Full information maximum likelihood estimation of quantitative spatial economics models

Shunsuke Segi*

1Graduate School of Engineering, Kobe University, Kobe, Kobe, Japan

Correspondence
*Shunsuke Segi, Graduate School of Engineering, Kobe University, Rokkodai-cho 1, Nada-ku, Kobe, 657-0013, Japan. Email: segi@people.kobe-u.ac.jp

Abstract
This study proposes a novel method to estimate the parameters in quantitative spatial economics models with full information maximum likelihood. The method can be applied to models with a one-to-one mapping from the observed data on the endogenous variables to the exogenous structural residuals. Under the assumption that the structural residuals follow a probability distribution, the log-likelihood function can be constructed. The method can be applied to both cross-sectional data and panel data. In many cases, the method would restrict the estimated parameters to the domain in which the equilibrium is stable. This study confirms the validity of the proposed method by applying it to two kinds of data: experimental data generated from a model with known parameters; and observed data from Japan during 1975–2015.

KEYWORDS:
quantitative spatial economics, economic geography, spatial econometrics, full information maximum likelihood estimation, agglomeration, dispersion, equilibrium stability

JEL CLASSIFICATION:
C13; C31; C33; C35; C51; C62; R12; R23

1 | INTRODUCTION

The application of quantitative spatial economics models to empirical research has gained much attention in recent times. In this line of research, researchers develop quantitative models that describe the distribution of population and economic activities across locations. The distinctive feature of the models is that they can express the observed data of the endogenous variables as an equilibrium outcome. The trick is that each location in the model has exogenous variables to indicate the intrinsic advantage of the location (e.g., amenities and productivity), which are called structural residuals or structural fundamentals. The observed data are exactly rationalized as an equilibrium by calibrating the structural residuals appropriately. The quantitative spatial economics models can be used to analyze how the agglomeration forces and the dispersion forces in the economy have contributed to past changes in the spatial distribution of economic activities. These models can also be used to perform counterfactual exercises to predict the impacts of policies or infrastructure development. This approach has a great advantage over the other approaches for spatial analysis (e.g., difference-in-differences approach) in that it can take account of the spatial interactions between locations. The approach of quantitative spatial economics has achieved fruitful results in spatial models to study the distribution of economic activities across regions within a country as well as urban models to study the internal structure of economic activities within a metropolitan area.
Although the approach of quantitative spatial economics is very appealing, its adoption produces a technically difficult problem. The quantitative models have parameters to determine the size of agglomeration forces and dispersion forces. It is necessary to estimate these parameters appropriately because the result of a counterfactual analysis depends on the values of these parameters. Several studies have overcome this problem. Most of them have used panel data and exogenous instrumental variables to estimate parameters. Redding and Sturm\textsuperscript{14} used the division of Germany as an exogenous shock. They searched for the parameter values so that the relative decline of the East–West German border cities after the division matches the observed data. Rossi-Hansberg et al.\textsuperscript{16} used urban revitalization programs implemented in Richmond, Virginia as an exogenous shock. They searched for the value of the distance decay parameter so that the increase in land rent around the impacted areas of the programs matches the observed data. Ahlfeldt et al.\textsuperscript{1} applied generalized method of moments (GMM) estimation to panel data on city blocks in Berlin and assumed that the differences in the structural residuals were uncorrelated with the exogenous changes induced by Berlin’s division and reunification. Diamond\textsuperscript{7} adopted a similar approach using exogenous local productivity changes as instrumental variables.

This study proposes a novel method to estimate the parameters in a quantitative spatial economics model with full information maximum likelihood (FIML). The method can be applied to models with a one-to-one mapping from the observed data on the endogenous variables to the exogenous structural residuals, which is satisfied in general quantitative spatial economics models\textsuperscript{13}. Under the assumption that the structural residuals follow a probability distribution, the log-likelihood function can be constructed. The method can be applied to both cross-sectional data and panel data, which is advantageous when the available data is limited. The method has another advantage. Although it is possible to identify structural residuals that rationalize the observed data for any values of the model parameters, the stability of the equilibrium is not guaranteed. Therefore, researchers should be careful about whether the equilibrium is stable under the estimated parameters. Allen and Arkolakis\textsuperscript{2} solved this problem by analytically deriving the domain of the parameter space in which the equilibrium is unique and stable, and then restricted the parameters to the domain. However, it is not always possible to derive such a domain. In addition, this approach excludes the possibility that multiple equilibria exist. However, proposed estimation method in this study would restrict the estimated parameters to the domain of a stable equilibrium in many cases. It works even if the model has multiple equilibria.

The remainder of the paper is structured as follows. Section 2 formulates an illustrative quantitative spatial model to be estimated by the proposed method. Section 3 presents the estimation method with FIML and discusses its properties. Section 4 confirms the validity of the estimation method by applying it to the spatial model of Section 2 with two kinds of data: experimental data generated from a model with known parameters; and observed data from Japan during 1975–2015. Section 5 concludes.

2 QUANTITATIVE SPATIAL MODEL

In this section, I formulate an illustrative quantitative spatial model to be estimated by the method outlined in Section 3. The model is a variant of the canonical spatial model introduced in Redding and Rossi-Hansberg\textsuperscript{13}. The model in this section is useful for illustrating the role of structural residuals, which are a key concept not only in the quantitative spatial economics literature but also in the estimation method proposed in this study.

I consider an economy consisting of \(N_R\) regions indexed by \(r\). There is a unit measure of households in the economy. Each household has a unit measure of labor that is supplied inelastically to a firm located in the region where the household resides. Households are perfectly geographically mobile. Therefore, in the long run, the utility of residing in a region is equalized across all regions.

2.1 Household consumption

All the households in the model are homogeneous and have an identical preference. The preference is defined over “land-intensive goods” \((A_r)\), “residential land use” \((H_r)\), and “labor-intensive goods” \((M_r)\) and is assumed to take the following Cobb–Douglas utility function.

\[
v_r = \alpha \ln \frac{A_r}{\alpha} + \beta \ln \frac{H_r}{\beta} + \mu \ln \frac{M_r}{\mu} \quad (\alpha + \beta + \mu = 1)
\]

Here, \(v_r\) is the utility of a household in region \(r\). \(\alpha\), \(\beta\), and \(\mu\) are share parameters that sum up to one. \(A_r\), \(H_r\), and \(M_r\) are per capita consumption of the land-intensive goods, residential land use, and labor-intensive goods, respectively, in region \(r\).
Land-intensive goods are those whose production input is only land. I call them A-goods hereafter. The trade of A-goods follows the Armington trade model in which the goods are differentiated by region of origin. The consumption index of A-goods denoted by $A_r$ is defined over consumption of the goods from different regions as follows.

$$A_r = \left[ \sum_{r=1}^{N_r} \frac{a_{r'r}}{a_{rr'}} \right]^{\frac{1}{\sigma_A}}$$

Here, $a_{r'r}$ denotes the consumption of the A-goods produced in region $r'$. $\sigma_A$ is the elasticity of substitution between the A-goods produced in different regions.

Labor-intensive goods are those whose production input is only labor. I call them M-goods hereafter. The households have a “love-of-variety” preference over M-goods. The consumption index of M-goods denoted by $M_r$ is defined over consumption of the differentiated varieties of the goods as follows.

$$M_r = \left[ \sum_{r=1}^{N_r} n_r^M \frac{m_{r'r}}{m_{rr'}} \right]^{\frac{1}{\sigma_M}}$$

Here, $n_r^M$ denotes the number of varieties of M-goods produced in region $r$. $m_{r'r}$ denotes the consumption of a variety of M-goods produced in region $r'$. Because all the varieties produced in a region are symmetrical, I omit the index to denote a variety. $\sigma_M$ is the elasticity of substitution among differentiated varieties of M-goods.

Household income is derived from wage and land rent as follows.

$$I_r = w_r + \omega^A + \omega^H$$

Here, $I_r$ denotes the income per capita in region $r$. $w_r$ denotes the wage rate in region $r$. $\omega^A$ and $\omega^H$ denote the two types of income from land rent, which is formulated in Subsections 2.2 and 2.3.

The demand function of a household is derived as follows from the maximization problem of the utility function (1)–(3) subject to the household’s budget constraint.

$$A_r = \frac{\alpha I_r}{P_r^A}$$

$$P_r^A = \left[ \sum_{r'=1}^{N_r} \{ p_{r'}^A (1 + \delta_{r'r})^{\phi_A} \}^{1-\sigma_A} \right]$$

$$a_{r'r} = A_r \left[ \frac{P_r^A}{p_{r'}^A(1 + \delta_{r'r})^{\phi_A}} \right]^{\sigma_A}$$

$$M_r = \frac{\mu I_r}{P_r^M}$$

$$P_r^M = \left[ \sum_{r'=1}^{N_r} n_r^M \{ p_{r'}^M (1 + \delta_{r'r})^{\phi_M} \}^{1-\sigma_M} \right]$$

$$m_{r'r} = M_r \left[ \frac{P_r^M}{p_{r'}^M(1 + \delta_{r'r})^{\phi_M}} \right]^{\sigma_M}$$

$$H_r = \frac{\beta I_r}{P_r^H}$$

Here, $P_r^A$ and $P_r^M$ denote the price index of A-goods and M-goods, respectively. They are derived as shown in equations (6) and (9), respectively. $p_{r'}^A$ and $p_{r'}^M$ denote the free-on-board (FOB) price of A-goods and M-goods in region $r'$, respectively. To trade goods between different regions, they incur an iceberg trade cost. The trade cost depends on the travel time between regions. Let $\delta_{r'r}$ denote the exogenous travel time between regions $r$ and $r'$. One plus ad-valorem trade cost between regions $r'$ and $r$ is expressed as $(1 + \delta_{r'r})^{\phi_A}$ for A-goods and $(1 + \delta_{r'r})^{\phi_M}$ for M-goods, where $\phi_A$ and $\phi_M$ are parameters. The iceberg trade cost is assumed to be symmetrical, and thus, $\delta_{r'r} = \delta_{r'r'}$. In addition, the trade cost for intra-regional trade is assumed to be zero, and
thus, \( \delta_{rr} = 0 \). \( P_r^H \) denotes the land rent in region \( r \). As in Redding and Rossi-Hansberg\(^{13}\), the unit of residential land is quality adjusted. Therefore, \( H_r \) and \( P_r^H \) can be different from the observed area and rent of residential land.

### 2.2 Market of residential land

Residential land supplied to households in region \( r \) is assumed to be fixed and is expressed as \( \bar{S}_r e^{\varphi_1} \). Here, \( \bar{S}_r \) denotes the observable exogenous area of available land (not limited to residential land) in region \( r \). \( e^{\varphi_1} \) is the ratio of quality-adjusted residential land area to the available land area in region \( r \). \( \lambda_r \) is the exogenous amenities in region \( r \), which affect the supply of quality-adjusted residential land. Let \( \lambda_r \) denote the population of households in region \( r \). Then, the market-clearing condition of residential land in region \( r \) is written as follows.

\[
\lambda_r H_r = \bar{S}_r e^{\varphi_1} \quad (12)
\]

The ownership of the residential land in the economy is assumed to be equally shared by the households in the economy. The income of a household from residential land rent, denoted by \( \omega^H \), is expressed as follows.

\[
\omega^H = \sum_{r=1}^{N_R} \frac{P_r^H \bar{S}_r e^{\varphi_1}}{\sum_{r=1}^{N_R} \lambda_r} \quad (13)
\]

Note that the denominator of the right-hand side is unity.

### 2.3 Market of A-goods

A-goods are produced under conditions of perfect competition and constant returns to scale. A unit of land is necessary to produce a unit of A-goods. The land supplied for the production of A-goods in region \( r \) is assumed to be fixed and is expressed as \( \bar{a} \bar{S}_r \). Here, \( \bar{a} \) is a constant to denote the ratio of land area supplied for the production of A-goods to the available land area in region \( r \). The market-clearing condition of A-goods produced in region \( r \) is written as follows.

\[
\sum_{r=1}^{N_R} (1 + \delta_{rr})^\varphi \lambda_r a_{rr} = \bar{a} \bar{S}_r \quad (14)
\]

The ownership of the land supplied for the production of A-goods is assumed to be equally shared by the households in the economy. The income of a household from the rent of the land, denoted by \( \omega^A \), is expressed as follows.

\[
\omega^A = \sum_{r=1}^{N_R} \frac{P_r^A \bar{a} \bar{S}_r}{\sum_{r=1}^{N_R} \lambda_r} \quad (15)
\]

### 2.4 Markets of M-goods and labor

Each variety of M-goods is produced by a firm under conditions of monopolistic competition and increasing returns to scale, just as the manufactured goods in Krugman’s core–periphery model\(^{8}\). To produce a variety, a firm in region \( r \) must incur a fixed cost of \( F \) units of labor and a constant variable cost of \( e^{w^W} \) units of labor. \( w^W \) is the exogenous productivity in region \( r \), which affects the variable cost. The profit maximization of the firms implies that the FOB price of M-goods in region \( r \) is a constant markup over the marginal cost of production in the region.

\[
p_r^M = \frac{\sigma_M}{\sigma_M - 1} w^r e^{w^W} \quad (16)
\]

The zero-profit condition of the firms implies that the equilibrium output of each variety is equal to a constant that depends on the productivity of the region.

\[
\sum_{r=1}^{N_R} (1 + \delta_{rr})^\varphi \lambda_r m_{rr} = (\sigma_M - 1) F e^{w^W} \quad (17)
\]

---

\(^{1}\) For simplicity, I assume that the residential land and the land supplied for the production of A-goods are separate, and can never be converted.
Given this constant equilibrium output of each variety, the labor market-clearing condition implies that the total number of varieties of M-goods produced in region \( r \) is proportional to the population of households (workers) in the region.

\[
\lambda_r^M = \frac{\lambda_r}{\sigma_M F}
\]  

(18)

2.5 Spatial equilibrium

Equations (4)–(18) determine a unique short-term equilibrium of the model, given the fixed population distribution \( \lambda = [\lambda_1, \cdots, \lambda_N] \). In the equilibrium, the utility of a household in region \( r \) is expressed as follows.

\[
v_r = -\alpha \ln P_r^A - \beta \ln P_r^H - \mu \ln P_r^M + \ln I_r
\]

(19)

In the long run, households move to regions with higher utility until the utility is equalized across all the regions. This state is defined as a spatial equilibrium. From equation (12), \( H_r \) and \( v_r \) diverge to infinity as \( \lambda_r \) approaches zero. Therefore, all the regions have positive population in the spatial equilibrium. The condition of the spatial equilibrium is written as follows.

\[
v_1 = v_2 = \cdots = v_N = v_N
\]

(20)

There can be multiple equilibria that satisfy equation (20). In that case, some of the equilibria may be unstable. In an unstable equilibrium, when the population of some region is infinitesimally perturbed, the equilibrium collapses and the population converges to one of the stable equilibria. Because an unstable equilibrium cannot describe reality, it is important to check whether the equilibrium to be analyzed is stable or not. I discuss this issue in Subsection 3.6.

2.6 Structural residuals and model inversion

Suppose that the values of the parameters in the model of this section are known. In addition, suppose that there are available observed data on population (\( \lambda \)), wage (\( w = [w_1, \cdots, w_N] \)), area of available land (\( S = [S_1, \cdots, S_N] \)), and travel time (\( \delta = [\delta_1, \cdots, \delta_N, \delta_{11}] \)). Then, it can be shown that there is a one-to-one mapping from the observed data on the endogenous variables (\( \lambda \) and \( w \)) to the values of the exogenous amenities (\( \epsilon^A = [\epsilon_1^A, \cdots, \epsilon_N^A] \)) and the exogenous productivity (\( \epsilon^W = [\epsilon_1^W, \cdots, \epsilon_N^W] \)). In other words, the model can be inverted to recover the unique values of \( \epsilon^A \) and \( \epsilon^W \) that exactly rationalize the observed data as an equilibrium outcome.

In this context, \( \epsilon^A \) and \( \epsilon^W \) can be called “structural residuals,” because they stand in the unobservable determinants of the endogenous variables. In other words, a part of the variance of the endogenous variables can be explained by the model structure while the rest is explained by the structural residuals.

Although the unique values of \( \epsilon^A \) and \( \epsilon^W \) can be recovered, the expression is not precise. This is because the observed data of the endogenous variables are informative only on the differences of the structural residuals. In the equilibrium, \( \epsilon^A \) is used to match \( \lambda \) to the observed data. Suppose that \( \lambda \) satisfies equation (20) when \( \epsilon^A = 0 \). Then, equation (20) is still satisfied when \( \epsilon^A = 1 \), because it increases every \( v_r \) by \( \beta \). To recover the unique values of the exogenous amenities, a reference region should be chosen.

I use region \( N_R \) as the reference region in this study, and define the normalized amenities, denoted by \( \xi^A = [\xi^A_1, \cdots, \xi^A_{N_R-1}] \), as follows.

\[
\xi^A_r = \epsilon^A_r - \epsilon^A_{N_R}
\]

(21)

The values of \( \xi^A \) can be uniquely recovered. A similar discussion applies to \( \epsilon^W \). \( \epsilon^W \) is used to match \( w \) in the equilibrium to the observed data. Because the formulated model is a general equilibrium model, \( w \) in the model is determined as relative prices. Therefore, to match \( w \) to the observed data, a reference region should be chosen, of which wage is used as the numeraire. Again, I use region \( N_R \) as the reference region and suppose \( w_{N_R} = 1 \). Then, I define the normalized wage, denoted by \( \bar{w}_r = [\bar{w}_1, \cdots, \bar{w}_{N_R-1}] \), as follows.

\[
\bar{w}_r = \frac{w_r}{w_{N_R}}
\]

(22)

Because \( \bar{w} \) is an \((N_R - 1)\)-dimensional vector, one of the elements of \( \epsilon^W \) cannot be recovered. To recover the unique values of the exogenous productivity, I define normalized productivity, denoted by \( \xi^W = [\xi^W_1, \cdots, \xi^W_{N_R-1}] \), as follows.

\[
\xi^W_r = \epsilon^W_r - \epsilon^W_{N_R}
\]

(23)
The values of $\xi^W$ can be uniquely recovered.

### 3 | ESTIMATION METHOD WITH FIML

#### 3.1 Quantitative spatial economics model as a spatial econometric model

Quantitative spatial economics models can be implicitly written as the following simultaneous equations.

$$v_i(\lambda, \xi^\lambda, \epsilon^X) = v_N(\lambda, \xi^\lambda, \epsilon^X) \quad (1 \leq i \leq N - 1)$$

(24)

$$\sum_{i=1}^{N} \lambda_i = 1$$

(25)

$$f_j(\lambda, x, \xi^\lambda, \epsilon^X) = 0 \quad (1 \leq j \leq M)$$

(26)

Here, $\lambda = [\lambda_1, \ldots, \lambda_N]^T$ denotes an $N$-dimensional vector, and its element $\lambda_i$ is the share of households that choose alternative $i$. The elements of $\lambda$ sum up to one, as shown in equation (25). $x = [x_1, \ldots, x_M]^T$ denotes the $M$-dimensional vector of the observable endogenous variables. $\xi^\lambda = [\xi^\lambda_1, \ldots, \xi^\lambda_{N-1}]^T$ denotes the $(N-1)$-dimensional vector of the structural residuals to match $\lambda$ with the observed data. To recover the unique values of $\xi^\lambda$, alternative $N$ is used as the reference. $\epsilon^X = [\epsilon^X_1, \ldots, \epsilon^X_M]^T$ denotes the $M$-dimensional vector of the structural residuals to match $x$ with the observed data. $v_i$ denotes the indirect utility function of the households that choose alternative $i$. $f_1, \ldots, f_M$ denote functions that compose simultaneous equations (26) to determine $x$ given $\lambda, \xi^\lambda$, and $\epsilon^X$. $x$ is uniquely determined by these equations. Equations (24) and (25) are the condition of spatial equilibrium in which the utility of mobile households is equalized among all the choice alternatives. $\lambda$ is determined by these equations given $\xi^\lambda$ and $\epsilon^X$. There can be multiple solutions (equilibria) of $\lambda$ that satisfy the equations.

In the model of Section 2, $x$ and $\epsilon^X$ correspond to $w$ and $\xi^W$, respectively. Note that $x$ cannot include unobservable endogenous variables, for which data are not available. Unobservable endogenous variables, like $P^A$ in the model of Section 2, should be expressed as functions of $\lambda, x, \xi^\lambda$, and $\epsilon^X$ to construct simultaneous equations in the form of equations (24)—(26).

This study contends that equations (24)—(26) can be used as a spatial econometric model by defining the probability distribution of the structural residuals. In other words, the structural residuals can be regarded as the error terms of an econometric model. The first step is to formulate the joint probability density function of the structural residuals denoted by $\psi(\xi^\lambda, \epsilon^X)$. Note that the likelihood of the model is not equal to $\psi$, because the random variables are transformed from $\xi^\lambda$ and $\epsilon^X$ to $\lambda$ and $x$ through the simultaneous equations. The determinant of the Jacobian of $\xi^\lambda$ and $\epsilon^X$ with respect to $\lambda$ and $x$ is necessary to transform the probability density function from that of the structural residuals to that of the endogenous variables.

From equation (25), $\lambda_N$ can be expressed as $1 - \sum_{i=1}^{N-1} \lambda_i$. I inject this relation into equations (24) and (26), erasing $\lambda_N$. Then, I derive $(N + M - 1)$-th-order simultaneous equations with respect to $\lambda_{-N} = [\lambda_1, \ldots, \lambda_{N-1}]^T$ and $x$. Here, $\lambda_{-N}$ is an $(N-1)$-dimensional vector of $\lambda$, excluding $\lambda_N$. I take the total derivative of the $(N + M - 1)$-th-order equations around the equilibrium, where the endogenous variables are identical to the observed data and the structural residuals satisfy equations (24)—(26). The following linearized simultaneous equations are derived through the total derivatives.

$$\sum_{i=1}^{N-1} \left( \frac{\partial v_N}{\partial \lambda_i} - \frac{\partial v_N}{\partial \lambda_N} - \frac{\partial v_i}{\partial \lambda_i} + \frac{\partial v_i}{\partial \lambda_N} \right) d\lambda_i = - \sum_{j=1}^{M} \left( \frac{\partial v_N}{\partial \xi^\lambda_j} - \frac{\partial v_i}{\partial \xi^\lambda_j} \right) d\xi^\lambda_j - \sum_{j=1}^{M} \left( \frac{\partial v_N}{\partial \epsilon^X_j} - \frac{\partial v_i}{\partial \epsilon^X_j} \right) d\epsilon^X_j \quad (1 \leq i \leq N - 1)$$

(27)
Calculation of partial derivatives and log-likelihood function

In this study, I use $\partial a / \partial b$ to denote the Jacobian of a vector $a$ with respect to a vector $b$. $f = [f_1, \cdots, f_M]^T$ is a vector function. $J^x$, $J^\xi$, and $J^\lambda$ are submatrices expressed as follows.

$$\left( J^\lambda \right)_{i,j} = \frac{\partial \lambda_N}{\partial \lambda_i} - \frac{\partial \lambda_N}{\partial \lambda_j} + \frac{\partial \lambda_i}{\partial \lambda_j} (30)$$

$$\left( J^\xi \right)_{i,j} = \frac{\partial \xi^\Lambda}{\partial \xi_i} - \frac{\partial \xi^\Lambda}{\partial \xi_j} (31)$$

$$\left( J^x \right)_{i,j} = \frac{\partial x^\Lambda}{\partial x_i} - \frac{\partial x^\Lambda}{\partial x_j} (32)$$

In this study, I use $(A)_{i,j}$ to denote the entry in the $i$-th row and $j$-th column of a matrix $A$. With this result, the Jacobian to transform the probability density function is written as follows.

$$\frac{\partial [\xi^\Lambda, x^\Lambda]}{\partial [\lambda_N, x]} = - \left[ J^\xi \right]^{-1} \left[ J^x \right] \left[ \frac{\partial f}{\partial x} \right] (33)$$

With this result, the log-likelihood function of the model of (24)-(26) is written as follows:

$$LLF = \ln \psi(\xi^\Lambda, x^\Lambda) + \ln \left| \frac{\partial [\xi^\Lambda, x^\Lambda]}{\partial [\lambda_N, x]} \right| = \ln \psi(\xi^\Lambda, x^\Lambda) + \ln |\det J^\Lambda| + \ln |\det \frac{\partial f}{\partial x}| - \ln \left| \det \left[ \frac{\partial f}{\partial x} \right] \right| (34)$$

The parameters of the economic model and the probability distribution $\psi$ are estimated by maximizing this log-likelihood function. This log-likelihood function is valid regardless of whether the model has multiple equilibria. When there are multiple equilibria, I assume that the equilibrium closest to the observed data is selected for some reason.

### 3.3 Calculation of partial derivatives and log-likelihood function

Although equations (24)-(25) are simple and easy to understand, general quantitative spatial economics models cannot be explicitly written in that way owing to unobservable endogenous variables. General models are explicitly written as follows.

$$\tilde{\ell}_i(\lambda, x, y, \xi^\Lambda, x^\Lambda) = \bar{e}_N(\lambda, x, y, \xi^\Lambda, x^\Lambda) \quad (1 \leq i \leq N - 1) \quad (35)$$

$$\sum_{i=1}^{N} \tilde{\ell}_i = 1 \quad (36)$$

$$\bar{f}_j(\lambda, x, y, \xi^\Lambda, x^\Lambda) = 0 \quad (1 \leq j \leq M) \quad (37)$$

$$\bar{g}_k(\lambda, x, y, \xi^\Lambda, x^\Lambda) = 0 \quad (1 \leq k \leq K) \quad (38)$$

2 $\det A^{-1} = 1 / (\det A)$ for a non-singular matrix $A$. The following identity equation holds for a block matrix, where $A$, $B$, $C$, and $D$ are the submatrices. $\det \left[ \begin{array}{cc} A & O \\ C & D \end{array} \right] = \det A \times \det D = \det \left[ \begin{array}{cc} A & B \\ O & D \end{array} \right]$.

3 For example, suppose that there are two random variables $x$ and $\varepsilon$ that satisfy $x^2 = \varepsilon$. $\varepsilon$ is always positive and its probability density function is expressed as $\psi(\varepsilon)$. Given $\varepsilon$, there are two values of $x$. One is positive and the other is negative. Here, I assume that positive $x$ is always selected for some reason. In this case, the log-likelihood function of the observation is written as follows.

$$LLF = \ln \psi(\varepsilon) + \ln \left| \frac{d \varepsilon}{d x} \right| = \ln \psi(\varepsilon^2) + \ln 2x$$

This is valid, because there is a one-to-one mapping from $x$ to $\varepsilon$. 

$$\bar{f}_j(\lambda, x, y, \xi^\Lambda, x^\Lambda) = 0 \quad (1 \leq j \leq M) \quad (37)$$

$$\bar{g}_k(\lambda, x, y, \xi^\Lambda, x^\Lambda) = 0 \quad (1 \leq k \leq K) \quad (38)$$
Here, \( y = [y_1, \ldots, y_K]^T \) denotes the \( K \)-dimensional vector of the unobservable endogenous variables. \( \bar{\epsilon}_1, \ldots, \bar{\epsilon}_N \) denote functions that determine the utility of \( N \) alternatives given the endogenous variables (\( \lambda, x, \) and \( y \)) and the structural residuals (\( \xi^h \) and \( \epsilon^X \)). \( \bar{f}_1, \ldots, \bar{f}_M \), and \( \bar{g}_1, \ldots, \bar{g}_K \) denote functions that compose simultaneous equations (37) and (38) to determine \( x \) and \( y \) given \( \lambda, \xi^h, \) and \( \epsilon^X \).

Equations (35)–(38) can be implicitly written as equations (24)–(26). Let \( x = X(\lambda, \xi^h, \epsilon^X) \), and \( y = Y(\lambda, \xi^h, \epsilon^X) \) denote a vector function that determines \( x \) and \( y \) given its arguments, respectively. These functions are derived by solving equations (37) and (38) with respect to \( x \) and \( y \). In addition, let \( y = Y_x(\lambda, x, \xi^h, \epsilon^X) \) denote a vector function that determines \( y \) given its arguments. This function is derived by solving equation (38) with respect to \( y \). With \( X, Y, Y_x, \bar{v}, \bar{f}_j, \bar{v}, \) and \( f_j \) in equations (24)–(26) are expressed as follows.

\[
v_j(\lambda, \xi^h, \epsilon^X) = \bar{v}_j(\lambda, X(\lambda, \xi^h, \epsilon^X), Y(\lambda, \xi^h, \epsilon^X)) \tag{39}
\]

\[
f_j(\lambda, x, \xi^h, \epsilon^X) = \bar{f}_j(\lambda, x, Y_x(\lambda, x, \xi^h, \epsilon^X), \xi^h, \epsilon^X) \tag{40}
\]

To calculate the log-likelihood function of equation (34), the partial derivatives of \( \bar{v} \), \( \bar{f} \), and \( \bar{g} \) are calculated around the equilibrium. Next, the partial derivatives of \( \bar{v} \) are calculated around the equilibrium. Next, the partial derivatives of \( \bar{v} \) and \( \bar{g} \) are calculated around the equilibrium. Next, the partial derivatives of \( \bar{v} \) and \( \bar{f} \) are calculated with equations (42) and (44). Finally, equation (34) is used to calculate the log-likelihood.
3.4 Application to panel data

FIML introduced in Subsections 3.1–3.3 is applied to cross-sectional data collected at one period of time. The method can be easily extended to be applicable to panel data collected at several periods of time. Suppose that the data are collected at $T$ periods of time. I use subscript $t$ ($1 \leq t \leq T$) to denote variables and functions at period $t$. Let $\psi(\xi^X_t, \ldots, \xi^X_T, \epsilon^X_1, \ldots, \epsilon^X_T)$ denote the joint probability density function of all the structural residuals from all the periods. It would be better to consider the temporal correlation of the structural residuals, because the intrinsic nature of a region would not change so quickly. One option for considering the temporal correlation is to use the random effects model as in standard panel data analyses. I introduce this approach in the next subsection.

Because a static economic model is assumed in this study, the Jacobian of the model to transform the probability density function can be calculated independently for each period. In the end, the log-likelihood function is written as follows.

$$L = \sum_{t=1}^T \ln |J^T_{t,t}| + \sum_{t=1}^T \ln \left| \det \left[ \frac{\partial \xi^X_t}{\partial \lambda_{N,N,t}} \right] \right| - \ln \left| \det \left[ J^T_{\epsilon^X_t}\epsilon^X_t J^T_{\epsilon^X_t}\epsilon^X_t \right] \right|$$

3.5 Formulating the probability density function of the structural residuals

To apply FIML to quantitative spatial economics models, it is necessary to formulate the probability density function of the structural residuals. Appropriate formulation would depend on the model. In this subsection, I use the model of Section 3.4 as a specific illustration, and formulate the probability density function of its structural residuals ($\xi^A$ and $\xi^W$). It would be possible to adopt a similar formulation in other models after appropriate modifications.

3.5.1 Cross-sectional data

First, I consider the case in which FIML is applied to cross-sectional data. In general, structural residuals in adjacent regions should have spatial correlation. Therefore, it would be better to introduce spatial correlation into the formulation. The first choice is the well-known spatial error model. I express $\epsilon^G$ ($G \in \{ A, W \}$) as follows.

$$\epsilon^G = \rho_G \Omega_G \epsilon^G + \tilde{\epsilon}^G$$

(46)

Here, $\rho_G$ is a parameter to determine the magnitude of the spatial correlation. $\Omega_G$ is a spatial weight matrix. $\tilde{\epsilon}^G$ is a vector of random variables that follow an independent and identically distributed (i.i.d.) normal distribution $N(0, s^2_G)$. I use exponential distance weights to construct the spatial weight matrix.

$$\Omega_G = \begin{cases} \frac{\exp(-\zeta_G \delta_{rr}^G)}{\sum_{r'=r}^{\exp(-\zeta_G \delta_{rr}^G)} (r \neq r')} & (r \neq r') \\ 0 & (r = r') \end{cases}$$

(47)

Here, $\zeta_G$ is a positive parameter. $\delta_{rr}^G$ is the travel time between regions $r$ and $r'$ in the spatial model. Equation (46) implies that the covariance matrix of $\epsilon^G$, denoted by $\Sigma_G$, is expressed as follows.

$$\Sigma_G = s^2_G \frac{(E - \rho_G \Omega_G)^{-1}}{\left( E - \rho_G \Omega_G \right)^{-1}}$$

(48)

$\Sigma_G$ is not the covariance matrix of $\xi^G$. Note that $\xi^G$ follows a multivariate normal distribution because the summation of two normal random variables is also a normal random variable. Let $N(0, K_G)$ denote the normal distribution of $\xi^G$, where $K_G$ is the covariance matrix of $\xi^G$. $K_G$ is expressed with $\Sigma_G$ as follows.

$$(K_G)_{r,r'} = E \left[ \xi_{r}^G \xi_{r'}^G \right] = E \left[ (\epsilon_{r}^G - \epsilon_{N_{r}}^G) (\epsilon_{r'}^G - \epsilon_{N_{r'}}^G) \right] = (\Sigma_G)_{r,r'} - (\Sigma_G)_{r,N_{r}} - (\Sigma_G)_{r',N_{r'}} + (\Sigma_G)_{N_{r},N_{r'}}$$

(49)

With this result, the log-probability density function of $\xi^G$ is written as follows.

$$-\frac{N_R - 1}{2} \ln(2\pi) - \frac{1}{2} \ln \det K_G - \frac{1}{2} \left( \xi^G \right)^T K_G^{-1} \xi^G$$

(50)

In this formulation, $\rho_G$, $\zeta_G$, and $s_G$ are the parameters of the density function.
I assume that $\xi^A$ and $\xi^W$ are independent of each other. Under this assumption, the log-probability density function of $\xi^A$ and $\xi^W$ is written as follows.

$$\ln \psi(\xi^A, \xi^W) = -\frac{N_R - 1}{2} \ln(2\pi) - \frac{\ln \det K_A}{2} - \frac{1}{2} (\xi^A)^T K_A^{-1} \xi^A - \frac{N_R - 1}{2} \ln(2\pi) - \frac{\ln \det K_W}{2} - \frac{1}{2} (\xi^W)^T K_W^{-1} \xi^W \quad (51)$$

### 3.5.2 Panel data

Next, I consider the case in which FIML is applied to panel data collected at $T$ periods of time. As mentioned in the previous subsection, it would be better to consider the temporal correlation of the structural residuals. Here, I use the random effects model as in standard panel data analyses. I express $\epsilon^G_i = (G \in \{A, W\}, 1 \leq t \leq T)$ as follows.

$$\epsilon^G_i = \epsilon_i^{G_{\text{const}}} + \epsilon_i^{G_{\text{var}}} \quad (52)$$

Here, $\epsilon_i^{G_{\text{const}}}$ denotes a vector of individual effects, which indicate the intrinsic nature of the regions. $\epsilon_i^{G_{\text{var}}}$ is common to all periods of time. $\epsilon_i^{G_{\text{var}}}$ denotes a vector of temporal error terms at period $t$.

I assume that $\epsilon_i^{G_{\text{const}}}$ follows a multivariate normal distribution $N(0, \Sigma_G)$ where $\Sigma_G$ is defined by equations (47) and (48). In this case, $1/T \sum_{t=1}^T \delta_{rr,t}$ is used in equation (47) instead of $\delta_{rr,t}$. Meanwhile, I assume that the elements of $\epsilon_i^{G_{\text{var}}}$ follow an i.i.d. normal distribution $N(0, s_{G_{\text{var}}}^2)$. In other words, I do not consider the spatial correlation of $\epsilon_i^{G_{\text{var}}}$. The reason is for computational efficiency, as discussed at the end of this subsection.

Let $N(0, K_{GT})$ denote the normal distribution of $\xi^G = ([\epsilon^G_1]^T, \ldots, [\epsilon^G_T]^T)^T$, where $K_{GT}$ is the $T(N_R - 1) \times T(N_R - 1)$ covariance matrix of $\xi^G$. $K_{GT}$ is expressed as follows with $\Sigma_G$ and $s_{G_{\text{var}}}^2$.

$$K_{GT}(r+(N_R-1)(t-1)+s(N_R-1)(r'-1), r'+s(N_R-1)(t'-1)) = E \left[ \xi^G_r \xi^G_{r'} \right] = E \left[ \left( \epsilon_i^G - \epsilon^G_N \right \left\{ \left( \epsilon_i^G \right)_r - \epsilon^G_N \right\right) \right] = \left( \Sigma_G \right)_{r', r} - \left( \Sigma_G \right)_{r, N} + \left( \Sigma_G \right)_{N, r'} + \left( \Sigma_G \right)_{N, N}$$

$$+ \text{if}(r = r' \text{ and } t = t') \cdot s_{G_{\text{var}}}^2 + \text{if}(t = t') \cdot s_{G_{\text{var}}}^2 \quad (53)$$

Here, if() is a logical function that returns 1 if the contents of the parentheses are true and returns 0 otherwise. With this result, the log-probability density function of $\xi^G$ is written as follows.

$$\ln \psi(\xi^G) = -\frac{T(N_R - 1)}{2} \ln(2\pi) - \frac{\ln \det K_{GT}}{2} - \frac{1}{2} (\xi^G)^T K_{GT}^{-1} \xi^G \quad (54)$$

In this formulation, $\rho_G$, $\zeta_G$, $s_G$, and $s_{G_{\text{var}}}$ are the parameters of the density function.

I assume that $\xi^A$ and $\xi^W$ are independent of each other. Under this assumption, the log-probability density function of $\xi^A$ and $\xi^W$ is written as follows.

$$\ln \psi(\xi^A, \xi^W) = -\frac{T(N_R - 1)}{2} \ln(2\pi) - \frac{\ln \det K_{AT}}{2} - \frac{1}{2} (\xi^A)^T K_{AT}^{-1} \xi^A$$

$$- \frac{T(N_R - 1)}{2} \ln(2\pi) - \frac{\ln \det K_{WT}}{2} - \frac{1}{2} (\xi^W)^T K_{WT}^{-1} \xi^W \quad (55)$$

### 3.6 Relation to the stability of equilibrium

As mentioned in Section 1 and Subsection 2.5, it is important to check whether the equilibrium is stable under the estimated parameters. However, when the equilibrium is unstable, there is nothing to deal with the problem. For example, it is impossible to optimize the objective function of the estimation method subject to the constraint that the equilibrium is stable. This is because the equilibrium is unstable on the boundary that separates the parameter space into the domain of stable equilibrium and the domain of unstable equilibrium. A continuous optimization problem with an open constraint cannot be solved.

In the case of FIML proposed in this study, this problem might not be a concern. In many cases, the maximum of the log-likelihood function would be within the domain of stable equilibrium. Here, I discuss this property.

To define stability, a dynamic should be formulated to govern the evolution of population distribution ($\lambda$) given the indirect utility function ($v$). Here, I use the projection dynamic [122] instead of the well-known replicator dynamic. The projection dynamic is described by the following differential equations.

$$\dot{\lambda}_i = v_i(\lambda) - \frac{1}{N} \sum_{i' = 1}^N v_{i'}(\lambda) \quad (1 \leq i \leq N) \quad (56)$$
I omit the structural residuals in the arguments of \( \nu \) because they are fixed when the stability is analyzed. This equation means that the households are attracted to regions with higher utility compared to the average, which is the same outcome as the replicator dynamic. The difference is how they calculate the average. The projection dynamic uses the simple average of utility across regions while the replicator dynamic uses the weighted average utilizing the population as the weighting factors.

Next, the Jacobian of the differential equations (56) is derived. Since \( \sum_{i=1}^{N} \lambda_i = 1 \) is always satisfied, one of the equations in (56) is redundant. Therefore, an equation should be dropped before deriving the Jacobian. Here, I drop the equation of \( \lambda_N \). I inject \( \lambda_N = 1 - \sum_{i=1}^{N-1} \lambda_i \) into the other equations, deriving \( (N-1) \)-th-order differential equations. The Jacobian of these \( (N-1) \)-th-order differential equations, denoted by \( J \), is expressed as follows.

\[
(J)_{i,i} = \frac{\partial v_i}{\partial \lambda_{i'}} - \frac{\partial v_i}{\partial \lambda_N} - \frac{1}{N} \sum_{i'=1}^{N} \left( \frac{\partial v_{i'}}{\partial \lambda_{i'}} - \frac{\partial v_{i'}}{\partial \lambda_N} \right)
\]  

(57)

The necessary and sufficient condition of the stability is that all the eigenvalues of \( J \) have negative real parts in the equilibrium.\(^4\) The Jacobian \( J \) has a close relation to Jacobian \( J_4^6 \) defined in equation (30). The following equation is an identity equation about these Jacobians.

\[
\frac{1}{N} \det J_4^6 = (-1)^{N-1} \det J
\]

(58)

The proof of equation (58) is shown in Appendix A. The right-hand side of equation (58) is the constant term of the characteristic polynomial of \( J \), which is expressed as \( \det(tE-J) \) using \( t \) to denote the variable. According to the Routh–Hurwitz stability criterion, a necessary condition of stability is for all the coefficients of the characteristic polynomial to be positive. Therefore, if the value of equation (58) is negative or zero, the equilibrium is unstable.

Identity equation (58) is the core of the discussion in this subsection. The value of equation (58) is always positive in the domain of the parameter space where the equilibrium is stable. Meanwhile, the value is either zero or positive on the boundary of the domain. When one or more eigenvalues with a zero real part is zero at a point on the boundary,\(^5\) then the value of equation (58) is zero, since the determinant of a matrix is the product of all its eigenvalues. When all the eigenvalues with zero real parts are conjugate pairs of imaginary numbers expressed as \( \pm ai (a > 0) \) at a point on the boundary, the value of equation (58) is positive.

If the determinant of \( J \) is always zero on the boundary, the maximum of the log-likelihood function of equation (34) or (45) is within the domain of stable equilibrium, because the function diverges to negative infinity as the parameters approach the boundary. Even if all the eigenvalues of \( J \) are complex on some part of the boundary, the maximum of the log-likelihood function would still be within the domain of stable equilibrium in many cases, because the zero real parts makes the determinant and the likelihood small on the boundary. This discussion is valid regardless of whether the model has multiple equilibria or not. Therefore, FIML is effective in the situation in which the agglomeration force is expected to be sufficiently strong to support multiple equilibria.

Equation (58) can also be used for the optimization of the log-likelihood function. First, the initial values of the parameters are set so that the agglomeration force is sufficiently weak and the equilibrium is stable. When a line search is performed during the optimization, the value of \( \det J_4^6 \) under the parameter values in the next step is checked. If \( \det J_4^6 \) is negative, it means that the optimization routine jumps into the domain of unstable equilibrium. In such a case, the step size of the line search is reduced until \( \det J_4^6 \) becomes positive. Of course, this does not guarantee that the parameters are within the domain of stable equilibrium, because \( \det J_4^6 \) can be positive even in the domain of unstable equilibrium. Therefore, the eigenvalues of \( J \) should be checked after the maximum of the log-likelihood function is found. I confirm the effectiveness of this strategy in the analyses of Section 4. If the unreliability of this strategy is a concern, the eigenvalues of \( J \) should be checked in a line search although it is more computationally costly to do so.

---

\(^4\) The Jacobian of the \( N \)-th-order differential equations of (56) always have a zero eigenvalue because of its redundancy. Therefore, it is necessary to drop one of the equations.

\(^5\) Since the eigenvalues of a matrix are continuous with respect to its entries, all the eigenvalues of \( J \) have either zero or negative real parts on the boundary as long as the partial derivatives of the utility are continuous with respect to the model parameters.

\(^6\) Since \( J \) is a real matrix, all of its eigenvalues are either real numbers or conjugate pairs of complex numbers.
4 | RESULT OF FIML APPLICATION TO QUANTITATIVE SPATIAL MODEL

In this section, I confirm the validity of FIML by applying it to the spatial model of Section 2 with two kinds of data: experimental data generated from a model with known parameters; and observed data of Japan from 1975 to 2015.

4.1 Estimation result with experimental data

Suppose an economy consisting of $N_R = 100$ regions arranged linearly. Each region has a unit of available land, and hence, $\bar{S}_r = 1$ for all the regions. I assume that the travel time between regions $r$ and $r'$ ($\delta_{rr'}$) is defined as follows.

$$\delta_{rr'} = \tau |r - r'|$$  \hspace{1cm} (59)

I use $\tau = 0.2$ as a constant to indicate the travel time to move from a region to its next region. I set the true parameter values of the spatial model as follows: $\alpha = 0.35$, $\beta = 0.1$, $\mu = 0.55$, $\phi_A = 0.5$, $\phi_M = 1$, $\sigma_A = \sigma_M = 5$.

Under the abovementioned setting, I generate cross-sectional data of this economy as follows. First, I assume that the structural residuals ($\Lambda$ and $W$) follow the spatial error model of equation (46). With this stochastic model, I generate $\Lambda$ and $W$ under the following parameter values: $\rho_{\Lambda} = 0.5$, $\zeta_{\Lambda} = 5$, $s_{\Lambda} = 0.3$, $\rho_{W} = 0.5$, $\zeta_{W} = 5$, $s_{W} = 0.03$. Then, I search for a spatial equilibrium with projection dynamic of equation (56). The population and wage in the equilibrium are supplied as cross-sectional data to FIML.

I find two equilibria, which means that the economy has multiple equilibria. I call them Equilibrium A and Equilibrium B, respectively. The geographical distribution of population and wage in these two equilibria are shown in Figure 1. The left graph is of Equilibrium A and the right graph is of Equilibrium B. The population data are shown by blue bars, and the wage data are shown by orange lines. Wage data are normalized so that their average across the regions is unity ($1/N_R \sum_{r=1}^{N_R} w_r = 1$). Note that Equilibrium A and B share the same structural residuals.

I apply FIML to the data of both Equilibrium A and B separately. The estimation result is shown in Table 1. Because either $\sigma_A$ or $\phi_A$ is unidentified, I fix the value of $\sigma_A$ at its true value of 5 and estimate the value of $\phi_A$. The share parameters of the utility function must sum up to one ($\alpha + \beta + \mu = 1$). Thus, I express $\mu$ as $1 - \alpha - \beta$ in the optimization routine. In this case, the standard error of $\mu$ is expressed as $\sqrt{\sigma_\alpha^2 + \sigma_\beta^2 + 2\sigma_{\alpha\beta}}$, where $\sigma_\alpha^2$ and $\sigma_\beta^2$ denote the variances of the estimators of $\alpha$ and $\beta$, and $\sigma_{\alpha\beta}$ denotes the covariance of these estimators. The column of “True value” shows the true values of the parameters. The columns of “Equilibrium A” and “Equilibrium B” show the estimation results with the data of Equilibrium A and Equilibrium B, respectively.

All the estimates of the parameters of the spatial model are significant at the 1% level. Many of the 95% confidence intervals of the parameters of the spatial model include their true values. This result confirms that FIML works even if the quantitative spatial economics model has multiple equilibria.

Although I find the local maxima of the log-likelihood function with respect to $\zeta_{\Lambda}$ and $\zeta_{W}$ in the abovementioned analysis, such local maxima do not necessarily exist. In some cases, the log-likelihood function is monotonically increasing with respect to $\zeta_{\Lambda}$ and $\zeta_{W}$.
TABLE 1 Estimation result with experimental data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Equilibrium A</th>
<th>Equilibrium B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.35</td>
<td>0.2249$^{**}$</td>
<td>0.3694$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0731)</td>
<td>(0.0390)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.1</td>
<td>0.1158$^{**}$</td>
<td>0.0915$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0115)</td>
<td>(0.0074)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.55</td>
<td>0.6593$^{**}$</td>
<td>0.5391$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0618)</td>
<td>(0.0321)</td>
</tr>
<tr>
<td>$\phi_A$</td>
<td>0.5</td>
<td>0.5546$^{**}$</td>
<td>0.7464$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1260)</td>
<td>(0.0828)</td>
</tr>
<tr>
<td>$\phi_M$</td>
<td>1</td>
<td>0.9542$^{**}$</td>
<td>1.0363$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0166)</td>
<td>(0.0271)</td>
</tr>
<tr>
<td>$\sigma_M$</td>
<td>5</td>
<td>5.2794$^{**}$</td>
<td>4.7769$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1062)</td>
<td>(0.0737)</td>
</tr>
<tr>
<td>$\rho_\Lambda$</td>
<td>0.5</td>
<td>0.2655$^*$</td>
<td>0.2275</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1559)</td>
<td>(0.1629)</td>
</tr>
<tr>
<td>$\zeta_\Lambda$</td>
<td>5</td>
<td>4.8556</td>
<td>4.7741</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.3438)</td>
<td>(3.7379)</td>
</tr>
<tr>
<td>$s_\Lambda$</td>
<td>0.3</td>
<td>0.2821$^{**}$</td>
<td>0.2777$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0205)</td>
<td>(0.0205)</td>
</tr>
<tr>
<td>$\rho_W$</td>
<td>0.5</td>
<td>0.5392$^{**}$</td>
<td>0.5452$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1071)</td>
<td>(0.1130)</td>
</tr>
<tr>
<td>$\zeta_W$</td>
<td>5</td>
<td>7.4178$^*$</td>
<td>7.1742$^*$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.2262)</td>
<td>(4.2634)</td>
</tr>
<tr>
<td>$s_W$</td>
<td>0.03</td>
<td>0.0276$^{**}$</td>
<td>0.0283$^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0020)</td>
<td>(0.0021)</td>
</tr>
</tbody>
</table>

Full information maximum likelihood estimates. Standard errors from the negative Hessian matrix of the log-likelihood function are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

to $\zeta_\Lambda$ or $\zeta_W$. In other words, these parameters can be unidentified. In such cases, $\zeta_\Lambda$ or $\zeta_W$ must be exogenously fixed by the researcher to perform FIML estimation.

4.1.1 Monte Carlo experiment

I also perform a Monte Carlo experiment to investigate the efficiency of FIML estimators. Again, I consider an economy consisting of $N_R$ regions arranged linearly, each of which has a unit of available land. I consider five values as $N_R$: $N_R = 100$, $N_R = 200$, $N_R = 300$, $N_R = 400$ and $N_R = 500$. $\delta_{rr'}$ is defined by equation (59) and $\tau = 20/N_R$ is used. Under this setting, one to four agglomerations of population appear in the spatial equilibrium. I generate 100 cross-sectional data sets for each value of $N_R$ with the same parameter values used in the abovementioned analysis. Then, FIML is applied to estimate the parameter values for each data set. Table 2 summarizes the result of the experiment. The table shows the sample averages and standard deviations of the estimates calculated from the 100 data sets for each value of $N_R$. Unlike the analysis of Table 2, I fixed $\zeta_\Lambda$ and $\zeta_W$ at their true values in the estimation process. This is because these parameters are unidentified with some data sets although they are identified with most of the data sets.

Table 2 confirms that the estimators of $\alpha$, $\beta$, $\mu$, $\sigma_M$, $\rho_\Lambda$, and $\rho_W$ are biased even in the case of $N_R = 500$. This is because the generated cross-sectional data are highly spatially correlated, reducing the effective sample size. Nevertheless, the size of bias is not that large. If the bias of the estimators is of concern, parametric bootstrap can be used to quantify the size of bias, just as in this Monte Carlo experiment. Then, the bias can be corrected with the result.

Table 2 confirms that the variance of the estimators may be too large to use the estimation result for counterfactual exercises. However, this also applies to the other estimation methods. To reduce the variance, additional information should be supplied...
TABLE 2  Sample averages and standard deviations of FIML estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>( N_R = 100 )</th>
<th>( N_R = 200 )</th>
<th>( N_R = 300 )</th>
<th>( N_R = 400 )</th>
<th>( N_R = 500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.35</td>
<td>0.3914</td>
<td>0.3870</td>
<td>0.3728</td>
<td>0.3775</td>
<td>0.3780</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0896)</td>
<td>(0.0792)</td>
<td>(0.0677)</td>
<td>(0.0696)</td>
<td>(0.0641)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.1</td>
<td>0.0908</td>
<td>0.0920</td>
<td>0.0955</td>
<td>0.0948</td>
<td>0.0950</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0134)</td>
<td>(0.0116)</td>
<td>(0.0104)</td>
<td>(0.0103)</td>
<td>(0.0094)</td>
</tr>
<tr>
<td>( \mu )</td>
<td>0.55</td>
<td>0.5178</td>
<td>0.5210</td>
<td>0.5318</td>
<td>0.5277</td>
<td>0.5270</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0767)</td>
<td>(0.0682)</td>
<td>(0.0579)</td>
<td>(0.0596)</td>
<td>(0.0548)</td>
</tr>
<tr>
<td>( \phi_A )</td>
<td>0.5</td>
<td>0.5478</td>
<td>0.5329</td>
<td>0.5099</td>
<td>0.5001</td>
<td>0.4974</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1456)</td>
<td>(0.1395)</td>
<td>(0.1215)</td>
<td>(0.0840)</td>
<td>(0.0792)</td>
</tr>
<tr>
<td>( \phi_M )</td>
<td>1</td>
<td>1.0146</td>
<td>1.0064</td>
<td>1.0075</td>
<td>1.0008</td>
<td>1.0032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0260)</td>
<td>(0.0186)</td>
<td>(0.0204)</td>
<td>(0.0154)</td>
<td>(0.0149)</td>
</tr>
<tr>
<td>( \sigma_M )</td>
<td>5</td>
<td>4.9294</td>
<td>4.9471</td>
<td>4.9605</td>
<td>4.9675</td>
<td>4.9607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1028)</td>
<td>(0.0854)</td>
<td>(0.0762)</td>
<td>(0.0766)</td>
<td>(0.0777)</td>
</tr>
<tr>
<td>( \rho_A )</td>
<td>0.5</td>
<td>0.3833</td>
<td>0.4144</td>
<td>0.4103</td>
<td>0.4061</td>
<td>0.4184</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1434)</td>
<td>(0.1025)</td>
<td>(0.1303)</td>
<td>(0.1065)</td>
<td>(0.1132)</td>
</tr>
<tr>
<td>( s_A )</td>
<td>0.3</td>
<td>0.2892</td>
<td>0.2949</td>
<td>0.2958</td>
<td>0.2995</td>
<td>0.2987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0230)</td>
<td>(0.0141)</td>
<td>(0.0119)</td>
<td>(0.0111)</td>
<td>(0.0092)</td>
</tr>
<tr>
<td>( \rho_W )</td>
<td>0.5</td>
<td>0.4344</td>
<td>0.4317</td>
<td>0.4374</td>
<td>0.4499</td>
<td>0.4453</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1241)</td>
<td>(0.0988)</td>
<td>(0.1062)</td>
<td>(0.0990)</td>
<td>(0.0989)</td>
</tr>
<tr>
<td>( s_W )</td>
<td>0.03</td>
<td>0.0292</td>
<td>0.0300</td>
<td>0.0299</td>
<td>0.0299</td>
<td>0.0299</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0023)</td>
<td>(0.0015)</td>
<td>(0.0014)</td>
<td>(0.0012)</td>
<td>(0.0011)</td>
</tr>
</tbody>
</table>

The sample size is 100 for each value of \( N_R \). The values without parentheses are averages. The values in parentheses are standard deviations.

from outside of the data. For example, if the prior distribution of the parameters is known, Bayesian estimation could be used to derive the posterior distribution of the parameters with small variance.

In addition, the counterfactual exercises might not be affected so much even if the estimates deviate from the true values. This is because the deviation of a parameter is offset by the deviation of another parameter. The estimators of the parameters of the spatial model are correlated. Specifically, when the estimate of \( \alpha \) is greater than its true value, the estimates of the other parameters are generally less than their true values, and vice versa. In other words, when the estimate of \( \alpha \) is large, \( \phi_A \) is reduced to adjust the size of the global dispersion force (the global dispersion force ceases when \( \phi_A \) is zero), \( \beta \) is reduced to adjust the size of the total dispersion force, and \( \sigma_M \) is reduced to adjust the agglomeration force (the agglomeration force is strengthened as \( \sigma_M \) decreases). In the end, the behavior of the model under the estimated values of the parameters is not that different from its behavior under the true values of the parameters.

4.2 Estimation result with cross-sectional data of Japan in 2015

Next, I introduce the result of FIML application to cross-sectional data of Japan in the year 2015. I use the data of \( N_R = 1,376 \) municipalities (cities, towns, villages, and special wards of Tokyo) located on Honshu and Kyushu, the two largest islands of Japan. These 1,376 municipalities are indicated by the green polygons in Figure 2. The municipalities on the other islands (Hokkaido, Shikoku, Okinawa, and other small islands) are not considered in the analysis. This is because I use the data on the shortest travel time on the road network in 2015 as the travel time between municipalities (\( \delta_{rt} \)). Basically, those islands are not accessible from Honshu and Kyushu via road. Although Shikoku was accessible from Honshu via highways on bridges in 2015, I do not consider the municipalities in Shikoku in the analysis. This is because in the following subsection, I also apply FIML to panel data of the same municipalities from 1975 to 2015 to compare the result with that of the cross-sectional data. Shikoku was not accessible from Honshu via road before 1988. I also do not consider some municipalities in Fukushima Prefecture, because the population in these municipalities were zero or much smaller than that before 2011 due to evacuation following the Fukushima Daiichi nuclear disaster. The 1,376 municipalities held 91.2% of the Japanese total population in 2015, and 89.9% in 1975.
FIGURE 2 1,376 municipalities considered in the analysis (indicated by green polygons)

The data on the population (including the non-working-age population) are taken from the Japanese Census held in 2015\cite{12}. The population data are normalized so that $\sum_{r=1}^{N_R} \lambda_r = 1$. The data on wage at municipality level are not available in Japan. Instead, I use the data on taxable income in 2015\cite{10}. The total taxable income of each municipality is divided by the number of taxpayers in the municipality. Then, the quotient is used as the proxy for the wage ($w_r$). The data on the area of available land ($\bar{S}_r$) in 2015 are taken from the publication of the Statistics Bureau of Japan\cite{17}. The definition of available land area in the statistics is the total area subtracted by the undeveloped area of forests, plains, and lakes. Most Japanese territory consists of steep mountains, which are unsuitable for residential and commercial land use. Therefore, it is better to use the data on the available land area as $\bar{S}_r$ rather than to use the data on the total area.

The data on the travel time between municipalities ($\delta_{rr'}$) are collected with software called “NITAS,” developed by the Ministry of Land, Infrastructure, Transport and Tourism of Japan\cite{11}, to allow users to perform shortest path analysis on the Japanese transportation network. With this software, I collect the data on the shortest travel time between the municipal offices of the 1,376 municipalities on the road network in 2015. I use “hour” as the unit of travel time.

For the estimation, I assume that the probability density function of the structural residuals ($\xi^\Lambda$ and $\xi^W$) is expressed as equation (51). In the estimation process, I fix the value of $\sigma_A$ at 5. The estimation result is shown in Table 3. All the estimates of the share parameters ($\alpha$, $\beta$, and $\mu$) are statistically significant. The estimates of the parameters to determine the spatial correlation of the structural residuals ($\rho_\Lambda$ and $\rho_W$) are statistically significant, which means that the intrinsic attractiveness of the regions is spatially correlated. This is an expected result.

FIML maximizes the fitness of the model to the observed data. This can be confirmed by the graphs shown in Figure 3. These graphs are the scatter plots of the fitted values and the observed values of the endogenous variables in logarithmic scale. The left plot is about $\lambda_r$ and the right plot is about $w_r$. The straight lines in the graphs indicate the points where the fitted values and the observed values coincide. The fitted values are defined as the values of the endogenous variables in the equilibrium with zero structural residuals ($\epsilon^\Lambda = 0$, $\epsilon^W = 0$). I use $\hat{\lambda}_r$ and $\hat{w}_r$ to denote the fitted values of $\lambda_r$ and $w_r$, respectively. $\lambda_r$ and $w_r$ are normalized so that their averages across the regions are unity. I use the projection dynamic of equation (56) to search for the equilibrium with zero structural residuals. It appears that there is only a single equilibrium under the estimated parameter values of Table 3.

Figure 3 confirms clear correlations between the fitted values and the observed values. The correlation coefficients between the logarithms of the fitted values and the observed values are 0.689 for $\lambda_r$ and 0.558 for $w_r$. This implies that the spatial model can well explain the static geographical distribution of population and income despite its structural simplicity. This result also implies that FIML is effective for exploiting the ability of the spatial model to explain reality.
TABLE 3  Estimation result with cross-sectional data of Japan in 2015

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.1008**</td>
</tr>
<tr>
<td></td>
<td>(0.0421)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.2234***</td>
</tr>
<tr>
<td></td>
<td>(0.0127)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.6758***</td>
</tr>
<tr>
<td></td>
<td>(0.0342)</td>
</tr>
<tr>
<td>$\phi_A$</td>
<td>1.3441***</td>
</tr>
<tr>
<td></td>
<td>(0.2016)</td>
</tr>
<tr>
<td>$\phi_M$</td>
<td>0.8408***</td>
</tr>
<tr>
<td></td>
<td>(0.0571)</td>
</tr>
<tr>
<td>$\sigma_M$</td>
<td>3.9583***</td>
</tr>
<tr>
<td></td>
<td>(0.2587)</td>
</tr>
<tr>
<td>$\rho_\Lambda$</td>
<td>0.7774***</td>
</tr>
<tr>
<td></td>
<td>(0.0273)</td>
</tr>
<tr>
<td>$\zeta_\Lambda$</td>
<td>5.8891***</td>
</tr>
<tr>
<td></td>
<td>(0.4625)</td>
</tr>
<tr>
<td>$\varsigma_\Lambda$</td>
<td>0.5377***</td>
</tr>
<tr>
<td></td>
<td>(0.0110)</td>
</tr>
<tr>
<td>$\rho_W$</td>
<td>0.6740***</td>
</tr>
<tr>
<td></td>
<td>(0.0313)</td>
</tr>
<tr>
<td>$\zeta_W$</td>
<td>7.2600***</td>
</tr>
<tr>
<td></td>
<td>(0.7401)</td>
</tr>
<tr>
<td>$s_W$</td>
<td>0.1074***</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
</tr>
</tbody>
</table>

Full information maximum likelihood estimates. Standard errors from the negative Hessian matrix of the log-likelihood function in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

4.3  Estimation result with panel data of Japan from 1975 to 2015

Next, I introduce the result of FIML application to panel data of Japan from 1975 to 2015. I use 10-year interval data at 1975, 1985, 1995, 2005, 2015 ($T = 5$). I use the data of $N_R = 1,376$ municipalities shown in Figure 2. I fix the structure of the municipalities at that of 2015 although the municipalities had been restructured during the period since 1975. The data on the population (including the non-working-age population) are taken from the historical data of the Japanese Census. The population data are normalized for each year so that $\sum_{r=1}^{N_R} \lambda_{r,t} = 1$. I use the historical data on taxable income as a proxy for the wage, just as in the previous subsection. I use the data on the area of available land in 2015 for all the years, because the changes in the data during this period are small. The historical data on the travel time between municipalities are collected with NITAS. I collect the data on the shortest travel time between the municipal offices on the road network for each year. When the road network data in some year is unavailable, I use the data in the closest year. In the end, I use the road network data in 1976, 1986, 1996, 2006, and 2015. With this historical data of travel time and FIML estimation, I can investigate how the past road development in Japan has contributed to the changes in the geographical distribution of population and income.

For the estimation, I assume that the probability density function of the structural residuals ($\xi_{\Lambda T}$ and $\xi_{W T}$) is expressed as equation (55). In the estimation process, I fix the value of $\sigma_A$ at 5. The estimation result is shown in the middle column of Table 4. All the estimates of the parameters are significant at the 1% level. The estimates of $\alpha$ and $\beta$ are larger than those of Table 3.

---

7 The data of some municipalities decreased during this period, probably because of the renewed survey data. Therefore, I use the data of the newest period.
This means that both the local dispersion force and the global dispersion force are more important for explaining the transition of the geographical distribution of population and income than for explaining the static distribution.

To investigate the fitness of the spatial model to the observed data, I use scatter plots of the fitted values and the observed values of the endogenous variables. In this subsection, I define the fitted values as the values of the endogenous variables in the equilibrium with time-invariant structural residuals. In other words, the fitted values are calculated under the condition that the temporal error terms shown in equation (52) are zero (\(G_{\text{var}} = 0\)) but the individual effects (\(G_{\text{const}}\)) are not zero. The true values of the individual effects are unknown. Hence, I use the temporal mean of the structural residuals (\(1/T \sum_{t=1}^{T} G_{r,t}\)) under the estimated parameter values as the proxy for \(G_{\text{const}}\).

Since the interest here is how well the spatial model can explain the transition of the geographical distribution of population and income, I plot the changes in the endogenous variables. The graphs in Figure 4 are the scatter plots of the changes in the fitted values and the observed values of the endogenous variables from 1975 (\(t = 1\)) to 2015 (\(t = 5\)). The estimation result of \(\alpha > 0\) in Table 4 is used to calculate the fitted values shown in Figure 4. The left plot is about \(\ln(\lambda_{r,5}/\lambda_{r,1})\) and the right plot is about \(\ln(w_{r,5}/w_{r,1})\). \(\lambda_{r,t}\) and \(\hat{w}_{r,t}\) are normalized so that their averages across the regions are unity for each period. Again, it appears that there is only a single equilibrium for each period under the estimated parameter values of Table 4.

---

*The process I follow here can be clarified with the following example. The standard linear equations model with individual effects is written as follows.

\[
y_{it} = x_{it}'\beta + a_i + \epsilon_{it}
\]

Here, subscript \(i\) denotes the \(i\)-th observation. Subscript \(t\) denotes the observation at period \(t\). \(y_{it}\) is a dependent variable. \(x_{it}\) is a vector of independent variables. \(\beta\) is a vector of coefficients. \(a_i\) denotes the individual effect of the \(i\)-th observation. \(\epsilon_{it}\) is a temporal error term. Random effects estimation assumes that a single regression line cannot explain the observed data well. It assumes that regression lines should be differentiated for each observation with unique intercepts (\(a_i\)). Therefore, the fitness of the abovementioned linear model cannot be checked with the condition that \(a_i\) is common to all the observations.
TABLE 4 Estimation result with panel data of Japan from 1975 to 2015

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.2705***</td>
</tr>
<tr>
<td></td>
<td>(0.0196)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.3289***</td>
</tr>
<tr>
<td></td>
<td>(0.0092)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.4006***</td>
</tr>
<tr>
<td></td>
<td>(0.0147)</td>
</tr>
<tr>
<td>$\phi_A$</td>
<td>1.2453***</td>
</tr>
<tr>
<td></td>
<td>(0.0365)</td>
</tr>
<tr>
<td>$\phi_M$</td>
<td>0.3788***</td>
</tr>
<tr>
<td></td>
<td>(0.0424)</td>
</tr>
<tr>
<td>$\sigma_M$</td>
<td>5.9006***</td>
</tr>
<tr>
<td></td>
<td>(0.6576)</td>
</tr>
<tr>
<td>$\rho_\Lambda$</td>
<td>0.9002***</td>
</tr>
<tr>
<td></td>
<td>(0.0133)</td>
</tr>
<tr>
<td>$\zeta_\Lambda$</td>
<td>5.0427***</td>
</tr>
<tr>
<td></td>
<td>(0.2664)</td>
</tr>
<tr>
<td>$s_\Lambda$</td>
<td>0.5201***</td>
</tr>
<tr>
<td></td>
<td>(0.0108)</td>
</tr>
<tr>
<td>$s_{\Lambda var}$</td>
<td>0.1616***</td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
</tr>
<tr>
<td>$\rho_W$</td>
<td>0.8342***</td>
</tr>
<tr>
<td></td>
<td>(0.0216)</td>
</tr>
<tr>
<td>$\zeta_W$</td>
<td>5.5160***</td>
</tr>
<tr>
<td></td>
<td>(0.4562)</td>
</tr>
<tr>
<td>$s_W$</td>
<td>0.0795***</td>
</tr>
<tr>
<td></td>
<td>(0.0025)</td>
</tr>
<tr>
<td>$s_{W var}$</td>
<td>0.0694***</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
</tr>
</tbody>
</table>

LLF 66617.00

Full information maximum likelihood estimates. Standard errors from the negative Hessian matrix of the log-likelihood function are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

The fitness of the model is not good, especially for the changes in income. The correlation coefficient between the fitted values and the observed values of $\ln(w_{r,5}/w_{r,1})$ is 0.062. The reason for the bad fitness to income data is the weak correlation between population growth and income growth in the observed data. The correlation coefficients between the observed values of $\ln(\lambda_{r,t+1}/\lambda_{r,t})$ and $\ln(w_{r,t+1}/w_{r,t})$ are 0.524 at $t = 1$ (1975), 0.031 at $t = 2$ (1985), −0.097 at $t = 3$ (1995), and 0.304 at $t = 4$ (2005). Meanwhile, there is moderate correlation between the population and income in the observed data. The correlation coefficients between the observed values of $\ln \lambda_{r,t}$ and $\ln w_{r,t}$ are 0.499 at $t = 1$, 0.534 at $t = 2$, 0.501 at $t = 3$, 0.515 at $t = 4$, and 0.591 at $t = 5$. It is difficult for the spatial model to rationalize both static correlation and growth correlation simultaneously. The compromise of FIML is to reduce the temporal changes in the fitted values of the endogenous variables. Then, the forecast errors in the growth of the endogenous variables are lowered while the moderate correlation between the population and income is reproduced. For this reason, the vertical axes and the horizontal axes of the graphs in Figure 4 have different scales. The correlation coefficients between the fitted values of $\ln \hat{\lambda}_{r,t}$ and $\ln \hat{w}_{r,t}$ are 0.545 at $t = 1$, 0.549 at $t = 2$, 0.554 at $t = 3$, 0.554 at $t = 4$, and 0.556 at $t = 5$. If it is necessary to improve the fitness of the model further, the model should be able to deal with the observed small (or negative) correlation between $\ln(\lambda_{r,t+1}/\lambda_{r,t})$ and $\ln(w_{r,t+1}/w_{r,t})$ at $t = 2$ and 3. One way to do so may be to
allow $\sigma_M$ to vary temporally, since it determines the size of correlation between population and income. Therefore, if the values of $\sigma_M$ are differentiated according to the periods, the fitness is expected to be improved.

Although the fitness of the model to the changes in population is not good in terms of magnitude, the model roughly reproduces the trend of the changes. The correlation coefficients between the fitted values and the observed values of $\ln(\lambda_{r,5}/\lambda_{r,1})$ are 0.418. Note that I consider only the changes in the travel time on the road network in the analysis. Of course, they are poor approximations of the actual changes in the transportation cost, since the cost also depends on other transportation infrastructure, like rail, ports, and airports. Therefore, the rough fitness to the changes in population is not surprising. Nevertheless, the model can explain some portion of the variance of $\ln(\lambda_{r,5}/\lambda_{r,1})$. The variance of the observed values of $\ln(\lambda_{r,5}/\lambda_{r,1})$ is 0.166. Meanwhile, the variance of the prediction errors of $\ln(\lambda_{r,5}/\lambda_{r,1}) - \ln(\lambda_{r,5}/\lambda_{r,1})$ is 0.153. This implies that $1 - 0.153/0.162 = 8.2\%$ of the observed variance can be explained\(^9\) by the road development from 1975 to 2015.

The result shown by the scatter plots about population can be confirmed by Figures 5 and 6. These figures show three maps of the 1,376 municipalities considered in the analysis. The municipalities are classified by color according to the value of $\ln(\lambda_{r,5}/\lambda_{r,1})$. The municipalities indicated by darker red have higher population growth rate from 1975 to 2015 while the municipalities indicated by darker blue have lower (negative) population growth rate. The map of Figure 5 is depicted with the observed values and the maps of Figure 6 are depicted with the fitted values under the estimation results in Table 4. Note that each map uses a unique scale to classify the municipalities, and hence, the values indicated by specific color differ for each map. The maps can be used to grasp the correlation between the observed values and the fitted values. The map of Figure 5 confirms that the population has concentrated in large metropolitan areas, such as Tokyo, Osaka, Nagoya, Sendai, Hiroshima, and Fukuoka. This observed trend is roughly reproduced in the map of Figure 6.

5 | CONCLUSIONS

This study proposed a novel method to estimate the parameters in quantitative spatial economics model with FIML. The method can be applied to general static quantitative models. The method has various advantages, such as its relation to stability analysis. The results of the application confirmed the validity of the proposed method.

ACKNOWLEDGMENTS

I express my gratitude to Dr. Yuki Takayama for his insightful comments on this research, which helped me to enrich the contents of this paper, especially regarding global dispersion force and stability analysis. I also thank the Ministry of Land, Infrastructure, Transport and Tourism of Japan for lending the software “NITAS,” which I used to collect the historical data of travel time between municipalities. I would like to thank Editage (www.editage.com) for English language editing.

APPENDIX

A PROOF OF EQUATION (58)

For notational simplicity, I define a new variable $u_{it}$ as follows.

$$u_{it} = \frac{\partial v_i}{\partial \lambda_i} - \frac{\partial v_i}{\partial \lambda_N}$$  \hspace{1cm} (A1)

\(^9\)In addition, the road network data in NITAS do not cover the changes in the local roads. The data cover only the changes in highways. For local roads, the network data in 2015 are also used in the network data of the past periods.

\(^{10}\)In a linear regression model expressed as $y = \beta x + \epsilon$, the variance explained by $x$ is calculated as the error variance of $\epsilon$ subtracted from the total variance of $y$.

\(^{11}\)I classify the municipalities into 32 classes so that each class contains almost the same number of municipalities.
With this variable, $J$ and $J'_{J/v}$ are expressed as follows.

\[
(J)_{i,i'} = \frac{1}{N} \left[ -\sum_{i''=1}^{i-1} u_{i'i''} + (N-1)u_{i'i'} - \sum_{i''=i+1}^{N} u_{i'i''} \right] \tag{A2}
\]

\[
(J'_{J/v})_{1,i'} = u_{N1i'} - u_{i'i'} \tag{A3}
\]

I apply row elementary operations to $J'_{J/v}$ to convert it into $J$. I focus on the first column of the matrix, because the same operations are applied to all the columns. I omit the column index $i'$ for the same reason. During the proof, I use equations in
the following style.

\[ \det J'_J = \det \begin{bmatrix} u_N - u_1 & \cdots \\ \vdots & \ddots \\ u_N - u_{N-1} & \cdots \end{bmatrix} \] (A4)

First, I multiply all the rows by \(-1\), then multiply the first row by \(N - 1\).

\[ (-1)^{N-1}(N - 1) \det J'_J = \det \begin{bmatrix} (N - 1)u_1 - (N - 1)u_N & \cdots \\ u_2 - u_N & \cdots \\ \vdots & \ddots \\ u_{N-1} - u_N & \cdots \end{bmatrix} \] (A5)

Next, I subtract all the rows second and upward from the first row.

\[ (-1)^{N-1}(N - 1) \det J'_J = \det \begin{bmatrix} (N - 1)u_1 - \sum_{i'=2}^{N-1} u_{i'} - u_N & \cdots \\ u_2 - u_N & \cdots \\ \vdots & \ddots \\ u_{N-1} - u_N & \cdots \end{bmatrix} \] (A6)

From this point, I repeatedly apply the same operations to all the rows second and upward in ascending order. I call the step on which I apply the operations to the \(i\)-th row the \(i\)-th step. Suppose that at the beginning of the \(i\)-th step, the equation is expressed as follows.

\[ (-1)^{N-1} N^{i-2}[N - (i - 1)] \det J'_J = \det \begin{bmatrix} (N - 1)u_1 - \sum_{i'=2}^{N-1} u_{i'} - u_N & \cdots \\ \vdots & \ddots \\ -\sum_{i'=1}^{i-2} u_{i'} + (N - 1)u_{i-1} - \sum_{i'=1}^{N-1} u_{i'} - u_N & \cdots \\ u_i - u_N & \cdots \\ \vdots & \ddots \\ u_{N-1} - u_N & \cdots \end{bmatrix} \] (A7)

Note that this applies to the second step \((i = 2)\) from equation (A6). I add the rows from first to \((i - 1)\)-th multiplied by \(-a_i\) to the \(i\)-th row. Then, I add the rows from \((i + 1)\)-th to \((N - 1)\)-th multiplied by \(-N a_i\) to the \(i\)-th row. Here, \(a_i\) is a constant expressed as follows.

\[ a_i = \frac{1}{(N - 1)[N - (i - 1)] - (i - 1)} = \frac{1}{N(N - i)} \] (A8)

After the operations, the determinant is unchanged and the \(i\)-th row is expressed as follows.

\[ \frac{-1}{(N - 1)[N - (i - 1)] - (i - 1)} \sum_{i'=1}^{N-1} u_{i'} + \frac{(N - 1)[N - (i - 1)]}{(N - 1)[N - (i - 1)] - (i - 1)} u_i \]

\[ = \frac{N - (i - 1)}{N(N - i)} \left[ -\sum_{i'=1}^{i-1} u_{i'} + (N - 1)u_i - \sum_{i'=i+1}^{N-1} u_{i'} - u_N \right] \] (A9)

Then, I divide the \(i\)-th row by \([N - (i - 1)]a_i\) to complete the \(i\)-th step.

\[ (-1)^{N-1} N^{i-1}[N - i] \det J'_J = \det \begin{bmatrix} (N - 1)u_1 - \sum_{i'=2}^{N-1} u_{i'} - u_N & \cdots \\ \vdots & \ddots \\ -\sum_{i'=1}^{i-1} u_{i'} + (N - 1)u_i - \sum_{i'=i+1}^{N-1} u_{i'} - u_N & \cdots \\ u_i - u_N & \cdots \\ \vdots & \ddots \\ u_{N-1} - u_N & \cdots \end{bmatrix} \] (A10)
Here, I confirm that equation (A7) holds at the beginning of the \((i + 1)\)-th step. Repeating the abovementioned operations, I derive the following equation.

\[
(-1)^{N-1} N^{N-2} \det J^p = \begin{vmatrix}
(N - 1)u_1 - \sum_{\rho=2}^{N-1} u_\rho - u_N & \cdots \\
\vdots & \ddots & \ddots \\
-\sum_{\rho=1}^{N-2} u_\rho + (N - 1)u_{N-1} - u_N & \cdots
\end{vmatrix}
\]

(A11)

Finally, I divide all \(N - 1\) rows by \(N\) to derive equation (58).

\[
(-1)^{N-1} \frac{1}{N} \det J^p = \frac{1}{N} \begin{vmatrix}
(N - 1)u_1 - \sum_{\rho=2}^{N-1} u_\rho - u_N & \cdots \\
\vdots & \ddots & \ddots \\
-\sum_{\rho=1}^{N-2} u_\rho + (N - 1)u_{N-1} - u_N & \cdots
\end{vmatrix} = \det J
\]

(A12)

References


