} = \mathbf{1}_T \otimes \mathbf{u} + \mathbf{v}.$$

The variance of $\mathbf{w}$ is obtained as follows:

$$\begin{aligned}\Omega \equiv Var(\mathbf{w}) &= (\mathbf{1}_T\mathbf{1}_T') \otimes \sigma_u^2\Sigma + \sigma_v^2 I_{mT} \\ &= \sigma^2[(\mathbf{1}_T\mathbf{1}_T') \otimes \rho\Sigma + (1-\rho)I_{mT}],\end{aligned}$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\rho = \sigma_u^2/\sigma^2$. We define R as: $R = (\mathbf{1}_T\mathbf{1}_T') \otimes \rho\Sigma + (1-\rho)I_{mT}$. Then, the likelihood function of the parameters to be estimated is:

$$\begin{aligned}\ln L(\beta, \lambda, \sigma^2, \rho) &= -\frac{mT}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| - \frac{1}{2}(\mathbf{y} - X\beta)'\Omega^{-1}(\mathbf{y} - X\beta) \\ &= -\frac{mT}{2}\ln 2\pi - \frac{mT}{2}\ln\sigma^2 - \frac{1}{2}\ln|R| - \frac{1}{2\sigma^2}(\mathbf{y} - X\beta)'R^{-1}(\mathbf{y} - X\beta).\end{aligned}$$

Since the MLEs of σ^2 and β are obtained as

$$\hat{\sigma}^2 = \frac{1}{mT}(\mathbf{y} - X\beta)'R^{-1}(\mathbf{y} - X\beta),$$

and

$$\hat{\beta} = (X'R^{-1}X)^{-1}X'R^{-1}\mathbf{y},$$

respectively, the concentrated likelihood function is:

$$\ln L(\lambda, \rho) = -\frac{mT}{2}\ln\left[\mathbf{y}'(R^{-1} - R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1})\mathbf{y}\right] - \frac{1}{2}\ln|R|.$$

Since the R is an $mT \times mT$ matrix, it is burdensome to calculate its determinant and inverse matrix repeatedly in every step of the maximization procedure. To decrease the computational burden, we use the following transformations for R^{-1}:

$$R^{-1} = \frac{1}{1-\rho}\left(I_{mT} + \frac{\rho}{1-\rho}(\mathbf{1}_T \otimes \Sigma^{1/2})(\mathbf{1}_T \otimes \Sigma^{1/2})'\right)^{-1}$$

$$
\begin{aligned}
&= \frac{1}{1-\rho} I_{mT} - \frac{\rho}{(1-\rho)^2} (\mathbf{1}_T \otimes \Sigma^{1/2}) \left(I_N + \frac{\rho T}{1-\rho} \Sigma \right)^{-1} (\mathbf{1}_T \otimes \Sigma^{1/2})' \\
&= \frac{1}{1-\rho} I_{mT} - \frac{\rho}{(1-\rho)^2} (\mathbf{1}_T \mathbf{1}_T') \otimes \left[\Sigma - \frac{\rho^2}{(1-\rho)^3} \Sigma \left(I_N + \frac{\rho T}{1-\rho} \Sigma \right)^{-1} \Sigma \right].
\end{aligned}
$$

Here, we do not have to take the inverse of an $mT \times mT$ matrix but only the $m \times m$ matrix.

Even in the SCREM, we can discuss which areas have large area-effects and/or which areas have correlations in effects using the conditional mean and variance of $\mathbf{u}$, which are regarded as realizations of the random area-effects. They are obtained as:

$$
E(\mathbf{u}|\mathbf{y}) = \rho(\mathbf{1}_T \otimes \Sigma)' R^{-1} (\mathbf{y} - X\beta),
$$

and

$$
E(\mathbf{u}\mathbf{u}'|\mathbf{y}) = \rho\sigma^2 \left[\Sigma - \rho(\mathbf{1}_T \otimes \Sigma)' R^{-1} (\mathbf{1}_T \otimes \Sigma) \right].
$$

When we substitute the estimates for the parameters, we can obtain their estimates. However, we do not show these values in the empirical studies as the estimation results are not as good as those of the SCFEM.

3 SIMULATIONS AND ESTIMATION

In this section, we will examine whether the methods proposed in the previous section would work well, and we apply them to analyze municipality-based data on housing construction in the TMA. In the simulations, we will compare the estimates of the SCFEM with within estimates and OLS estimates with true clusters being known (OLSTrue). In the OLSTrue model, we calculate the OLS estimates based on the true model, where we know the structure of cluster. In the estimation of the housing-construction function, we apply both the SCFEM and the SCREM to examine the main factors that affect housing construction, how large the unobservable supply-side factors, like regulations on housing developments, are and which areas constitute a cluster.

3.1 SIMULATIONS

In the simulations, we generate the necessary data based on eq.(2) for three years (T=3). We use a 6×6 lattice for the total region to be examined, where there are 36 areas ($m = 36$ in eq.(2)). First, we have to define which areas are neighbors of a particular area. We set as neighbors the left, right, upper and lower adjacent areas of an area. Note that there are just two neighbors for four corner areas in this region and three neighbors for edge areas. We make sequential numbers for the areas in an order so that the i, jth cell of the lattice should be $6 \times (i-1) + j$ (see Fig. 1). We set three clusters, the upper left cluster consisting of nine areas (1,2,3,7,8,9,13,14,15), the upper right cluster consisting of nine areas (4,5,6,10,11,12,16,17,18), and the rest consisting of 18 areas. The cluster-effects are set as $u = (2, 5, 10)$ for the upper left, upper right and the remaining clusters, respectively.

The explanatory variable and the errors, x_{it} and v_{it}, are independently drawn from $N(3, 9)$ and $N(0, 4)$, respectively. The parameter β is set to be two. We conducted 1000-times repetitions.

First, we evaluate how accurately we can find the clusters among the area-effects and how efficiently we can estimate the values of the area-effects, $\mathbf{u}$, with our proposed method. For the first point, we examine the distribution of the selected number of clusters in the simulations. In addition, we consider the expectation regarding how many estimated clusters lie across the true clusters. For the second point, we evaluate the efficiency from the mean squared error for each of the 36 areas, comparing the three estimates, namely the SCFEM, the within- and the OLSTrue estimates.

Table 1 shows descriptive statistics of the distribution of the estimated number of clusters and the number of estimated clusters lying across the true clusters, for both the FS and the BS method. The mean and median of the estimated number of clusters obtained with 1000-times repetitions are 9.459 and 9 respectively for the FS method, and 9.672 and 10 respectively for the BS method. The standard deviations are 1.897 and 1.889 for each method respectively, whereas for both methods about 80 % of the estimates are in a region from seven to 12. With this simulation, we know both of the methods tend to select a larger number of clusters.

Even if the estimated number of the clusters is larger than the true clusters, this will not cause a bias of the estimates as long as the column vectors of the true cluster matrix, D_0 of eq.(3), are expressed as linear combinations of the estimated cluster matrix $\hat{D}$. In other words, if the estimated clusters are derived by dividing the true clusters and the estimated clusters do not lie across the true clusters, the estimates of the area-effects are still unbiased. The expectation regarding the number of estimated clusters that lie across the true clusters is less than 0.5, its median being 0 and the 90 percentile being 1 in the FS method. These results are nearly the same for the BS method. Thus, the probability of the estimated clusters lying across the true clusters is somewhat small for both methods. However, note that, even if they are unbiased, the estimates with $\hat{D}$ are less efficient than the estimates when the true clusters, D_0, are known.

In table 2, we show the simulation means of the area-effects, $\hat{\mathbf{u}}$, and the mean squared errors (MSE) of the SCFEM, the OLSTrue estimates and the within estimates. The estimates of the SCFEM, with the third column relating to the FS method, and the fourth related to the BS method, are almost unbiased. The MSEs of the SCFEM, for both the FS and the BS method, of the OLSTrue estimates and the within estimates are shown in the fifth column to the eighth column. Of course, the MSEs of the OLSTrue estimates are uniformly the smallest. The MSEs of the SCFEM, for both the FS and the BS methods, are smaller than those of the within estimates without one exception (Area 16). Note that the MSE of the SCFEM consists of two parts, that is, the squared bias caused by mis-clustering and the variance of the estimate. The within estimate has a larger variance since the model has larger parameters to be estimated than the other models. This is why the MSEs of the SCFEM estimates are smaller than those of the within estimates, in general. The first part of the MSE is negligibly small, as the results in table 1 show.

Second, we compare the estimates of the β of the SCFEM, for both the FS and the BS methods, with the OLSTrue estimate and the within estimate. Table 3 shows the means, standard deviations and MSEs of these estimates. The means of the estimates are nearly the same and show no remarkable bias. Needless to say, the OLSTrue estimate is the most efficient among the estimates. The estimates of the SCFEM for both methods are more efficient than the within estimates. Their MSEs are also smaller than for the within estimates. Thus, the estimates of the β of the SCFEM are superior to the within estimates, both in terms of efficiency and MSE.

From the results of the simulations, we are able to conclude as follows. First, both the FS and BS methods tend to select a larger model than the true model. However, as the estimated clusters seldom lie across the true clusters, estimates of the cluster-effects are almost unbiased. In addition, they are more efficient than the within estimates. Second, the estimate of the parameter of the other explanatory variable except the cluster-effects is more efficient than the within estimate. Thus, the estimates proposed in this paper are more promising than the within estimates when clusters exist.

3.2 ESTIMATION OF A HOUSING-CONSTRUCTION FUNCTION

In this subsection, we will estimate a housing-construction function and examine the magnitudes of the factors affecting the function. We use data from 1996 to 1998 for the municipalities in the TMA, which encompasses areas that are located within 60-minutes-distance from Tokyo Station. There are 88 areas in the region. Note that Yokohama and Kawasaki cities each have 18 and seven areas, respectively, although these areas are only administrative branches of the cities and hence have no jurisdictional power. We regard them as areas because their populations are as large as an average municipality. In addition, they have distinguishable socio-economic characteristics, which are observable from the data we used, which mark them as independent areas. On the other hand, the Tokyo-23-discricts, which are located in the economic center of Tokyo, surrounding Tokyo Station, have limited but not full jurisdictional power. They can decide on public spending, urban planning and regulations on housing development, but they cannot decide policies relating to local taxes. The other areas examined are municipalities that are regarded as fully jurisdictional areas.

We use a logarithm of housing construction per household in a municipality as a dependent variable.[1] We assume that housing construction in an area is determined by the following fac-

[1] The number of households is the number registered in the beginning of a year, whereas the number of housing constructions is calculated from the notifications of housing construction at the end of a year.

tors. First, the socio-economic characteristics of an area that affect the housing demand of the area are major factors. If the areas are the same with respect to the socio-economic characteristics of their residents, housing demand is determined only by the number of households in the area, so that housing construction per household is constant. In fact, the socio-economic characteristics are very different between the areas. We decompose these socio-economic characteristics into three categories, namely, *economic attributes*, *housing stock* and *amenity*. The *economic attributes* consist of income per household and the time distance to Tokyo Station, the *housing stock*, or the number of houses constructed per household and age of housing stock. In addition, they include the *amenity* of population density, public spending on education per household, the college advancement rate of the high schools, the number of hospitals/clinics per thousand households and the number of parks per household. All of these variables are taken into account in the logarithm.

Second, we consider supply-side effects like the regulations for housing development and spillover effects from public or private infrastructures located near an area that should affect the housing construction within the area. These effects are captured by the cluster-effects in the SCFEM and by the spatial-correlation parameters in the SCREM since they are unobservable or it is difficult to measure their strength.

Table 4 shows descriptive statistics, namely the means, standard deviations and the coefficients of variation (CV) of the dependent and independent variables from 1996 to 1998. The coefficient of variation is a measure for evaluating spatial unevenness among the areas.

The number of houses constructed per household decreased from 1996 to 1998, although the standard deviation increased and thus the absolute value of CV increased. This implies that the decline of the housing boom occurred unevenly among the areas. In addition, household income decreased in 1998, whereas the CV increased. The time distance to Tokyo Station is fixed in 1997 as no new lines were constructed in these years. The *housing stock* is measured by the number of houses per household and by the age of housing stock. The former is a proxy for the abundance of houses with respect to the number of households, while the latter measures the need to rebuild houses. The age of the housing stock is calculated using a weighted average of the time-lengths after construction, with weights being the proportion of houses built in certain construction years. The population density and its CV hardly changed in these years. Public spending for education per household decreased constantly, as did its CV, which implies that public spending evened out among the areas. The college advancement rate of high schools is measured by the proportion of students who attend colleges or universities. This rate also evened out among the areas, as did the number of hospitals per thousand households. However, this pattern did not hold for the number of clinics. Overall, the differences of *amenity* among areas came to be small.

In order to estimate the SCREM, Σ of eq.(6) should be well defined so that the range of λ is restricted according to the values of the smallest and the largest eigenvalues of C. In our case, the smallest eigenvalue of C, namely e_1, is negative, whereas the largest one, namely e_m, is positive. Thus, we have to restrict the parameter λ because $e_1^{-1} < \lambda < e_m^{-1}$ in the numerical maximization procedure of the likelihood function. In the procedure, we reformulate λ as $\lambda = \frac{e_m^{-1}+e_1^{-1}}{2} + \frac{e_m^{-1}-e_1^{-1}}{\pi} \arctan(\gamma)$, $-\infty < \gamma < \infty$, where γ is the substituted parameter for λ.

In the same way, we reformulate ρ with the substitution parameter η as $\rho = \frac{1}{1+\exp(\eta)}$. We evaluated the standard errors of λ and ρ with the delta method as $s.e.(\hat{\lambda}) = \frac{e_m^{-1}-e_1^{-1}}{\pi} \frac{1}{1+\gamma^2}\pi \times s.e.(\hat{\gamma})$ and $s.e.(\hat{\rho}) = \hat{\rho}(1-\hat{\rho}) \times s.e.(\hat{\eta})$.

When we estimate the SCFEM with the BS method, we have to make the dimension of the cluster-dummy matrix less than $m-3$ because three explanatory variables, the time distance to Tokyo, the number of houses per household and the age of the housing stock, are time-invariant variables. We combine four pairs of areas with the estimation results from the within model. We obtain estimates of the fixed-effects using the within model and then we combine the adjacent areas by selecting the four that have the smallest differences of effects between areas. Consequently, we make four clusters, each of which is composed of two adjacent areas.

In table 5, we show the estimation results of the SCFEM, using both the FS and the BS methods, and of the SCREM. In addition, we show the estimation results of the linear regression (LR) model and the within model. First, we compare the estimates of the within model with the LR model. The within estimates are unbiased, although not efficient, in the cases where either the SCFEM or the SCREM are true, whereas the LR estimates are unbiased if the SCREM is true but are biased if the SCFEM is true. The signs of the LR estimates of important explanatory variables, like household income and population density, are the same as the estimates of the within model. Household income has no significant effect in either of the models, while population density has a significantly negative effect on housing construction both in the LR and the within models. The magnitudes of population density are bigger in the within model than in the LR model. Based on both of the model selection criteria, adjusted R^2 and the *aggregate prediction error*, the within model is superior to the LR model. As a result, we discard the LR model.

Second, we compare the remaining three models, beginning with the SCREM. The estimates of household income are significantly positive, whereas those of population density are significantly negative. The estimates and values of the SCFEM are very similar to those of the SCREM in this case. The components of the *housing stock* are insignificant, and public spending for education per household has significant negative effects on housing construction. This result differs from the estimation results of the SCFEM and the within models. Furthermore, the adjusted R^2 of the SCREM is somewhat lower than the other models. Thus, we can conclude that the SCREM is the least preferable model, compared with the other models for this data set.

Third, we compare the within model with the SCFEM. The within model cannot employ the time-invariant explanatory variables, like time distance to Tokyo Station or variables relevant to the *housing stock*, although they are captured in the fixed-effects, as well as in other area-specific effects. It is a defect of the within model that we cannot discuss the effects of these variables. The signs and values of the estimates using the within model are nearly the same as for the SCFEM, except for the value of the population density coefficient. However, unlike the SCFEM estimate, the estimate of household income is not significant. Moreover, the *aggregate prediction error* is bigger for the within model than for the SCFEM. Thus, the SCFEM is preferable to the within model.

Finally, we examine which method of the SCFEM is better using our data set, the FS or the BS method. The signs of the estimates are different for the number of houses per household, the number of hospitals and the number of clinics, as they are negative for the BS method but positive using the FS method. The number of clusters with the BS method is 69, whereas with the FS method it is 50. Note that the estimates become less efficient as the number of clusters becomes larger. In particular, when we want to discuss the structure of the clusters, the efficiency of the estimates is important. The adjusted R^2 with the BS method is larger than for the FS method and the within model. However, using the BS method, the *aggregate prediction error* is bigger than for the FS method, but smaller than for the within model. With these facts, we conclude that the SCFEM with the FS method is the best of these models using our data set.

Now, we examine the effects of the factors influencing housing construction using the estimation results of the SCFEM based on the FS method, which were shown in table 5. In relation to the factors for *economic attributes*, the coefficient of household income was significantly positive, whereas that of distance to Tokyo Station was significantly negative. The number of houses constructed per household was larger in areas that had higher household incomes and were nearer to Tokyo Station.

The estimation results of factors in the *housing stock* are interesting. The signs of the number of houses per household and the age of the housing stock are significantly positive and negative, respectively. These results reflect the difficulties of housing construction in areas that developed in earlier periods. Because a huge number of individual landowners own small portions of land, large-scale development of houses in these areas is costly. Most of the landowners are old, live on pensions and have a deep affection for their town, so that they have no incentive to either re-build or sell their houses.

The coefficients of the variables in *amenity* are significant except for number of hospitals per thousand households. Population density, the college advancement rate of high schools and the number of parks all have negative effects. By contrast, public spending for education per household and the number of clinics per thousand households have positive effects. The estimation results seem to be reasonable, except for the college advancement rate and number of parks. In general, the high schools with high college advancement rates locate in areas that are centers for wider areas. Students with high scores commute from the suburbs to the center. Housing development in this center is restricted or difficult because it was developed in an earlier period. Thus, the college advancement rate is not a good proxy of the educational environment of an area. The coefficient of the number of parks is negative because parks are not in such high demand in the suburbs as in business districts. This is because more natural environments remain in the suburbs, whereas business districts tend to have more parks.

Figures 2 and 3 are Choropleth maps that show the spatial distribution of the estimated values of the cluster-effects and the location of clusters, respectively. Table 6 shows the values of the cluster-effects. The cluster-effects term represents unobservable area-specific factors, beyond the explanatory variables, that affect housing construction in the area. Larger positive values of the effects mean that unobservable factors do enhance housing construction. From figure 2 and table 6, we find that unobservable effects are large in the east, west and north of central Tokyo and in the Tokyo-23-districts. However, they are small in the south.

Our explanatory variables do not incorporate factors that would reduce the construction cost of houses. In general, houses are supplied in the TMA, not in the form of detached houses but in the form of condominiums. Although large developments decrease the cost of housing construction and often attract consumers, it is often subject to restrictive regulations. Since municipalities can impose a new regulation or change the coverage of regulations regarding land use, typically zoning, construction costs are different among the municipalities. These regulations underlie the unobserved factors represented by the area-effects. The estimated area-effects mean that it is easier to undertake large developments of condominiums in the east, west and north of the TMA. Because the south of the TMA was developed earlier than were other areas, it has tended to suffer from over-population and congestion for a long time. The regulations against large development are more severe than in the other areas. The exceptional municipality that has high area-effects in the south is Kamakura, which is the famous old capital and a strongly preferred residence.

In figure 3, we find that the biggest cluster is located in Yokohama city and is composed of eight out of the 18 areas of the city. The next largest cluster lies across Yokohama and Kawasaki cities, which are adjacent each other. Two clusters composed of five areas are located in the center of the Tokyo-23-districts and of boundary areas in the west of the Tokyo-23-districts. We cannot say exactly what factors assist in creating clusters from these results, but the larger clusters tend to locate in the Tokyo-23-districts and the cities of Yokohama and Kawasaki. This is consistent with our intuition, as there are well-developed public transportation networks in the Tokyo-23-districts and the areas of Yokohama and Kawasaki have no independent jurisdictional powers. Smaller clusters tend to be formed along the main railways.

4 CONCLUSION

In this paper, we have proposed two types of econometric models, a *spatially clustered fixed-effects model* (SCFEM) and a *spatially correlated random-effects model* (SCREM), to examine area-based panel data. We have observed that housing construction is spatially uneven in the Tokyo Metropolitan Area (TMA). We investigate what factors affect housing construction, incorporating unobservable factors, such as local differences in regulations governing housing developments and spillovers of local public or private goods, which may cause spatial clustering or correlation of housing construction.

The SCFEM is a type of fixed-effects model where a cluster has the same effects, so that we have to find which areas constitute a cluster. The issue of finding clusters can be regarded as a problem of model selection from too many possible models. We adopt an *aggregate prediction error* as a model selection criterion, which is estimated by a resampling method, namely *leave-one-out cross-validation*. Forward- and backward-stepwise methods are employed for the searching procedure. We show by simulations that these methods work well and the estimated parameters of concern are more efficient than the within estimates. The SCREM is a model where the random-effects are spatially correlated. We use the concentrated maximum likelihood method for the estimation.

We estimate housing-construction functions and find that the SCFEM using the forward-stepwise method is the best model since its *aggregate prediction error* is the smallest among the models and the signs and magnitudes of the estimates are rationally accountable. The unobservable area-effects are large in the east, west and north areas of the TMA but small in the south. This may result from the fact that the south was developed earlier than the other areas, so that regulations against housing development are more severe. Clusters are found in huge cities like Yokohama, Kawasaki and the Tokyo-23-districts. This is partly because Yokohama and Kawasaki have no independent jurisdictional powers and partly because there are spillover effects from the well-developed public transportation networks in the Tokyo-23-districts. Smaller clusters tend to be formed along the main railways.

Appendix:

In this appendix, we provide an example of how to construct the cluster-dummy matrix, $D(k)$, and the corresponding adjacent matrix, $A(k)$ with dimension k. Let $A(m)$ be the matrix of the largest model, which means it is the matrix representing adjacency of the municipalities of the region. In this example, there are seven municipalities and the adjacency is represented by $A(7)$ as:

$$A(7) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 1 & 0 \\ & & 1 & 0 & 0 & 1 & 1 \\ & & & 1 & 1 & 1 & 0 \\ & & & & 1 & 1 & 0 \\ & & & & & 1 & 1 \\ & & & & & & 1 \end{pmatrix},$$

where the lower-left elements are omitted for simplicity since $A(m)$ is symmetric. Starting with this state, assume that we can combine the second and sixth area into a cluster. We impose a rule that the column vector with the larger column number is merged into the smaller-number column vector when combining areas/clusters. Then, the cluster-dummy matrix is obtained as:

$$D(6) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The corresponding adjacent matrix is constructed as follows. Let $a(k)_j$ be the jth column vector of $A(k)$. Define a logical operator of dummy vectors with elements being 1 or 0, $\vee$, that creates a vector which takes a disjunction of two vectors. For example, when $x = \{0\ 1\ 0\ 1\}$ and $y = \{1\ 1\ 0\ 0\}$, then $x \vee y$ gives $\{1\ 1\ 0\ 1\}$. With this operator, replace the second column and row vectors of $A(7)$ as follows:

$$a(7)_{C2} \Leftarrow a(7)_{C2} \vee a(7)_{C6},$$

and

$$a(7)_{R2} \Leftarrow a(7)_{R2} \quad \vee \quad a(7)_{R6},$$

where subscripts represent the numbers of the columns or the rows, respectively. Then, the newly created adjacent matrix is obtained as:

$$A(7) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ & 1 & 1 & 1 & 1 & 1 & 1 \\ & & 1 & 0 & 0 & 1 & 1 \\ & & & 1 & 1 & 1 & 0 \\ & & & & 1 & 1 & 0 \\ & & & & & 1 & 1 \\ & & & & & & 1 \end{pmatrix}.$$

The adjacency of the created cluster, which is composed of what were previously the second and sixth municipalities, is represented by the second column of the above matrix. The sixth column is now redundant. Hence, the adjacent matrix of dimension six is obtained by deleting both the sixth column and row vectors as follows:

$$A(6) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ & 1 & 1 & 1 & 1 & 1 \\ & & 1 & 0 & 0 & 1 \\ & & & 1 & 1 & 0 \\ & & & & 1 & 0 \\ & & & & & 1 \end{pmatrix}.$$

Acknowledgements

This research was supported by a grant-in-aid from the Zengin Foundation for Studies on Economics and Finance and by MEXT.KAKENHI 15200021.

References

[1] Aaronson, D., 2001, Neighborhood dynamics, Journal of Urban Economics 49, 1-31.

[2] Anselin, Luc, 1988, Spatial econometrics: methods and models (Kluwer Academic Publishers, Dordrecht).

[3] Cressie, N. A. C., 1993, Statistics for spatial data (John Wiley and Sons, New York).

[4] Davison, A. C., and D. V. Hinkley, 1997, Bootstrap methods and their application (Cambridge University Press, Cambridge).

[5] DiPasquale, D. and W. C. Wheaton, 1994, Housing market dynamics and the future of housing prices, Journal of Urban Economics 35, 1-27.

[6] Gramlich, E. M. and D. L. Rubinfeld, 1982, Microestimates of public spending demand functions and test of the Tiebout and median-voter hypothesis, Journal of Political Economy 90, 536-560.

[7] Geisser, S. 1975, The predictive sample reuse method with application, Journal of the American Statistical Association 70, 320-328.

[8] Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).

[9] Krugman, P. R., 1993, First nature, second nature, and metropolitan location, Journal of Regional Science 33, 293-298.

[10] Krugman, P. R., 1996, The self-organizing economy (Blackwell, Cambridge, MA).

[11] Lütkepohl, H., 1991, Introduction to multiple time series analysis, 2nd edition (Springer-Verlag, Berlin).

[12] Mayer, C. J. and C. T. Somerville, 2000, Residential construction: using the urban growth model to estimate housing supply, Journal of Urban Economics 48, 85-109.

[13] Netz, J. S. and B. A. Taylor, 2002, Maximum or minimum differentiation? Location patterns of retail outlets, The Review of Economics and Statistics 84, 162-175.

[14] Pinkse, J. and M. E. Slade, 1998, Contracting in space: an application of spatial statistics to discrete-choice models, Journal of Econometrics 85, 125-154.

[15] Stone, M., 1974, Cross-validatory choice and assessment of statistical predictions, Journal of the Royal Statistical Society Ser. B, 36, 111-147.

[16] Stone, M., 1977, An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, Journal of the Royal Statistical Society Ser. B, 39, 44-47.

[17] Tiebout, C. 1956, A pure theory of local expenditure, *Journal of Political Economy* 64, 718-740.

[18] Topel, R. and S. Rosen, 1988, Housing investment in the United States, Journal of Political Economy 96, 718-740.

Table 1: Descriptive Statistics of Selecting Clusters

	Forward-stepwise method		Backward-stepwise method	
	No. of Clusters	No. of Estimated Clusters Lying across True Clusters	No. of Clusters	No. of Estimated Clusters Lying across True Clusters
Mean	9.459	0.479	9.672	0.485
s.d.	1.897	0.593	1.889	0.583
5 percentile	6	0	7	0
10 percentile	7	0	7	0
Median	9	0	10	0
90 percentile	12	1	12	1
95 percentile	13	1	13	1

Table 2: Estimates of the Area-Effects and Mean Squared Errors

	Area	Mean of SCFEM Estimates: FS	Mean of SCFEM Estimates: BS	MSE of SCFEM Estimates: FS	MSE of SCFEM Estimates: BS	MSE of OLS Estimates with True Model	MSE of Within Estimates
UPPER-LEFT CLUSTER	1	1.968	1.963	1.157	1.204	0.158	1.612
	2	2.030	2.034	0.971	1.023	0.158	1.335
	3	2.099	2.100	1.279	1.309	0.158	1.406
	7	1.921	1.928	1.156	1.167	0.158	1.529
	8	2.009	2.000	0.993	0.970	0.158	1.318
	9	2.106	2.111	1.333	1.276	0.158	1.405
	13	1.983	1.974	1.089	1.066	0.158	1.393
	14	1.973	1.980	1.027	1.051	0.158	1.352
	15	2.055	2.053	1.149	1.128	0.158	1.316
UPPER-RIGHT CLUSTER	4	4.873	4.871	1.417	1.489	0.185	1.569
	5	4.916	4.903	1.070	1.160	0.185	1.438
	6	5.038	5.048	0.967	1.057	0.185	1.379
	10	4.874	4.896	1.276	1.285	0.185	1.393
	11	4.945	4.955	0.961	0.965	0.185	1.250
	12	5.011	4.994	0.902	0.954	0.185	1.324
	16	4.893	4.879	1.711	1.609	0.185	1.572
	17	5.005	4.991	1.191	1.064	0.185	1.348
	18	5.011	4.988	1.232	1.142	0.185	1.369
LOWER CLUSTER	19	10.038	10.033	1.046	1.158	0.112	1.450
	20	10.052	10.068	0.930	1.059	0.112	1.460
	21	10.057	10.043	0.837	0.996	0.112	1.373
	22	9.892	9.912	1.095	1.198	0.112	1.503
	23	9.964	9.957	0.918	1.140	0.112	1.389
	24	10.010	10.019	0.914	1.036	0.112	1.313
	25	10.050	10.075	0.978	1.117	0.112	1.548
	26	10.022	10.020	0.922	1.126	0.112	1.476
	27	9.924	9.914	0.789	0.992	0.112	1.318
	28	9.998	10.000	0.847	1.012	0.112	1.373
	29	9.918	9.930	0.812	1.069	0.112	1.436
	30	9.947	9.965	0.793	0.972	0.112	1.278
	31	10.057	10.077	0.954	1.113	0.112	1.395
	32	9.956	9.949	0.810	1.052	0.112	1.374
	33	9.972	9.961	0.874	1.050	0.112	1.354
	34	10.015	9.992	0.964	1.120	0.112	1.509
	35	9.941	9.963	0.905	1.093	0.112	1.438
	36	9.989	9.967	0.778	0.962	0.112	1.237

Notes: SCFEM, FS and BS are abbreviations of spatially clustered fixed-effects model, forward-stepwise and backward-stepwise, respectively.

Table 3: Comparison of the Estimates

	SCFEM estimate with forward-stepwise method	SCFEM estimate with backward-stepwise method	OLS Estimates with True Model Known	Within estimate
Mean	2.0016	2.0016	1.9995	1.9973
S.D.	0.0736	0.0754	0.0633	0.0783
MSE	0.0054	0.0057	0.0040	0.0061

Note: S.D. is an an abbreviation of standard deviation.

Table 4: Descriptive Statistics of the Data

	Mean			S.D.			Coefficient of Variation		
	1998	1997	1996	1998	1997	1996	1998	1997	1996
Dependent Variable									
No. of Housing Construction per Household (log)	-3.591	-3.494	-3.395	0.310	0.273	0.252	-0.086	-0.078	-0.074
Independent Variable									
Economic Attributes									
Household Income (log)	1.595	1.610	1.604	0.121	0.116	0.115	0.076	0.072	0.072
Time Distance to Tokyo Station (log)		5.212			3.041			0.583	
Housing Stock									
No. of Houses per Household (log)		0.029			0.059			2.024	
Age of Housing Stock (log)		-0.368			0.082			-0.224	
Amenity									
Population Density (log)	9.002	8.996	8.991	0.488	0.491	0.494	0.054	0.055	0.055
Public Spending for Education per Household (log)	-1.510	-1.457	-1.412	1.240	1.305	1.328	-0.821	-0.896	-0.941
College Advancement Rate of the High Schools (log)	3.750	3.678	3.613	0.383	0.435	0.473	0.102	0.118	0.131
No. of Hospitals per 1000 household (log)	-2.328	-2.319	-2.261	0.542	0.574	0.561	-0.233	-0.247	-0.248
No. of Clinics per 1000 household (log)	0.578	0.586	0.583	0.479	0.482	0.489	0.827	0.823	0.840
No. of Parks per household (log)		0.029			0.059			2.024	

Table 5: Estimates of SCFE and SCRE Models

	Spatially Clustered Fixed-Effects Model		Spatially Correlated Random-Effects	Linear Regression Model	Within Model
	FS	**BS**			
constant			-2.088	-2.580	
			1.086	*0.691*	
Economic Attributes					
Household Income (log)	1.008	0.904	0.586	0.181	0.335
	0.137	*0.176*	*0.317*	*0.209*	*0.745*
Time Distance to Tokyo Station (log)	-0.008	-0.196	-0.001	0.001	
	0.003	*0.013*	*0.008*	*0.005*	
Housing Stock					
No. of Houses per Household (log)	0.564	-0.927	0.687	0.600	
	0.255	*0.291*	*0.522*	*0.335*	
Age of Housing Stock (log)	-0.343	-2.615	-0.227	-0.435	
	0.186	*0.284*	*0.438*	*0.281*	
Amenity					
Population Density (log)	-0.245	-0.950	-0.220	-0.193	-2.909
	0.040	*0.064*	*0.087*	*0.052*	*0.888*
Public Spending for Education per Household (log)	0.208	0.274	-0.136	-0.071	0.263
	0.025	*0.054*	*0.044*	*0.029*	*0.077*
College Advancement Rate of the High Schools (log)	-0.175	-0.128	-0.038	0.089	-0.135
	0.029	*0.046*	*0.063*	*0.045*	*0.090*
No. of Hospitals per 1000 household (log)	0.047	-0.065	0.085	0.019	-0.117
	0.021	*0.032*	*0.053*	*0.037*	*0.092*
No. of Clinics per 1000 household (log)	0.184	-0.451	0.145	0.250	-0.120
	0.038	*0.061*	*0.089*	*0.051*	*0.369*
No. of Parks per household (log)	-0.877	-0.896	-0.190	-0.076	-0.895
	0.038	*0.039*	*0.033*	*0.023*	*0.049*
Spatial error term					
λ			0.168		
			61.432		
ρ			0.577		
			1.524		
σ^2			0.067		
No. of Clusters	50	69		1	88
Log Likelihood			-1.293		
Adjusted R^2	0.842	0.873	-0.015	0.249	0.869
Aggregate Prediction Error	0.012	0.015		0.067	0.017

Note: The italics are standard errors of the estiamtes. The standard error of σ^2 is not calculated because it is not so important nor easy for calculation. FS and BS are abbreviations of forward-stepwise and backward-stepwise, respectively.

Table 6: Structure of Clusters and Effects of Unobservable Factors

Area	Estimates	s.e.	Cluster no.	no. of compo-nents	Area	Estimates	s.e.	Cluster no.	no. of compo-nents	Area	Estimates	s.e.	Cluster no.	no. of compo-nents
1	1.323	0.490	34	1	**31**	0.704	0.529	19	5	**61**	-0.259	0.514	9	1
2	1.380	0.520	35	2	**32**	1.517	0.546	40	1	**62**	0.599	0.522	16	1
3	1.380	0.520	35	2	**33**	1.132	0.522	28	5	**63**	0.392	0.515	12	2
4	1.227	0.512	31	3	**34**	1.531	0.517	41	3	**64**	-2.514	0.501	2	3
5	1.227	0.512	31	3	**35**	0.860	0.526	24	4	**65**	-2.193	0.504	5	1
6	1.227	0.512	31	3	**36**	0.294	0.534	11	1	**66**	-2.467	0.505	3	8
7	0.858	0.498	23	3	**37**	0.860	0.526	24	4	**67**	-2.467	0.505	3	8
8	0.858	0.498	23	3	**38**	0.860	0.526	24	4	**68**	-2.467	0.505	3	8
9	1.235	0.523	32	1	**39**	0.704	0.529	19	5	**69**	-2.351	0.519	4	1
10	2.131	0.511	49	1	**40**	0.416	0.506	14	2	**70**	-2.762	0.503	1	2
11	1.206	0.508	30	1	**41**	0.704	0.529	19	5	**71**	-2.762	0.503	1	2
12	0.729	0.493	20	1	**42**	0.416	0.506	14	2	**72**	-2.008	0.501	6	7
13	0.588	0.517	15	1	**43**	0.979	0.506	25	1	**73**	-2.467	0.505	3	8
14	1.471	0.508	38	2	**44**	0.406	0.512	13	1	**74**	-2.467	0.505	3	8
15	0.812	0.494	22	1	**45**	1.471	0.508	38	2	**75**	-2.467	0.505	3	8
16	0.688	0.497	18	1	**46**	1.028	0.532	27	1	**76**	-2.008	0.501	6	7
17	0.673	0.513	17	1	**47**	1.720	0.521	45	1	**77**	-2.467	0.505	3	8
18	0.858	0.498	23	3	**48**	1.132	0.522	28	5	**78**	-2.467	0.505	3	8
19	1.132	0.522	28	5	**49**	1.531	0.517	41	3	**79**	-2.008	0.501	6	7
20	0.994	0.512	26	1	**50**	1.744	0.513	46	1	**80**	-1.769	0.495	7	2
21	1.531	0.517	41	3	**51**	1.237	0.534	33	1	**81**	-2.514	0.501	2	3
22	1.684	0.529	44	1	**52**	0.860	0.526	24	4	**82**	-2.514	0.501	2	3
23	1.139	0.518	29	1	**53**	2.204	0.533	50	1	**83**	-2.008	0.501	6	7
24	1.562	0.505	42	1	**54**	1.509	0.521	39	1	**84**	-1.769	0.495	7	2
25	0.787	0.502	21	1	**55**	1.997	0.534	48	1	**85**	-2.008	0.501	6	7
26	1.132	0.522	28	5	**56**	1.387	0.518	36	1	**86**	-2.008	0.501	6	7
27	0.242	0.565	10	1	**57**	1.627	0.516	43	1	**87**	-2.008	0.501	6	7
28	1.132	0.522	28	5	**58**	1.795	0.517	47	1	**88**	1.401	0.519	37	1
29	0.704	0.529	19	5	**59**	-0.332	0.512	8	1					
30	0.704	0.529	19	5	**60**	0.392	0.515	12	2					

Note: The s.e. is the abbreviation of standard error. All areas/clusters are numbered with the cluster number.

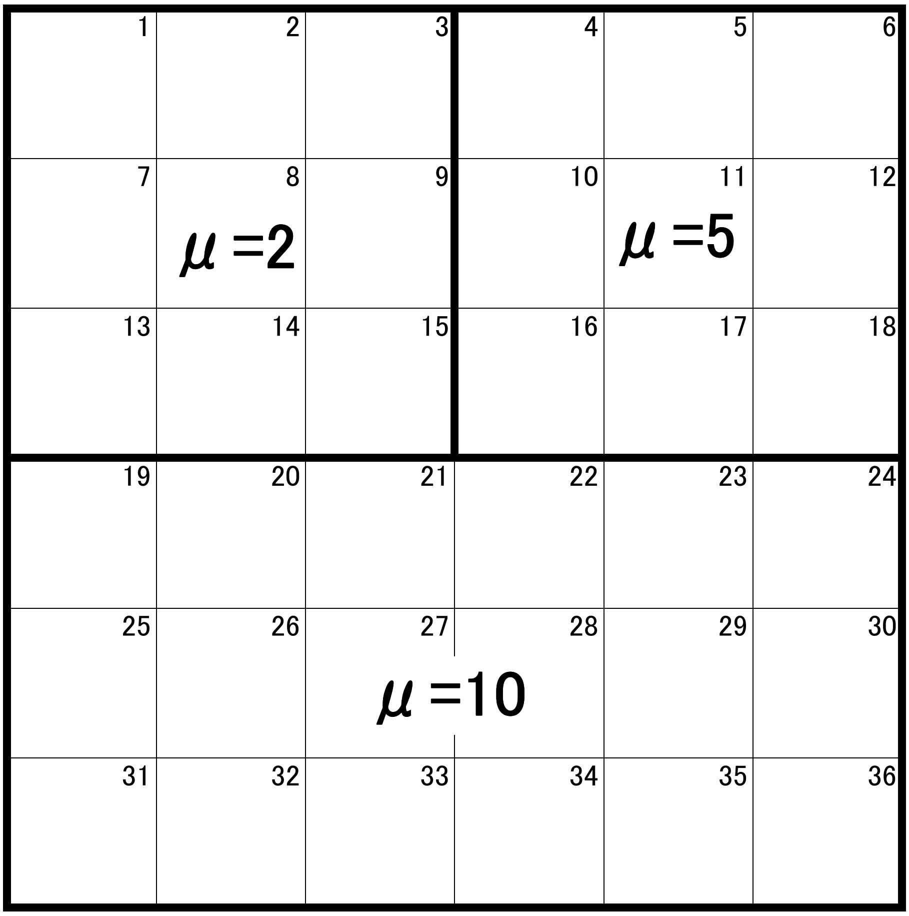

Fig.1 True Structure of Clusters

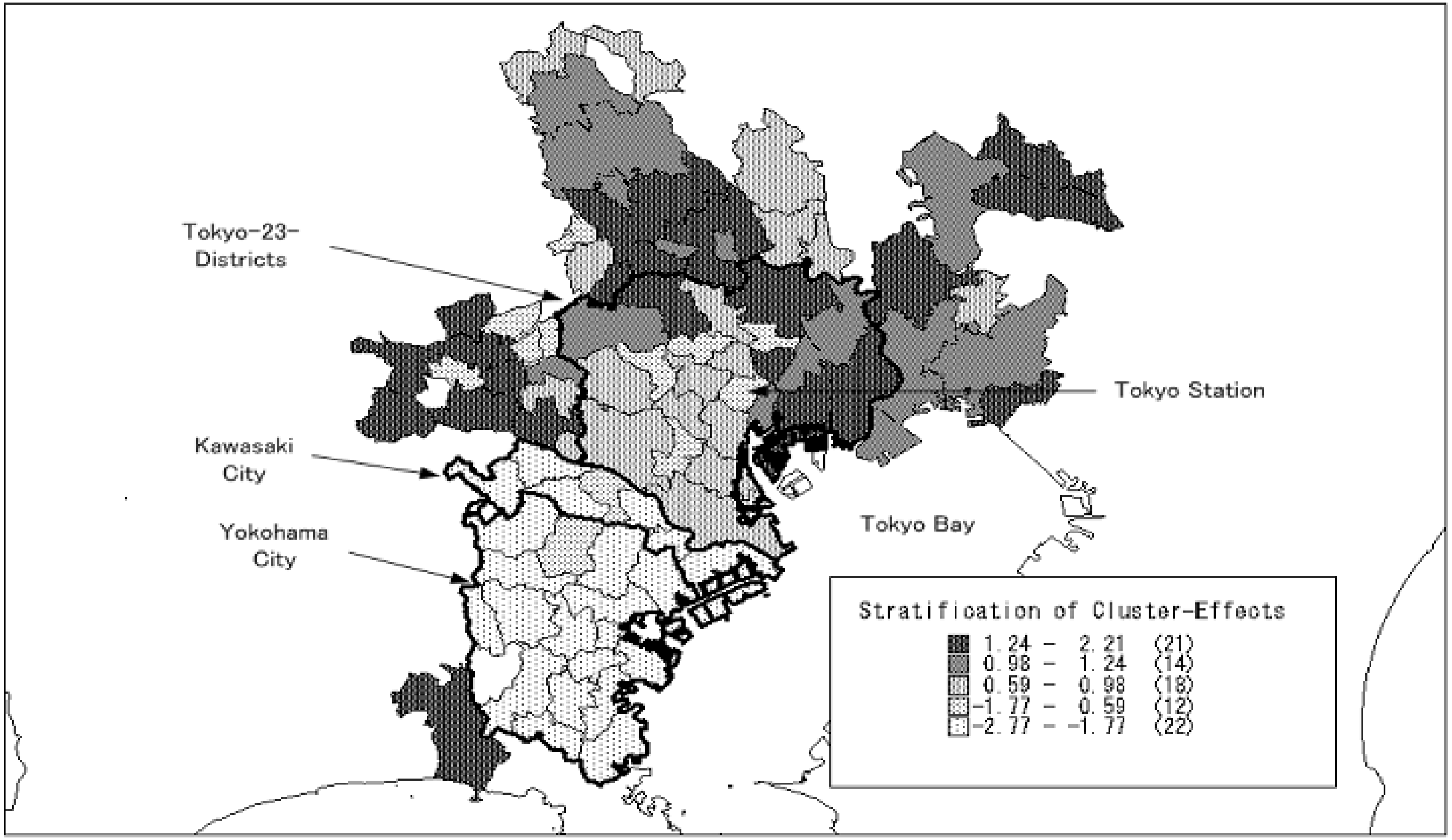

Fig. 2 Spatial Distribution of the Estimated Cluster-Effects

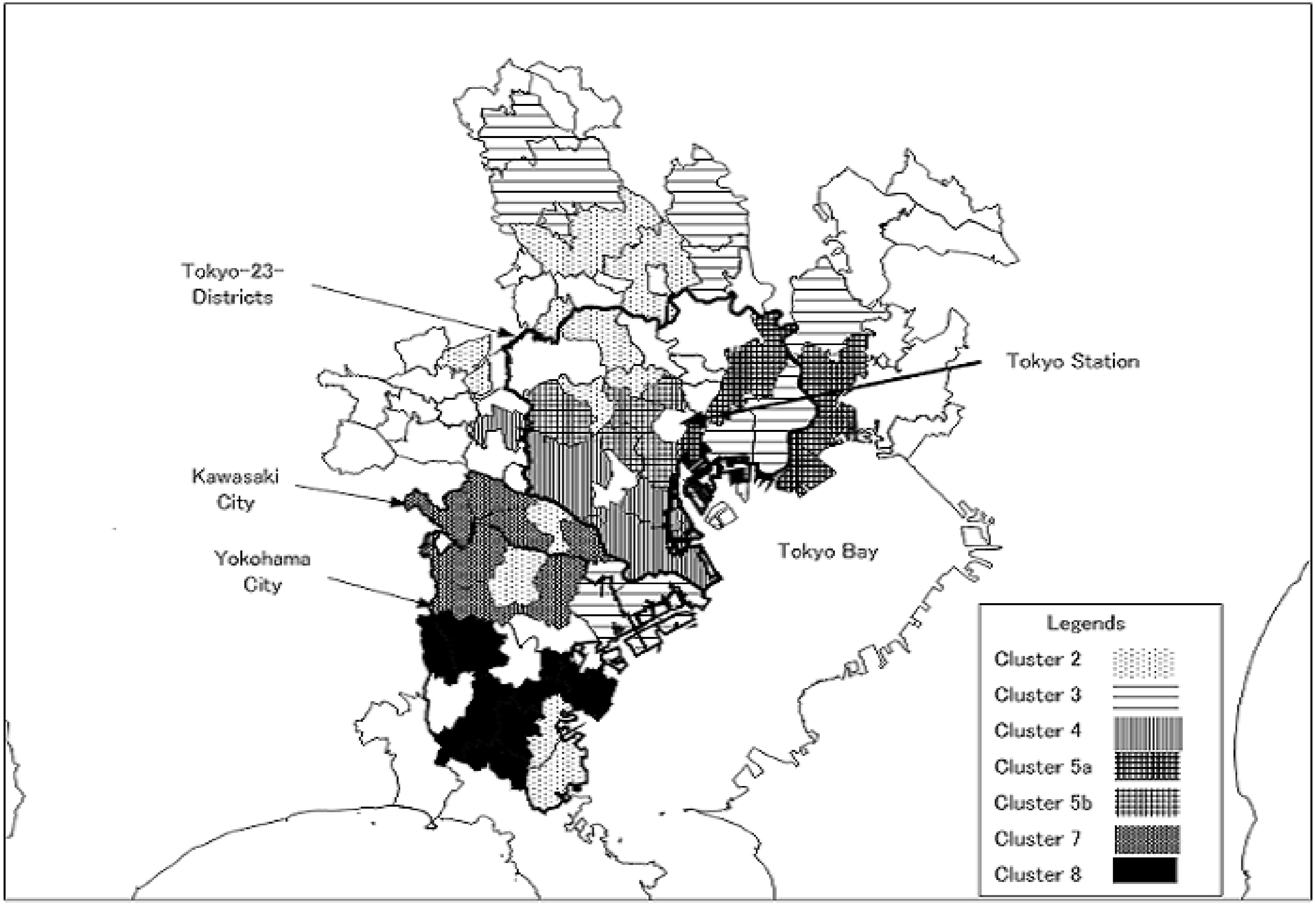

Fig. 3 Spatial Distribution of Clusters