PRI Discussion Paper Series (No.21A-3)

Population Mobility Structural Analysis and Population Estimation Using a Quantitative Spatial Model

Economist, Policy Research Institute, Ministry of Finance

Wataru Takahashi

May 2021

Research Department Policy Research Institute, MOF

3-1-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-8940, Japan

TEL 03-3581-4111

Population Mobility Structural Analysis and Population Estimation Using a Quantitative Spatial Model

Wataru Takahashi[1,2]
Policy Research Institute, Tokyo, Japan.

This article proposes a population forecasting using endogenous population model, which incorporates a spatial model considering the spatially heterogeneous feature of agents and the economy by employing quantitative spatial models. In this endogenous population model, agents migrate to maximize their utility. This model is estimated using the two-stage estimation approach, which is extensively used in quantitative spatial literature. Estimated parameters are significant and almost consistent with the economic and demographic stylized facts.

Using the parameters concerning migration and local utilities, we conduct projection analyses for 2015-2125 across all prefectures in Japan, which is now experiencing regionally asymmetric population decline. In the baseline projections, the population in less populated prefectures is mitigated slightly by introducing the young generation's migration behavior. Counterfactual analyses are then conducted to break down the factors of population decline in Japan. Among several factors, birth abandonments due to some constraints and slow productivity growth after 1995 turned out to have severely impacted demographics. The development of networks resulted in having negative impacts on demographics though they had positive impacts on the welfare of economic agents in many aspects.

Keywords: endogenous population model, spatial economics, quantitative spatial economics, population mobility.

JEL classification: C21, J61, R10, R12, R23 P25.

---

[1] Office of Econometric Analysis for Fiscal and Economic Policy, Research and Administration Department, Policy Research Institute, Ministry of Finance, 3-1-1, Kasumigaseki, Chiyoda-Ku, Tokyo.
E-mail: wataru.takahashi@mof.go.jp.

[2] The author is grateful to Prof. Motohiro Sato at Hitotsubashi University for his guidance and advice. The author also would like to thank Prof. Shinji Yamashige for his advice on prior research. Many feedbacks in Economic Theory and Policy Workshop(Aoyama Gakuin University),2019 Japanese Economic Association Spring Meeting(Musashi University), Research in the Frontiers of Modern Economics Workshops(Aoyama Gakuin University), are also helpful. We also received valuable comments from professor Hirokuni Iiboshi (Tokyo Metropolitan University), professor Hisakazu Kato (Meiji University), associate professor Masaki Nakahigashi (Niigata University) and professor Shinichi Nishiyama (Kyoto University).The opinions in the article are solely of the author and do not represent the official views of the institution to which he belongs. This work was supported by KAKENHI 19160971.(Grant-in-Aid for Young Scientists B).

1. Introduction

It has been a long time since Japan became an aging society with a declining birthrate. The population of Japan began to decline in 2015. Economics tells us that a change in demographic structure will have a major impact on the economy and public finance. In particular, in finance, the multigenerational overlapping generations model developed by Auerbach and Kotlikoff (1987) provides a detailed analytical framework to examine the impact of changes in demographic structure, which has several applications in Japan.[3]

However, the issue of sustainability has significant implications not only for the nation as a whole, but also for regions. In 2007, the city of Yūbari in Hokkaido declared financial bankruptcy and was designated by the government as a financial reconstruction organization, an event that is still fresh in our memories. After the designation, Yūbari's expenditure for public services was considerably restrained and the welfare of its inhabitants worsened. Thus, the lack of sustainability of local governments and economies contributes to regional inequality. In addition, lack of sustainability leads to inequality between residents with different characteristics, as it is known that inhabitant traits affect interregional migration and the attributes of local inhabitants (Moretti, 2011). Therefore, regional populations and economies should be analyzed under a framework similar to that of macro-models.

Three factors should be considered in assessing sustainability in regions in this way. First, it is necessary to consider the migration of economic agents. This is because interregional migration is more intense than international migration. Second, as new economic geography[4] points out, different regions in the same country have different economic structures. The heterogeneity of these regional economies as well as the of demographic structures should be taken into account. Third, the interrelationship between regional populations and regional economies should be formulated. For example, De la Croix and Gobbi (2017) argues that population density adversely affects fertility. In addition, some studies regard population concentration as agglomeration and analyze its relationship with economic activity. Based on these studies, it is clear that there is a complex relationship between current population, fertility rate, and economic activity in all regions.

Several studies have addressed regional economies in the study of sustainability using an overlapping generations model. Börsch-Supan, Ludwig, and Winter (2006) analyzed the impact of pension reforms in the three countries (Germany, France, and Italy) using the multiregional overlapping generations model of INGENUE (2001). Fougerè et al. (2009) quantitatively assessed how immigrant acceptance and educational spending affects the financial condition of Canadian states and territories. However, these studies assume that the region is connected to larger regional units (the

---

[3] Braun and Joines (2015) is a typical example of a study in this area. The conclusion is that Japan's fiscal sustainability cannot be achieved without fundamental structural reform, but it is also characterized by strong assumptions about the discounted present value of public debt.

[4] An early study is Krugman (1999). Comparatively recent theoretical developments include Fujita et al. (1999) and Baldwin et al. (2003).

world and states) via capital and financial markets, and that equilibrium is achieved through interregional capital transfers and fund transfers induced by interest rates that vary regionally. In these studies, the immobility of economic agents is assumed, or the movement of people restored from demographic statistics is merely given exogenously. That is, although these previous studies consider the heterogeneity of regional economies and demographic composition (the second factor), they do not consider the endogenous migration of people (the first factor) or interactions between regional populations and regional economies (the third factor).

The contribution of this paper is primarily related to the above-mentioned "first factor" and "third factor." In this paper, we analyze a model combining the discrete selection model by McFadden (1974) with a simple household utility maximization problem that is endogenous to the population. This analytical method is in line with recent progress in quantitative spatial modeling (Redding and Rossi-Hansberg, 2017). For this model, the indirect utility of inhabiting each prefecture in Japan at a particular point in time is estimated by performing a guided analysis using panel data on two types of inter-prefectural movement and traffic costs generated from a detailed traffic network. In addition, parameters in the household model are estimated by structural estimation. These results are mostly consistent with the intuition of the overlapping generations model.

This paper is structured as follows. In Chapter 2, two relevant research areas are surveyed. Chapter 3 formulates a theoretical model based on the above survey. Chapter 4 describes the estimation policies of the formulated models and the estimation methods at each stage and describes the estimation results. Chapter 5 conducts a counterfactual analysis of projections and populations based on the projected total population and population by prefecture based on the various estimated parameters. Finally, Chapter 6 describes conclusions and future policies.

2. Previous research

This chapter describes previous studies of spatial equilibrium and endogenous population models. The first part discusses the spatial equilibrium model often used in evaluating regional economic activity and local government policies and the quantitative spatial model developed in recent years as an extension of it. The second part introduces examples of applying the endogenous population theory model, including an overlapping generations model, to multiple regions.

The purpose of the spatial equilibrium model is to derive the economic value of factors specific to a region (regional characteristics) through selection of economic agents capable of moving between regions. Early studies by Rosen (1979) and Roback (1982) formulated economic agents to obtain utility from consumer goods circulated between regions, (immobile) consumer goods fixed to regions, and factors specific to regions. In this spatial equilibrium, the utility of economic agents in all regions is equalized by migration; furthermore, wages and non-migratory consumer goods prices are adjusted to enable the derivation of the value of region-specific factors. Moretti (2011) and Diamond (2016) discuss regional productivity using different types of regional characteristics, with the latter particularly identifying differences in regional characteristics between cities in addition to wage disparities. The spatial equilibrium model has also been applied in the field of public economics; as Diamond (2017) pointed out, the taxable rent of local governments depends on regional characteristics as well as the elasticity of housing supply. In addition, Brueckner and Neumark (2014) showed that economic agents' preferences for desirable regional characteristics (proximity to the ocean, long daylight hours) hampered "voting with their feet" and resulted in higher public sector wages.

The spatial equilibrium framework used in these studies consists of formulating relatively simple equilibrium conditions and empirically evaluating policies using these formulated conditions. In contrast, the quantitative spatial model used in this paper performs counterfactual hypothetical analysis and simulation using structural estimation. Redding and Rossi-Hansberg (2017), who conducted a survey of this study, show the applications of quantitative spatial modeling to various fields. The commonality of these studies is that economic variables are stochastically formulated and follow an extreme value distribution.[5] Introducing such a probability distribution eliminated the complexity of the equilibrium condition of the new economic geography and facilitated multiregional and multinational analyses.

Bryan and Morten (2015) and Morten and Oliveira (2016) can be cited as remarkable prior studies in similar fields as this paper. Both studies analyzed the interrelationship between economic development and labor migration by formulating the choice of employment area for economic agents. These studies feature a detailed analysis of migration costs. Bryan and Morten (2015) estimate the migration costs of Indonesian regions by individual records and examine the impact of migration costs on economic development through a comparison with the United States. Morten and Oliveira (2016) use Brazilian traffic network statistics to derive migration costs through the application of the fast-marching method (FMM) by Allen and Arkolakis (2014). Since endogeneity is assumed for economic development and the development of traffic networks, natural experiments are being carried out using data before and after the reorganization of traffic networks accompanying the construction of the new capital. As will be explained later, detailed time-series data on traffic networks are available in Japan. Therefore, a panel analysis applying Morten and Oliveira's (2016) method using traffic network as an operational variable will be conducted.

In order to analyze the population structure cited as a problem in the introduction, this paper introduces a negative correlation between birthrate and population density, in addition to the relationship between population migration and economic development. This negative correlation has been pointed out by De la Croix and Gobbi (2017) on a global scale, and Kato (2017) reported that a

[5] For example, in the new economic geographic model, the formulation of international trade using the Fréchet distribution by Eaton and Kortum (2002) is notable. On the other hand, in quantitative spatial analysis in recent years, McFadden (1974) has carried out analysis using the standard Gumbel distribution (Weibull distribution), and the latter framework is also followed in this paper, as will be described later.

strong correlation was observed in Japan. In addition, since Japan's birthrate is declining and its population is aging, a multiregional endogenous population model that formulates this correlation is presented. Sato and Yamamoto (2005) conducted an early study in this area that formulates demographic structures in economies with agglomeration effects on wages and congestion effects on child-rearing costs using representative household models in two regions.

In the spatial equilibrium model introduced at the beginning of this chapter, equilibrium is achieved by concordance of utility in all regions. Therefore, it is possible to formulate a regional endogenous population model such as the one proposed by Sato and Yamamoto (2005) that describes household behavior by maximizing utility under a spatial equilibrium model, and from which an endogenous population-space equilibrium model is formulated. Sato (2007) formulated a birth-endogenous multigenerational overlapping generations model based on a spatial equilibrium model. In the study, consumption and child rearing of representative households in the region reflect additional congestion-effective additional costs, and productivity of representative enterprises in the region reflects agglomeration effects. Therefore, the relationship between the regional population and birthrates considered in this study depends on the relative strengths of the agglomerated economy and the dissipated diseconomy. If the productivity improvement by economies of scale is sufficiently large in comparison with the child-rearing cost of the region, the population is concentrated in the region with large scale, and the population density positively affects the scale of wages and population inflow, while the birthrate is negatively affected.

Ishida, Oguro, and Yasuoka (2018) formulated an overlapping generations model under a spatial equilibrium model. A two-region model is formulated here, but the formalized household behavior is more detailed. The policy analysis of public childcare policy formulates congestion effects associated with land availability as well as housing costs, educational investment, and child-rearing costs. The analysis shows that increasing population density negatively affects birthrate through congestion effects in child-rearing costs. Here, the local government's public childcare policy obtains the result of reducing the negative effect on birth rate, while accelerating population inflow through the improvement of local welfare. Therefore, this paper aims to formulate a multiregional endogenous population model within the framework of quantitative spatial analysis, referring to Sato (2007) and Ishida, Oguro, and Yasuoka (2018), and to gain an understanding of demographic structure and economic households in quantitative terms.

3. Theoretical model

In this chapter, we formulate a multiregional population endogenous model that deals with household birth behavior and migration. In choosing a residential area, we adopt Morten and Oliveira's (2016) formula based on McFadden's (1974) conditional logit model. This study formulates individual preferences for each region by extreme value distributions and analyzes utility-based migration quantitatively. In contrast, Sato (2007) and Ishida, Oguro, and Yasuoka (2018) theoretically derive a negative relationship between regional birthrates and populations by formulating an overlapping generations model under a spatial equilibrium model. Based on these studies, this paper aims to gain a quantitative understanding of the sustainability of a region by estimating the parameters of population migration and population structure.

Formulation of population migration First, in accordance with Morten and Oliveira (2016), a conditional logit model as a discrete selection as shown by McFadden (1974) is used to formulate the regional utility ν_{jt} at time t in region j and the utility ν_{ljt} of an individual l residing in the region at that time. Here, individual utility ν_{ljt} is described by:

$$\nu_{ljt} = \omega + \nu_{jt} + \varepsilon_{ljt}. \tag{1}$$

ε_{ljt} may be regarded as the fitness[6] of individual l for region j. Before choosing a residential area, the fitness of an individual for the entire area $j \in M$ including their place of birth $\{\varepsilon_{l1t}, \cdots, \varepsilon_{ljt}, \cdots, \varepsilon_{lMt}\}$ is observed.

Individuals choose their areas of residence after birth and before engaging in economic activities in specific areas. The residential area selection problem is formulated as follows:

$$\max_j \{\nu_{jt} + \varepsilon_{ljt} - c_{ijt}\}. \tag{2}$$

Equation (2) determines whether an individual l who resides in region j at the beginning of time t stays in region j or migrates to another region. If individual l who was born in region i migrates from their place of birth, they must pay migration costs c_{ijt} to initiate economic activity in region j.

In solving the maximization problem, it is assumed that the fitness ε_{ljt} of individual l for region j follows a standard type I extreme distribution (standard Gumbel distribution). That is, a random variable ε_{ljt} follows a distribution defined by a probability density function $f(\varepsilon_{ljt})$ and a cumulative density function $F(\varepsilon_{ljt})$.

[6] Based on interpretations of Moretti (2011), this term can also be interpreted as an additional preference by the individual for a region.

$$f(\varepsilon_{ljt}) = \exp(-\varepsilon_{ljt})\exp\{\exp(-\varepsilon_{ljt})\}.$$
$$F(\varepsilon_{ljt}) = \exp\{\exp(-\varepsilon_{ljt})\}.$$

This assumption can be regarded as equal to an idiosyncratic term in the utility of an individual in McFadden's (1974) formulation, and the corresponding indirect utility ν_{ljt} of individual l as adding idiosyncratic preferences (fitness) ε_{ljt} to the population expected utility ν_{jt} of region j. Therefore, the model is consistent with the discrete selection problem from with the conditional logit model.

If we follow McFadden (1974), the probability π_{ijt} that an economic entity residing in region i chooses region j can be defined as follows:[7]

$$\pi_{ijt} = \frac{\exp(\nu_{jt} - c_{ijt})}{\sum_{m=1}^{M} \exp(\nu_{mt} - c_{imt})}. \tag{3}$$

Since $\sum_{m=1}^{M} \pi_{imt} = 1$ is fulfilled from the definition of π_{ijt}, the probability of migration π_{ijt} satisfies the assumption of a random variable. In addition, due to the existence of migration costs c_{ijt}, regional indirect utility ν_{jt} differs between places of birth, unlike conventional spatial equilibrium formulations.[8]

In addition, the regional labor force L_{jt} at time t is determined depending on the individual's regional distribution at the start of time t, since the migration of population from region i to region j is directly governed by indirect utility. This regional distribution is equal to the young population at the end of last year and can be defined as $\{N_{y,1t}, \cdots, N_{y,Mt}\}$ using the young population $N_{y,jt}$ in each region $j \in M$.[9] Therefore, the entry of the labor force L_{jt}^{+} into region j is given by Equation (4) below.

$$L_{jt}^{+} = \sum_{m=1}^{M} \pi_{mjt} N_{y,mt}. \tag{4}$$

Labor demand A single consumer good is produced in region j. Technologies of typical firms in region j are expressed by the Cobb-Douglas production function.

---

7 See Appendix 1.2 for detailed derivation assumptions.

8 In the spatial equilibrium model, migration occurs until the indirect utility of all regions is consistent. On the other hand, in the quantitative spatial model by Morten and Oliveira (2016), indirect utility among regions does not necessarily coincide because of the constant movement of populations according to fitness ε_{ljt} and the existence of travel costs.

9 Since demographic dynamics are analyzed in Chapter 5, the various demographic indicators should be taken as given here.

$$Y_{jt} = A_{jt} L_{jt}^{\kappa} K_{jt}^{1-\kappa}. \tag{5}$$

Here, A_{jt} is the productivity at time t of the typical firms in region j, and it is assumed that company revenue belongs to an absentee company head and that the indirect utility of the individual is not affected.

Capital markets are integrated across regions, and regional labor markets are perfectly competitive. Therefore, regional wages w_{jt} and capital rental costs r_t can be defined as follows. The subscript j is omitted because the rental costs are shared by the entire country.

$$\log w_{jt} = \log A_{jt} + \log \kappa + (1-\kappa) \log \frac{K_{jt}}{L_{jt}}, \tag{6}$$

$$\log r_t = \log A_{jt} + \log(1-\kappa) - \kappa \log \frac{K_{jt}}{L_{jt}}. \tag{7}$$

Equations (6) and (7) can be combined to obtain Equation (8) corresponding to the labor demand curve in region j. κ_t is a function of capital rental costs r_t and can be regarded as a national trend.

$$\log w_{jt} = \kappa_t + \frac{1}{\kappa} \log A_{jt}, \qquad where\ \kappa_t = \log \kappa + \frac{1-\kappa}{\kappa} \log(1-\kappa) - \frac{1-\kappa}{\kappa} \log r_t. \tag{8}$$

In accordance with general assumptions made in urban economics, all real estate in region j belongs to absent landlords, and the land and other rent shall not affect the individual's utility function.

Housing market According to studies of spatial equilibrium analyses of population migration, such as Moretti (2011) and Diamond (2017), housing costs h_{jt} at time t shall be defined by the trend term η_t and the reciprocal elasticity of housing supply to residential populations η^h. Here we assume that the working population L_{jt} of region j is entirely resident in the region[10] and that the resident population in the region is consistent with the working population. Consequently, the housing market in region j is given as follows:

$$\log h_{jt} = \eta_t + \eta^h \log L_{jt}. \tag{9}$$

Formulation of utility The endogenization of the population makes household utility somewhat more complicated than in Morten and Oliveira (2016). Individuals who migrate to region j at the beginning of time t begin economic activity, earning income w_{jt} in exchange for labor, and

[10] As will be explained later, this process is somewhat valid because the "region" in this paper is a prefecture.

distributing income to consumption and child rearing. Let C_{jt} be the individual's consumption of consumer goods and let them raise n_{jt} children. In addition, it is assumed that each individual gains utility through the regional characteristics X_{jt}. Thus, the expected utility of a representative individual at time t in region j is defined as follows:

$$U_{jt} = C_{jt}^{\alpha} n_{jt}^{\beta} H_{jt}^{1-\alpha-\beta} \cdot \exp\left(X_{jt}^{\gamma}\right). \tag{10}$$

In addition, the budget constraint of this representative individual is defined as follows. In addition to w_{jt} referring to regional wages, D_{jt} and h_{jt} refer respectively to regional population density and housing costs in region j. Based on demographic assumptions and the results of De la Croix and Gobbi (2017), child-rearing costs $f(D_{jt})$ increase as the population density of the region D_{jt} increases.

$$w_{jt} - h_{jt} H_{jt} = f(D_{jt}) n_{jt} + C_{jt}. \tag{11}$$

Solving the optimization problem defined by utility (10) and budget constraints (11), we can derive the conditions represented by Equations (12)–(14). These show the level of demand for consumer goods C_{jt}, regional birthrate n_{jt}, and regional housing demand H_{jt} in region j at the time of utility maximization of the representative individual.

$$C_{jt} = \alpha w_{jt} \tag{12}$$

$$n_{jt} = \beta \frac{1}{f\left(D_{jt}\right)} w_{jt} \tag{13}$$

$$H_{jt} = (1 - \alpha - \beta) \frac{w_{jt}}{h_{jt}} \tag{14}$$

The indirect utility function V_{jt} can be derived by substituting Equations (12)–(14) into the utility function (10):

$$V_{jt} = \alpha^{\alpha} \beta^{\beta} (1 - \alpha - \beta)^{1-\alpha-\beta} w_{jt} \left\{f\left(D_{jt}\right)\right\}^{-\beta} h_{jt}^{-(1-\alpha-\beta)} \cdot \exp\left(X_{jt}^{\gamma}\right) \tag{15}$$

The indirect utility of the representative individual refers to the expected utility level for choosing region j at the beginning of time t. In addition, the logarithm of Equation (15) gives the logarithmic indirect utility function v_{jt} of region j at time t.

$$v_{jt} = \log\{\alpha^{\alpha}\beta^{\beta}(1-\alpha-\beta)^{1-\alpha-\beta}\} + \log w_{jt}$$
$$-\beta \log\{f(w_{jt})\} - (1-\alpha-\beta)\log h_{jt} + \gamma X_{jt}. \qquad (16)$$

Based on the fact that v_{jt} is the expected utility of moving to region j at time t, the following Equation (17) can be obtained by summarizing and simplifying the equation.

$$v_{jt} = \omega + \nu_{jt}, \qquad where\ \ \omega = \log\{\alpha^{\alpha}\beta^{\beta}(1-\alpha-\beta)^{1-\alpha-\beta}\}. \qquad (17)$$

The constant term ω is common among regions, and ν_{jt} is a time-region idiosyncratic term. ν_{jt} is given by Equation (18).

$$\nu_{jt} = \log w_{jt} - \beta \log\{f(D_{it})\} - (1-\alpha-\beta)\log h_{jt} + \gamma X_{jt}. \qquad (18)$$

Formulation of individual behavior The time-region idiosyncratic term ν_{jt} resulting from the utility maximization problem discussed above corresponds to the regional utility of migration as defined at the beginning of this chapter. In other words, it is possible to show the utility v_{ljt} of residing in region j at time t for an individual l as follows:

$$\begin{aligned} v_{ljt} &= \omega + \nu_{jt} + \varepsilon_{ljt} \\ &= \omega + \log w_{jt} - \beta \log\{f(D_{jt})\} - (1-\alpha-\beta)\log h_{jt} + \gamma X_{jt} + \varepsilon_{ljt}. \end{aligned} \qquad (19)$$

As initially defined, ε_{ljt} is the fitness of the individual l to region j at time t and follows the type I extreme value distribution. The problem of region selection performed by the individual l at the beginning of time t is defined in Equation (2), and the probability of migration of an individual from region i to region j (the migration probability) is derived through this maximization problem.

$$\begin{aligned} \pi_{ijt} &= \frac{\exp(\nu_{jt} - c_{ijt})}{\sum_{m=1}^{M}\exp(\nu_{mt} - c_{imt})}, \\ &= \frac{\exp(\log w_{jt} - \beta \log\{f(D_{jt})\} - (1-\alpha-\beta)\log h_{jt} + \gamma X_{jt} - c_{ijt})}{\sum_{m=1}^{M}\exp(\log w_{mt} - \beta \log\{f(D_{mt})\} - (1-\alpha-\beta)\log h_{mt} + \gamma X_{mt} - c_{imt})} \end{aligned} \qquad (20)$$

Equation (20) is the migration probability π_{ijt} corresponding to the two-stage optimization problem of the regional selection problem and the utility maximization problem by individuals. Here, the probability that an individual residing in region i will move to region j at time t is described using the economic variables of each region $j \in M$ and costs of migrating c_{ijt} from region i to region j.

Spatial equilibrium In this chapter, quantitative spatial modeling is formulated in accordance with Morten and Oliveira (2016) and others. It is possible to formulate the spatial equilibrium by integrating the discussions so far. The equilibrium conditions consist of four equations. The regional equilibrium wage $\{w^*_{it}\}_{i\in M}$, regional equilibrium migration probability $\{\pi^*_{ijt}\}_{i,j\in M}$ from region i to region j, regional equilibrium labor force $\{L^{+*}_{it}\}_{i\in M}$, and regional equilibrium housing price $\{h^*_{it}\}_{i\in M}$ are determined for a given regional productivity $\{A_{it}\}_{i\in M}$, young population at the start of the period $\{N_{y,it}\}_{i\in M}$, and regional characteristics $\{X_{it}\}_{i\in M}$ by the equilibrium equations for them.

1. The regional labor market is in a spatial equilibrium state, and the regional equilibrium wage $\{w^*_{it}\}_{i\in M}$ for region j is given by the following equation:

$$\log w^*_{jt} = \kappa_t + \frac{1}{\kappa}\log A_{jt}\,. \tag{21}$$

2. The regional housing market is in a spatial equilibrium state, and the regional equilibrium housing price $\{h^*_{it}\}_{i\in M}$ is given by the following equation:

$$\log h^*_{jt} = \eta_t + \eta^h \log L^*_{jt} \tag{22}$$

3. The regional equilibrium migration probability $\{\pi^*_{ijt}\}_{i,j\in M}$ from region i to region j is given by the following equation:

$$\pi^*_{ijt} = \frac{\exp\left(\log w^*_{jt} - \beta\log\{f(D^*_{jt})\} - (1-\alpha-\beta)\log h^*_{jt} + \gamma X_{jt} - c_{ijt}\right)}{\sum_{m=1}^{M}\exp\left(\log w^*_{jt} - \beta\log\{f(D^*_{jt})\} - (1-\alpha-\beta)\log h^*_{jt} + \gamma X_{jt} - c_{ijt}\right)} \tag{23}$$

4. The regional equilibrium labor force entry $\{L^{+*}_{it}\}_{i\in M}$ is given by the following equation:

$$L^{+*}_{jt} = \sum_{m=1}^{M}\pi^*_{mjt}\,N_{y,mt}. \tag{24}$$

4. Estimations

In this section, based on previous studies such as Ahlfeldt et al. (2015), Diamond (2016), and Morten and Oliveira (2016), we conduct a two-stage estimation using the above spatial equilibrium conditions. In the first stage, the expected indirect utility v_{jt} in each region is derived using migration costs derived from two types of domestic migration statistics and traffic network time series data. In the second stage, we derive parameters for analysis by applying spatial equilibrium conditions to this indirect utility and by making structural estimates.

4.1 Stage 1: Induced estimation of indirect utility

In the first phase, we use migration statistics to estimate the expected indirect utility of Japan's 47 prefectures.

Estimation equation From the discussion so far, the probability π_{ijt} that an individual living in region i chooses region j as a destination is given by Equation (21):

$$\pi_{ijt} = \frac{\exp(\nu_{jt} - c_{ijt})}{\sum_{m=1}^{M} \exp(\nu_{mt} - c_{imt})}. \tag{25}$$

As can be seen from this formulation, costs of migration c_{ijt} from region i to region j have a significant impact on interregional migration. Therefore, it is necessary to formulate these migration costs c_{ijt} in the estimation.

$$c_{ijt} = \lambda^f \mathbb{I}\{j \neq i\} + \lambda^R \mathbb{I}\{j \in R\} + \lambda^t T_{ijt} \tag{26}$$

$\mathbb{I}\{j \neq i\}$ is a fixed migration cost dummy, indicating that individuals need to pay when migrating outside their home region. $\mathbb{I}\{j \in R\}$ is the same district dummy and indicates that migration between prefectures in the same district[11] is comparatively easy. T_{ijt} is the migration time between region i and region j at time t and is derived from the time series data for the traffic network by a method that will be described later. Equation (22) is substituted into Equation (21) and the logarithm of both sides is taken to obtain the following equation.

$$\log \pi_{ijt} = \nu_{jt} - \{\lambda^f \mathbb{I}\{j \neq i\} + \lambda^R \mathbb{I}\{j \in R\} + \lambda^t T_{ijt}\} - \log\left\{\sum_{m=1}^{M} \exp(\nu_{mt} - c_{imt})\right\}$$

By rewriting this equation, we obtain Equation (23), which is the first-stage estimation equation.

$$\log \pi_{ijt} = \nu_{jt} + \lambda^f \mathbb{I}\{j \neq i\} + \lambda^R \mathbb{I}\{j \in R\} + \lambda^t T_{ijt} + \Pi_{it} + \zeta_{ijt}. \tag{27}$$

Here, $\Pi_{it} = -\log\{\sum_{m=1}^{M} \exp(\nu_{mt} - c_{imt})\}$ is a fixed effect specific to the origin i and is absorbed in the fixed migration cost part in the estimation. The indirect effect $\{\nu_{jt}\}$, which is the object of the first-stage estimation, can be regarded as a fixed effect specific to the destination j at time t. Furthermore, ζ_{ijt} is the measurement error of the individual's probability of migration.

Intuitively, the sign of the coefficients is expected to be $\lambda^f < 0, \lambda^R > 0,\ \lambda^d < 0,\ \lambda^t < 0$. It is

[11] Here, we used a division that divided the 47 prefectures into eight districts. Except for the Hokkaido district, each district includes multiple prefectures. The details are as follows: 1. Hokkaido district: Hokkaido; 2. Tohoku district: Aomori, Iwate, Akita, Miyagi, Yamagata, Fukushima; 3. Kanto district: Ibaraki, Tochigi, Gunma, Saitama, Chiba, Tokyo, Kanagawa; 4. Chubu district: Yamanashi, Nagano, Niigata, Toyama, Ishikawa, Fukui, Shizuoka, Aichi, Gifu, Mie; 5. Kinki district: Shiga, Kyoto, Osaka, Hyogo, Nara, Wakayama; 6. Chugoku district: Tottori, Shimane, Okayama, Hiroshima, Yamaguchi; 7. Shikoku district: Kagawa, Ehime, Tokushima, Kochi; 8. Kyushu district: Fukuoka, Saga, Nagasaki, Kumamoto, Oita, Miyazaki, Kagoshima, Okinawa.

difficult for an individual living in a particular prefecture to migrate to another prefecture. In addition, the effort required for migration increases depending on the distance to the destination. Prefectures in the same district have similar cultural backgrounds, and it is assumed that this effort will decrease when the individual migrates within the same district.

Migration probability Migration probability π_{ijt} is calculated from two types of data on inter-prefectural migration: the "Basic Resident Register Migration Report" (Ministry of Internal Affairs and Communications) and the "National Census Migration Tabulation" (Ministry of Internal Affairs and Communications). The former has been reported every year since the 1950s but does not include the statistics for Okinawa until the second half of the 1970s. The latter has been reported once every decade since 1990[12] based on the census. The migration probability based on the Basic Resident Register Migration Report π_{ijt}^{R} is defined as the proportion of the tabulated population[13] that migrated to a given other prefecture, relative to the population of each prefecture at the start of the period of coverage every five years from 1975 to 2015. The probability of migration based on the National Census Migration Tabulation π_{ijt}^{C} was calculated by tabulating the permanent residence five years ago for each prefecture, and then calculating the ratio of the population migrating to each prefecture to the population thus tabulated. This probability of migration is derived for 1990, 2000, 2010, and 2015, when demographic migration was tabulated. However, it should be noted that the applicable period corresponding to the survey year differs due to differences in the tabulation method in the Resident Basic Register Migration Report.

The Basic Resident Register Migration Report tabulates totals for all migrations from 1975 to 2015, including those in Okinawa, and the time series for estimation is long. The National Census Migration Tabulation has a short time series and does not capture all population movements until 2010,[14] but is accurate because it is based on the National Census, which is a full census.

Travel time Here, we use the abundant time series data on the traffic network contained in the National Land Numerical Information (Ministry of Land, Infrastructure, Transport and Tourism) and two pieces of legislation on the development of the traffic network to generate panel data on travel time T_{ijt}. The data produced by the Ministry of Land, Infrastructure, Transport and Tourism contains time-series geographical data for various traffic networks including information on regular and irregular routes on Shinkansen (bullet train), trunk railways, highways, national roads, and maritime

[12] While Japan conducts a census every five years, it conducts the migration tabulation only once every decade. However, since 2010, population tabulation and migration tabulation have been implemented every five years, with the most recent being in 2015.

[13] Since statistics for each fiscal year are totaled for five years, there is a possibility that the same individual migrates a plurality of times.

[14] Population migration from 1990 to 1995 and from 2000 to 2005 is not included here.

traffic. Following Allen and Arkolakis (2014),[15] we applied the fast-marching method (FMM), a type of image analysis method, to these geographical data and calculated travel times between prefectural capitals every five years from 1975 to 2015.[16] One of the contributions of this paper is that we obtained annual data from a single source and used it to ensure consistency between the time series of the explained variable and the time series of the travel time.[17] Geographic information descriptions and methodologies are given in the appendix, and here we present the geographic information used to derive travel time for 1975 and 2015 with the aim of demonstrating intuition about the procedure.

[Figure 1]

Figure 1 shows Japan's transport networks in 1975 and 2015. Basically, dark colors are allocated to locations where there is a means of transportation with a high speed of travel. In order to guarantee the existence of travel time as a solution, a constant speed is defined for land where there is no means of transportation. However, we considered it impossible to pass through the sea. Comparing the traffic network in 1975 (Figure 1a) with that in 2015 (Figure 1b), it is apparent that the traffic network is being developed because the darkest lines, indicating the Shinkansen network, can be confirmed as extending.

Endogenous problems Utilizing the traffic network involves the problem of endogeneity. Discussions of economic geography show that regional economies will develop as transportation costs decline due to the development of traffic networks. However, it is also conceivable that an economically developed interregional transport network will be further implemented. As a result, the unobservable elements of the migration probability ζ_{ijt} are assumed to correlate with indicators of the traffic network. Morten and Oliveira (2016) avoided this problem by carrying out natural experiments[18] on the construction of a new capital and the accompanying development of a new transportation network and using data from before and after the development.

In this paper, we adopt an approach of constructing operational variables from two development

[15] Allen and Arkolakis (2014) is an early study that used geographic data on traffic networks to calculate travel times between two U.S. municipalities. This algorithm is developed with computational physics and calculates the transport cost (transport time) using the least expensive route from all points to all points. ArcGIS is used to process geographical data.

[16] For derivation, see Appendix 2.

[17] For example, Allen and Arkolakis (2014) uses interregional logistics volume data from 2007 and economic and demographic variables from the 2000 U.S. Census. In addition, the geographic information on transport networks is sourced from different statistics aggregated at different times. (It uses data on river and canal networks from 1999, data on railway networks from 2003, and data on highway networks from 2005.) Morten and Oliveira (2016) use panel data on highway networks from 1940 to 2000, but does not consider other modes of transport.

[18] Such endogenous treatment by natural experiments is often seen in previous studies. Ahlfeldt et al. (2015) uses data from the Berlin traffic network before and after the city was divided to control the causality of economic variables and traffic networks.

plans established before the period covered by the first-stage estimation (1975 to 1985). The first development plan is based on the National Shinkansen Railway Development Act (governed by the Ministry of Land, Infrastructure, Transport and Tourism) enacted in 1970, and in 1973, the development plan for each of five planned Shinkansen lines, such as the Tohoku Shinkansen line, was decided. The purpose of this law is to construct a "Shinkansen railway network that organically and efficiently connects core cities throughout the nation, while being sufficient to form a national trunk railway network." In addition to covering all eight districts of Japan, the planned route network connects almost all prefectural capitals.[19] The second development plan was established under the National Highway Act (governed by the Ministry of Land, Infrastructure, Transport and Tourism) enacted in 1957, and the development plans were decided from 1962 to 1987, taking more time than the development plan of the planned Shinkansen lines. The objectives of these development plans are similar to those of the Shinkansen network, aiming to construct highway networks that "constitute an essential part of the national automotive transport network and that communicate areas of particular political, economic and cultural importance and otherwise have particularly significant relationships with national interests." These two maintenance plans were initiated before 1975 at the start of the period covered by the explanatory and explained variables.

As indicated in the purpose provisions cited above, each development plan is intended to connect central cities in each district in Japan rather than connecting economically developed areas. Therefore, these transport network development plans can be regarded as having been established independently from the economic conditions of the time, and the transport network can be regarded as independent of the economic characteristics of the region. In addition, these are development plans, and it is assumed that they will not be affected by contrary causes and effects caused by economic activities after their formulation.[20] That is to say, although these development plans affect subsequent interregional travel time, it is assumed that they are unlikely to be affected by future economic activities that are likely to correlate with individual migration probabilities.

The operational variable Z_{ijt} in the first stage is created by the following procedure. First, the development plan drawing is prepared by connecting the prefectural offices connected in each development plan with a straight line.[21] Then, the same procedures as in Allen and Arkolakis (2014) are applied to the development plan diagram to find the shortest route between cities in the

[19] The Shinkansen network connects 43 prefectural capitals, while Naha (Okinawa), Nara (Nara), Tsu (Mie) and Mito (Ibaraki) are not connected.

[20] From this point of view, the exogenous nature of the second development plan, the National Highway Act, is dubious because it has been subject to revisions several times over a long period of time. The last revision was made in 1987 with the aim of dealing with the unipolar concentration in Tokyo and the depopulation of rural areas. Consequently, the results of applying the development plan in the two-stage least squares method described below discuss only the Census Migration Tabulation, which covers the period from 1990 to 2015.

[21] The purpose of this procedure is to extract information on which cities are connected in development plans. Previous studies such as Morten and Oliveira (2016) create links between all cities connected in the development plans (minimum spanning tree). Due to technical constraints, the above procedure is carried out in this paper.

development plan. This procedure is applied to each of the above two maintenance plans, and two types of operating variables, "Shinkansen Development Plan" and "High-Standard Trunk Road Network Plan," are prepared. The "simplified" development plan drawing prepared in the first stage is shown in Figure 2.

[Figure 2]

Again, land is allocated a constant speed. However, in order to derive solutions for Hokkaido and Okinawa, which are not connected in the development plans, routes with the same speed as on land[22] were arranged.

First-stage estimation using the two-stage least squares method A two-stage least-square estimation (2SLS) is used to obtain the consequences responding to endogeneity for the transport network. In the first stage, we estimate the following equation.

$$T_{ijt} = \theta_{jt} + \theta^{f}\mathbb{I}\{j \neq i\} + \theta^{R}\mathbb{I}\{j \in R\} + \theta^{d}D_{ij} + \theta^{Z}Z_{ijt} + \theta_{it} + \vartheta_{ijt}. \tag{28}$$

Here, T_{ijt} is travel time and Z_{ijt} is an operational variable created by the procedure described immediately above. Other variables are identical to the estimation formula for migration probability.

First, we describe the descriptive statistics for the variables used in the first stage of the estimation (Table 1). The number of inter-prefectural migrants also counts individuals who remain in the same prefecture as having migrated. Thus, some values for migrant probabilities and number of migrants are significantly larger relative to the mean and median, indicating that most prefectural residents remain in the same prefecture. Since the locations of the prefectural capitals have not changed, the travel distance remains unchanged with regard to time, whereas the travel time varies according to the development of the traffic network.[23] The instrumental variables Shinkansen Development Plan and High-Standard Trunk Road Network Plan are invariant between times because they correspond to the development plan and have the same characteristics as travel time. In addition, the descriptive statistics of migration data from the Basic Resident Register Migration Report and the National Census Migration Tabulation differ somewhat depending on the source and survey date of primary information, and the method of calculating migration probability described above. Therefore, we must present the estimation results of both explained variables from the viewpoint of robustness.

[22] These are: 1. the route between Sapporo (Hokkaido) and Aomori (Aomori) and 2. the route between Kagoshima (Kagoshima) and Naha (Okinawa).

[23] Due to unit-time issues with the fast-marching method, the travel distance and travel time are equalized using their respective maximums, which are the distance from Sapporo (Hokkaido) to Naha (Okinawa) and the travel time from Sapporo (Hokkaido) to Naha (Okinawa) in 1975.

[Table 1]

Table 2 below describes the estimation results based on the following estimation methods for both migration probabilities: 1. ordinary least squares (OLS), 2. instrumental variables (IV), 3. random effect panel estimation (Panel), and 4. instrumental variables panel estimation (Panel-IV). Estimation results of migration probabilities based on the Basic Resident Register Migration Report are shown in the upper part, and estimation results of migration probabilities based on the National Census Migration Tabulation are shown in the lower part. In implementing IV and Panel-IV, the Shinkansen Development Plan and the High-Standard Trunk Road Network Plan are implemented in the first stage, and the corresponding final estimation results are shown in (2) and (3), respectively. In addition, the results of estimation using only "travel distance" in the first-stage estimation are shown in (1).

Regarding estimation results, all estimates are significantly negative for "fixed migration cost dummy" and significantly positive for "same district dummy," yielding intuitive results. In addition, while positively significant signs for "travel time" were derived in the estimations of OLS and Panel, which do not use the instrumental variable method, negative significant coefficients for "travel time" were obtained when using instrumental variable methods, i.e., in the cases of IV and Panel-IV, and results suggesting the theory were obtained here as well.

[Table 2]

These estimates include, with some exceptions, three fixed effects: (1) origin-time fixed effects (origin-year FE), (2) destination-time fixed effects (destination-year FE), and (3) origin-destination fixed effects (origin-destination FE). The regional time indirect utility v_{jt} of the theoretical model corresponds to (2) the destination-year FE. Regional time series data corresponding to this value are shown in Figure 3. Figure 3a shows the panel estimation result using the Shinkansen Development Plan as an operational variable (Panel IV-2) as the indirect utility, and Figure 3b shows the panel estimation result using the High-Standard Trunk Road Network Plan as an operational variable (Panel IV-3). The names appended to the time series data indicate the names of the prefectures.[24]

[Figure 3]

[24] As the estimation results for the Basic Resident Register Migration Report do not ensure the significance of using the Shinkansen Development Plan as an operational variable, and the estimation results for the National Census Migration Tabulation do not ensure the significance of Distance Only and the High-Standard Trunk Road Network Plan as an operational variable, other estimation results were used. The reasons the significance of these variables is not ensured are assumed to be: (1) the development plans were influenced by migration and economic development and (2) the development stage of the development plans was decided depending on the economic development of the connected regions. In particular, the High-Standard Trunk Road Network Plan may be affected by (1) for the reasons mentioned above.

If the indirect utility level described in Figure 3 is checked, utility is high in prefectures in the Kanto and Kansai districts and in prefectures with central cities in regions with populations exceeding 1 million. In particular, the strong downtrend observed in the migration probabilities of the Basic Resident Register Migration Report may be attributed to structural changes in migration due to the time-series data extending over a long term.

4.2 Stage 2: Structural estimation of model parameters

In the second stage, we use the indirect utility ν_{jt} estimated in the previous section to carry out structural estimates for each parameter of the theoretical model. The variables to be estimated are wage elasticity to productivity in the region $1/\kappa$, elasticity of housing costs to regional labor supply, η^h, and preference parameters of utility functions, $\{\alpha,\ \beta,\ \gamma\}$.

Functions for estimation The functions for structural estimation are derived from regional labor market equilibrium (Equation (21)), regional housing market equilibrium (Equation (22)), and indirect utility index ν_{jt} at time t in region j (Equation (18)). Following Ahlfeldt et al. (2015), the variables in each region are defined as relative to the geometric mean,[25] so the derived functions to be estimated are defined in Equations (29)–(31) below.

$$\Delta \log\left(\frac{w_{jt}}{\overline{w}_t}\right) = \Delta\kappa_t + \frac{1}{\kappa}\Delta \log\left(\frac{A_{jt}}{\bar{A}_t}\right) + \Delta\varepsilon_{jt}^{L}. \tag{29}$$

$$\Delta \log\left(\frac{h_{jt}}{\bar{h}_t}\right) = \Delta\eta_t + \eta^h \Delta \log\left(\frac{L_{jt}}{\bar{L}_t}\right) + \Delta\varepsilon_{jt}^{H}. \tag{30}$$

$$\Delta\bar{\nu}_{jt} = \Delta \log\left(\frac{w_{jt}}{\overline{w}_t}\right) - \beta\Delta\left(\frac{f'\left(D_{jt}/\overline{D}_t\right)}{f\left(D_{jt}/\overline{D}_t\right)}\right) - (1-\alpha-\beta)\Delta\log\left(\frac{h_{jt}}{\bar{h}_t}\right) + \gamma\Delta\bar{X}_{jt} + \Delta\varepsilon_{jt}^{V}. \tag{31}$$

Additional discussions to interpret these equations are made prior to making structural estimates. This is because, first, some parameters and economic variables are unknown. Second, additional

---

[25] We supplement for the regional indirect utility indicator ν_{jt} and regional characteristics X_{jt}. Since the utility level defined in Equation (18) has already been logarithmized, the regional indirect utility V_{kt} is restored by an inverse logarithmic transformation, and then relativized by the geometric mean and logarithmized again. Moreover, we carried out special leveling work because some regional characteristics X_{jt} took negative values. Describing these tasks in expressions produces the following:

$$\bar{\nu}_{jt} = \log\left(\frac{V_{jt}}{\bar{V}_t}\right) = \log\left(\frac{\exp(\nu_{jt})}{\overline{\exp(\nu_{jt})}_t}\right).$$

$$\bar{X}_{jt} = \left\{\frac{X_{jt} - X_t^{min}}{X_t^{max} - X_t^{min}}\right\}.$$

instrumental variables are required for structural estimation. Therefore, we will discuss two economic variables separately, namely, (1) birth rate, (2) regional productivity and manipulation variables.

Subsidiary estimates (1): Discussion of child-rearing expenses In this discussion, we make a simple estimate of the functional parameters of a function $f(D_{jt})$ that defines the birthrate by prefecture, incorporating the discussion of the negative causality between birthrate and population density in De la Croix and Gobbi (2017).

De la Croix and Gobbi (2017) combined demographic statistics and various economic statistics of 44 developing countries[26] to draw the conclusion that population density negatively impacts birthrates. Based on the estimation equations used in this study, we formulate the estimation equation in Equation (32) below.

$$n_{jt} = \beta_0 + \beta_1 \log\left(1 + D_{jt}\right) + \sum_{i=2}^{M} \beta_i X_{ij}\,. \tag{32}$$

Here, n_{jt} is the birthrate in region j, D_{jt} is the population density in region j, and X_{ij} represents explanatory variables.[27] Each variable is logarithmized, and the estimate β_1 shows the elasticity of the birthrate to population density. Following previous research, the "annual marriage rate by prefecture (per 1,000 people)" ("Demographic Survey" (Ministry of Health, Labour and Welfare)) and "gross prefectural product per capita" ("Annual Report on Prefectural Accounts" (Cabinet Office)) are used as explanatory variables. The birthrate of the explained variable n_{jt} uses the "number of births by prefecture" for the working-generation population (25–71 years old) ("Demographic Survey" (MHLW)). The period covered by the estimation is 1989–2015. Fixed effects by year are introduced to eliminate the effects of economic fluctuations and other elements.

Estimation results are shown in Table 3 below, which shows the estimation results of "basic estimation" but does not introduce explanatory variables, and "extended estimation," which includes explanatory variables.

[Table 3]

[26] The same relation is highly plausible, as Sato (2007) and Kato (2017) have shown that it exists in Japan, both theoretically and empirically.
[27] De la Croix and Gobbi (2017) uses discrete data for n_{jt} because individual data from each country are used. For this reason, it yields slightly different results to the analysis in this paper, such as making Poisson estimates and clustering by country and region.

In the basic estimation, positive coefficients were observed in the logarithmic term of population density, and in the extended estimation, negative coefficients were observed, and results consistent with the previous research were obtained in the extended estimation. For the extended estimation, a negative sign for population density is significant at the normal estimate, while the significance disappears in the panel analysis of random effects. Therefore, the values of the normal estimate are used in the structural estimates that follow.[28]

From the estimation results, the relationship between the birthrate indicator n_{jt} and population density D_{jt} can be defined as $\log n_{jt} = -3.5982 - 0.0136\log(1 + D_{jt})$. In addition, the corresponding childcare cost function $f(D_{jt})$ can be defined as:

$$f(D_{jt}) = \exp(3.5982) \cdot D_{jt}^{0.0136} \tag{33}$$

We use this result in the structural estimates that follow.

Subsidiary estimates (2): Regional productivity estimates In addition to estimating child-rearing costs by region, it is necessary to estimate regional productivity A_{jt} when making structural estimates. In this paper, we use the method of Hayashi and Prescott (2002), a prominent study of Japanese productivity.

This method restores the total factor productivity (Solow residual) by growth accounting, and here, the regional productivity A_{jt} is restored by applying the procedure by prefecture. The data used here are "production volume by prefecture (expenditure)" and "number of employees by prefecture" in the Annual Report on Prefectural Accounts (Cabinet Office) from 1955 to 2016.[29] Assuming that the production function here is of the Cobb-Douglas type, we calculate the capital intensity θ using the following equation:[30]

$$lnY_{jt} = A_{jt} + \theta lnK_{jt} + (1 - \theta)lnL_{jt} + \eta_j + \varepsilon_{jt}.$$

Table 4 contains information on the data used, the estimation method, and the results of the estimation. In particular, when panel estimation is carried out in the fixed effect model, capital intensity is $\theta = 0.358$, and regional productivity $A_{jt} = Y_{jt}/(K_{jt}^{\theta}L_{jt}^{1-\theta})$ is restored using this.

---

[28] Previous studies have yielded estimates of −0.110 in Sato (2007) and −0.141 or −0.157 in De la Croix (2017). Although there were qualitative differences between regional and individual survey analyses, the estimates were interpreted as valid. In addition, the significance problem of panel analyses is a future problem.

[29] There are several points to note in using these data. First, calculations in the Annual Report on Prefectural Accounts have been revised in accordance with revisions to the definitions and calculation methods for national accounting. Second, the aggregate value of "gross prefectural product" does not match the "gross domestic product," and there are concern of errors in the "gross prefectural product."

[30] The production function corresponding to the gross prefectural product Y_{jt} is $Y_{jt} = A_{jt}K_{jt}^{\theta}L_{jt}^{1-\theta}$. In addition, it is assumed that the capital intensity θ is the same in all prefectures.

[Table 4]

Operational variables The three variables used are: (1) exogenous productivity shock on the region (Bartik instrument), (2) labor market indicators, and (3) the ratio of available land to residential land.

Bartik instruments are indicators presented in Bartik (1991) and are often used in spatial economics. This indicator summarizes the impact on the regional economic structure of changes in the workforce on a country-by-country basis in each industry. That is to say, where the workforce of the industry i in prefecture j at time t is $L_{i,j,t}$, the Bartik instrument $\varDelta B_{jt} = B_{jt} - B_{jt-1}$ of prefecture j at time t of the prefecture is defined as follows:

$$\varDelta B_{jt} = \sum_{i=1}^{I} \left(\Delta \log L_{i,-j,t}\right) \frac{L_{i,j,0}}{L_{j,0}}. \tag{34}$$

In this paper, we use Bartik instruments as operational variables for regional labor markets and regional utility.

Morten and Oliveira (2016) applied the idea of Bartik instruments to compile integrated indicators for the labor market by prefecture. This indicator M_{jt} shows how strongly high wages in a particular prefecture attract workers from other prefectures.

$$M_{jt} = \sum_{k \neq j}^{M} \left\{ - \frac{L_{kt-1}}{\left(\widehat{\lambda^f} + \widehat{\lambda^d_{kj}} D_{kj} + \widehat{\lambda}^t Z_{kj}\right)} \right\}. \tag{35}$$

This indicator is useful as an operational variable because the responses of labor supplies to the same productivity shock to different prefectures are assumed to be different. This paper applies this operational variable to the regional housing market because an individual who purchases housing has already chosen the prefecture as the destination.

As a final operational variable, we use the ratio of the area available for residence to the area of the prefecture $x_j^{housing}$ in accordance with Diamond (2017). The use of this operational variable is supported by previous studies on the relationship between the elasticity of housing prices and land availability. This operational variable is used for individual indirect utility.

Taken together, the discussion of these identification restrictions allows the moment condition in the second stage structural estimation to be expressed by Equation (36) for each corresponding operational variable. The variations of each error term are identical to those defined in Equations (29)–(31).

$$\begin{cases} E\left(\Delta\varepsilon_{jt}^{L}\Delta Z_{jt}^{L}\right) = 0, & \Delta Z_{jt}^{L} \in \left\{\Delta B_{jt}\right\}, \\ E\left(\Delta\varepsilon_{jt}^{H}\Delta Z_{jt}^{H}\right) = 0, & \Delta Z_{jt}^{H} \in \left\{M_{jt}\right\}, \\ E\left(\Delta\varepsilon_{jt}^{V}\Delta Z_{jt}^{V}\right) = 0, & \Delta Z_{jt}^{V} \in \left\{\Delta B_{jt}, x_{house}\right\}. \end{cases} \tag{36}$$

In the following, we describe the estimation results for these moment conditions.

Results of estimates For the purpose of consistency with the first phase of this section, here we use indirect utility indicators restored from two statistics: the Basic Resident Register Migration Report and the National Census Migration Tabulation. This means that Structural Estimate A includes the time series from 1975 to 2015, and Structural Estimate B includes the time series from 1990 to 2010.[31]

Before presenting the estimation results, we will explain the economic variables used in the second stage. For the "wage rate" w_{jt}, we use the "average monthly cash earnings" from the Basic Statistical Survey of Wage Structure (MHLW). For regional productivity, we use the total factor productivity A_{jt} restored in accordance with Hayashi and Prescott (2002) in subsidiary estimation (2). "Labor population" is the population calculated by totaling the population of municipalities listed in the Basic Tabulation Results for Population in the National Census from the ages of 25 to 71. As "monthly rent/3.3 m^2" is included in the Annual Report of the Retail Prices Survey (Ministry of Internal Affairs and Communications), this is used as housing expenses h_{jt}.

Regarding regional characteristics X_{jt}, we prepared three indicators: "amenity 1," "amenity 2," and "social capital." A discussion of these indicators is given in the appendix, and an overview of the indicators is given in this section. In deriving each "amenity," we used various statistical indices from 1975 to 2015,[32] which are published in Forms of the Prefectures as Seen in Statistics" (Ministry of Internal Affairs and Communications). In addition, Social Capital Stock Estimation Data was used when compiling "social capital." In this data, the net stock of social capital for the prefectures from 1960 to 2014 is recorded in 16 categories. Since the traffic network is used as the first-stage operational variable, this paper examines 13 categories excluding "roads," "ports," and "air."

Diamond (2016) is used as a reference for the process of obtaining regional characteristics'

---

[31] Indirect utility indicators corresponding to the 2015 National Census Migration Tabulation have been omitted because the calculation interval differs from other fiscal years and is not suitable for structural estimation.

[32] In Forms of the Prefectures as Seen in Statistics (Ministry of Internal Affairs and Communications), the data series differ from year to year. The data available for Structural Estimate A and Structural Estimate B differ because the period covered differs. Twenty-one series were used for the Basic Resident Register Migration Report for 1975–2015, and 33 series were used for the National Census Migration Tabulation for 1990–2000. (See Appendix 3 for details.)

indicators from various statistical series. As the first stage, the series were divided by cases into sectors. Concretely, statistical indices by prefecture were divided into the four sectors of: (1) environment, (2) education, (3) urban, and (4) welfare, and social capital was divided into the five sectors of (1) transportation, (2) living, (3) education, (4) environmental, and (5) industry. As a second stage, principal component analysis (PCA) was performed for each of these sectors to extract the first principal component of each sector. PCA was then implemented for the principal components of these sectors, and the regional characteristics thus derived are defined as "amenity 1" and "social capital." In addition, "amenity 2" refers to an indicator extracted by performing PCA after extracting a statistical series that correlates significantly with the indirect utility function and compares the estimation result with the case when using "amenity 1" in order to confirm robustness.

Descriptive statistics for the second-stage estimates are presented in Table 5. Descriptive statistics for Structural Estimate A and Structural Estimate B are described in the upper and lower rows, respectively.

[Table 5]

Structural estimates of the second-stage estimates are presented in Table 6. Here, structural estimates are made using the two indirect utility indicators derived in the first-stage estimation, the economic variables by prefecture described above, and the three regional characteristics. There are eight estimation results for each indirect utility because we also derive baseline estimates that do not use regional characteristics.

[Table 6]

We consider the results of Table 6. The coefficient $1/\kappa$ in Equation (29) can be interpreted as indicating "elasticity to wage productivity,"[33] and the coefficients are positively significant near 1 in all cases, indicating that wages rise due to increased regional productivity. The coefficient η^h in Equation (30) can be interpreted as indicating "inverse elasticity of housing demand to housing costs." This coefficient η^h is also significantly positive in the vicinity of 1 in all cases of Structural Estimation A and in reference estimates without the regional characteristics of Structural Estimation B, and significantly positive in the vicinity of 0.5 in other cases of Structural Estimation B). [34]

Next, the results of Equation (31) will be explained. This result corresponds to each preference

[33] Interpreting the regional production function used in this paper (Equation (5)), it becomes $w = \kappa A k^{-\kappa}$. In the conventional formulation, κ can be interpreted as capital intensity, but since the total factor productivity is restored from the production function this time, it is probable that a value deviated from the normal value was derived.

[34] This is consistent with formulations such as Moretti (2011) and Diamond (2017), in which housing prices increase in response to population inflows.

parameter of the indirect utility of each consumer. "Preference for children" can be defined as an inverse function of the child-rearing cost function $f(D_{jt})$, and all coefficients β for this parameter are significant, with coefficients around 0.8 being significantly positive in estimates of Structural estimate A and 0.2–0.4 being significantly positive in estimates of Structural Estimate B. The difference between the two is thought to be due to the difference in the period covered, and the estimates in the National Census Migration Tabulation for which the period covered is shorter reflect the effects of the declining birthrate in recent years. The $-(1-\alpha-\beta)$ coefficient of h_{jt} in Equation (31) shows the proportion of income spent on housing, indicating "preference for housing." The coefficients are positively significant, approximately 0.35 for Structural Estimation A, 0.4 for Structural Estimation B, and 0.25–0.3 for the rest.

We will describe estimation results on regional characteristics. In the case of both Structural Estimation A and Structural Estimation B, there was no significance of regional characteristics.[35] The coefficients of explanatory variables for other economic variables are also robustly significant for the introduction and selection of regional characteristics. Regarding Structural Estimate A and Structural Estimate B, the coefficients of social capital are large for both. In contrast, in Structural Estimate A, the coefficients of "amenity 1" following Diamond (2017) are relatively large, as are the coefficients of "amenity 2," which extracts significant ones in Structural Estimate B.

Finally, we discuss the validity of the parameters. The bottom of Table 6 shows the values of the parameters α, β, γ restored from the respective estimation results. Regarding these values, it is desirable that the following conditions are satisfied from the formulation of the individual utility function in Equation (10):

$$\alpha > 0, \beta > 0, and\ 1-\alpha-\beta > 0.$$

In Structural Estimate A using the Basic Resident Register Migration Report, α, which corresponds to "preference for consumption," is negative and does not meet the requirements of the utility function. On the other hand, the restored values in Structural Estimate B using the Basic Tabulation Results for Population in the National Census satisfy all the above assumptions, indicating that the results are almost robust with regard to the selection of regional characteristics. Therefore, the following discussion uses the results for Structural Estimation B.

In this chapter, we conducted a two-stage analysis based on quantitative spatial economics. In the first-stage estimation, indirect utility indicators by prefecture were derived for the two types of domestic migration-related statistics by using the two-stage least-squares method with distance and travel time

[35] Demographic migration arguments often do not specify the coefficients or significance of regional characteristics (Amenity) (Ahlfeldt et al. 2015, Diamond, 2017). However, this paper discusses it because it is formulated with a certain level of interest.

as operational variables. In the second-stage estimation, structural estimation was executed on the equilibrium condition formulated by the spatial equilibrium model to estimate the internal structure of the utility of the individual who chooses residence between regions and to obtain various parameters. In the next chapter, we will use these estimates to describe regional demographics.

5. Analysis of regional populations

In this chapter, we discuss the population dynamics of Japan's prefectures using the spatial equilibrium model formulated in Chapter 3 and the model parameters derived in Chapter 4. In the first half of this chapter, we formulate population dynamics consistent with the spatial equilibrium model. In the second half, we derive the projected values of the population dynamics of the prefectures and confirm the significance of introducing housing choices based on utility.

5.1 Spatial equilibrium and population dynamics

In Chapter 3, we formulated migration based on the utility function and utility of a simplified household model that endogenizes population dynamics. If we describe our model based on the formulation of the spatial equilibrium in Equations (21)–(24), we see Figure 4 below.

[Figure 4]

In this figure, the squares represent the equations. In particular, those accompanied by numbers correspond to the equilibrium Equations (21)–(24). The solid arrows point to the variables derived in the spatial equilibrium model of this paper, while the dashed arrows indicate the variables that require additional assumptions for derivation. Looking at Figure 4, we can see that there are three relationships that determine the transition of regional economies: (1) population dynamics), (2) dynamics of regional productivity (regional labor markets), and (3) dynamics of geographic factors. Of these, the latter two are considered exogenously,[36] and a detailed discussion of population dynamics is presented here.

Discussion of demographic dynamics As we have seen in Chapter 2, there are many studies of multiregional demographic endogenous models in Japan. After describing how regional populations are addressed in these previous studies, we describe the population dynamics required to derive regional populations in accordance with our formulation.

Sato and Yamada (2005) and Sato (2007) described population dynamics in a generic spatial

[36] Geographic factors (regional characteristics X_{jt} and migration costs c_{ijt}) are of great significance for analysis because they capture regional elements of population migration, but they are beyond the scope of the analysis in this paper. In addition, following Hayashi and Prescott (2002), regional productivity is defined with the Solow residual. Therefore, in formulating the dynamics of regional productivity, the formulation of the corporate sectors, including additional data in the industrial sectors in particular, the relationship between household savings and investment, and the capital migration equation, is required. This work is another task for later study.

equilibrium model. Sato and Yamada (2005) dealt with a dynamic population equilibrium that states that the series of the populations in two regions N_{1t}, N_{2t} the total population N_t, and the indirect utility series of the regions V_{1t}, V_{2t} satisfy three conditions.[37] Taken together, these conditions lead to a balance between dynamic and spatial equilibrium of the population when the coincidence of the next population and the number of children in the economy as a whole coincide with the equivalence the indirect utility of each region.

In Sato (2007), demographic and economic steady-state conditions were analyzed, and the variables are time-independent. Since there are households and enterprises in the economy, wages and interest rates are determined from the conditions for maximizing enterprise profits, and consumption in each period and the birthrate are determined from the conditions for maximizing household utility. In Sato (2007), the number of young people $N_{i,y}$ in region i is determined based on (1) the regional utility-equalization condition and (2) the migration-stable condition.[38] Wages and birthrates are given by the number of young people $N_{i,y}$ in each region. Since the elderly population is consistent with the young population in the previous period, steady-state definitions can be made by these two conditions.

In the above spatial equilibrium model, indirect utility coincides by households choosing a region through utility maximization. On the other hand, the new economic geography (quantitative spatial) model corresponding to the present study has wider variables to be analyzed in equilibrium and derives an equilibrium in which, as seen in Morita and Yamamoto (2018), corporate profits are consistent in each region and an equilibrium in which, as shown in Goto and Minamimura (2018), real wages are consistent in each region.[39] New economic geography also uses replicator dynamics as a convergence process to equilibrium, whereas Goto and Minamimura (2018) formulated so that the regional population converges to a population level that corresponds to an equilibrium level in real wages.[40]

---

[37] For the populations in the two regions N_{1t}, N_{2t} and the total population N_t, and the indirect utility of each region V_{1t}, V_{2t}, (1) the congruence of the populations in the two regions and the total population, (2) the law of motion for population, and (3) the no-migration condition is defined.

$$N_t = N_{1t} + N_{2t}, \tag{1}$$
$$N_{t+1} = N_{1t+1} + N_{2t+1} = n_{1t}N_{1t} + n_{2t}N_{2t}, \tag{2}$$
$$V_{1t} = V_{2t}. \tag{3}$$

n_{1t}, n_{2t} are the birthrates in each region.

[38] (1) The regional utility-equalization condition and (2) the migration-stable condition is defined as follows. V_i is the young population and $N_{y,i}$ is the regional utility. The migration-stable condition was introduced by Sato (2007), which dealt with economies in three or more regions.

$$V_i = V_j = \bar{v}, \tag{1}$$
$$\frac{\partial V_i}{\partial N_{y,i}} < 0. \tag{2}$$

[39] For example, the real wage equilibrium condition in Goto and Minamimura (2018) is described in (1) below.

$$\left(\frac{\omega_i^s}{\overline{\omega}^s}\right) \leq 0, \quad and \quad \left(\frac{\omega_i^s}{\overline{\omega}^s} - 1\right)\lambda_i^s = 0. \tag{1}$$

ω_i^s is the real wages in region i and $\overline{\omega}^s$ is the average. Furthermore, λ_i^s is the population share of the region. The second equation is introduced, and its equilibrium condition also encompasses the case of full agglomeration.

[40] Goto and Minamimura's (2018) steady state of the population is formulated as follows:

$$\bar{n}^* = 1, \tag{1}$$
$$\left\{\frac{n_{it}^*}{\bar{n}_t^*} - 1 + \Xi\left(\frac{\omega_{it}^*}{\overline{\omega}_t^*} - 1\right)\right\}\lambda_{it}^* = 0. \tag{2}$$

What is particularly discussed in these formulations is the steady state of spatial equilibrium and population. Under spatial equilibrium, the real wages ω_{it}^{*} in each region coincide, and the birthrate n_{it}^{*} also coincides because opportunity costs coincide. Therefore, the steady-state condition of demographic dynamics in each region (Note 36, Equation (2)) is an identity, and under spatial equilibrium, the steady state of the population is always realized only by the fact that the birthrate satisfies the conditions. On the other hand, the steady-state conditions of the population are said to be met even when real wages ω_{it}^{*} do not coincide among regions.[41] Therefore, no spatial equilibrium can be achieved under a population steady state. In subsequent formulations of population dynamics, Japan's prefectural population and regional economy will be predicted using projections from the current economic conditions (2015) rather than balanced welfare assessments or population steady-state derivations,[42] taking into account this spatial equilibrium and the inconsistency with population steady states.

Population dynamics in this paper In the following, we discuss the population dynamics corresponding to the quantitative spatial model of this paper. Before discussing, we list the variables pertaining to population that we dealt with in Chapters 3 and 4 and their definitions.

1. Working population (L_{jt}):
   Population aged 25–75 years who commence work in region j after region selection at time t.
2. Entry into the labor force (L_{jt}^{+}):
   Population aged 25–29 years who select region j based on region selection at time t.
3. Younger population $(N_{y,mt})$:
   Population aged 25–29 years who participate in region selection at time t in region m.
4. Birthrate (n_{jt}):
   The number of children an individual chooses after region selection at time t in region j.
5. Population density (D_{jt}):
   Population density after region selection at time t in region j.

Various variables are added to these definitions and some assumptions are made to describe population dynamics. Since the time series of structural estimates up to Chapter 4 was in five-year units, we will also discuss in five-year units.

Here, the birthrate at time t in each region i is n_{it}, and the average birthrate among regions is $\bar{n}_t$. Ξ corresponds to the degree of freedom of migration and is a general formulation in new economic geography. Under these conditions, the number of adults and population share in each region remain constant.

41 In this case, the higher the real wage ω_{it}^{*} in a region, the lower the birthrate n_{it}^{*}. This corresponds to the current situation in Japan.

42 These theoretical considerations are challenges for the future.

First, we define the population by generation. In this paper, we define three generations: the juvenile generation population $N_{b,mt}$, the working generation population $N_{w,mt}$, and the elderly generation population $N_{o,mt}$. The juvenile generation population includes the population from 0 to 24 years old, the working generation population includes the population from 25 to 74 years old, and the elderly generation population includes the population 75 years and older.[43]

The juvenile generation population $N_{b,mt}$ is defined by the following Equation (37) with respect to the juvenile generation population in the previous period $N_{b,mt-1}$:

$$N_{b,mt} = \{n_{mt-1}L_{mt-1} - N_{y,mt}\} + N_{b,mt-1}. \tag{37}$$

Here, $n_{mt-1}L_{mt-1}$ corresponds to a population (cohort) of individuals born in the previous period $t-1$ in region m. In addition, the young population $N_{y,mt} = n_{mt-5}L_{mt-5}$ is the individuals who are 25 to 29 years old at time t in region m and who participate in region selection during this period.

Next, we define the working generation population $N_{w,mt}$. The population of this generation can be described by Equation (38) below using the entry of the working population L^{+}_{mt} at time t, the population of the retiring cohort $n_{mt-10}L^{+}_{mt-10}$, and the working generation population of the previous period $N_{w,mt-1}$.

$$N_{w,mt} = \{L^{+}_{mt} - n_{mt-10}L^{+}_{mt-10}\} + N_{w,mt-1}. \tag{38}$$

While region m welcomes the population of L^{+}_{mt} from all prefectures in Japan at time t, workers aged 75–79 will retire and transition to the elderly generation. We assume that the working generation population$N_{w,mt}$ in this region m is consistent with the working population L_{mt} in region m.[44]

$$N_{w,mt} = L_{mt}. \tag{39}$$

That is, housing demand in the regional housing market at time t is defined by this generation population.

Finally, we define the elderly population $N_{o,mt}$. This generation includes the population 75 years or older, and its dynamics are formulated as shown in Equation (40).

$$N_{o,mt} = n_{mt-10}L^{+}_{mt-10} + (1 - d_{mt})N_{o,mt-1}. \tag{40}$$

[43] Structural estimates assume that the working population is 25–71 years old, which is based on the age groups in the census. On the other hand, we made the above assumption because we needed to describe the population structure every five years.

[44] The consumption, child-rearing, and housing purchases described in the utility function of Equation (10) are considered to be reasonable assumptions because the majority are purchased during this generation (25 to 75 years old).

$n_{mt-10}L^{+}_{mt-10}$ is the population 75–79 years old at time t, which is consistent with the retirement population above. In addition, the population of $(1-d_{mt})N_{o,mt-1}$ from the elderly generation population in the previous period $N_{o,mt-1}$ shall remain in the economy at time t. Here, d_{mt} corresponds to the mortality rate of the elderly generation population.[45]

Finally, we formulate the migration of the young generation in a quantitative spatial economic model and connect the quantitative spatial model with population dynamics. The population entering the economy in region m at the beginning of time t is the young population $N_{y,mt}$. The definition is restated in Equation (41) below.

$$N_{y,mt} = n_{mt-5}L_{mt-5} \tag{41}$$

That is, individuals aged 25 to 29 who enter the economy at the beginning of time t make region selections through the utility maximization described in Chapter 3. As a result of the region selection, the probability π_{mjt} that an individual in region m will migrate to region j is determined, and as a result, the labor population entry in region j is defined as L^{+}_{jt} as shown in Equation (42).

$$L^{+}_{jt} = \sum_{m=1}^{M} \pi_{mjt}\, N_{y,mt}. \tag{42}$$

The above formulations fully describe the population dynamics by generation in the economy. By aggregating these, the total population N_{jt} of region m at time t is defined.

$$N_{jt} = N_{b,mt} + N_{w,mt} + N_{o,mt}. \tag{43}$$

As the total population of the region N_{jt} has been derived, it is possible to define the remaining population variables: 4. birthrate (n_{jt}) and 5. population density (D_{jt}). The area of the prefecture remains unchanged, and the child-rearing cost function is considered not to change. Under this assumption, the population density D_{jt} and birthrate n_{jt} are defined by the total population N_{jt} for the period.

$$D_{jt} = \frac{N_{jt}}{AR_j}. \tag{44}$$

[45] Needless to say, the mortality rate should also be defined for the juvenile generation population and the working generation population, but these are omitted here because they complicate the relationship between the regional economy and the regional population.

$$n_{jt} = \exp(-3.5982)\hat{\beta}\, D_{jt}^{-0.0136} w_{jt}. \tag{45}$$

Here, AR_j is the area of prefecture j and remains unchanged. In addition, since the wage rate w_{jt} is assumed to be determined by regional productivity, the birthrate n_{jt} in population dynamics is determined when the total regional population N_{jt} is given.

As noted above, this section derives population dynamics consistent with the spatial models formulated and estimated through Chapters 3 and 4. Finally, we present a conceptual diagram of the formulation of population dynamics.

[Figure 5]

The formulation given here makes it possible to describe the population structure in all prefectures consistently with spatial equilibrium.

Assumptions about population dynamics and economic models As a final discussion of population dynamics, we examine the assumptions made in Chapter 3 to accurately correspond the formulation of the individual utility function and indirect utility function[46] with population dynamics.

According to the formulation of population dynamics, those who select regions at time t are limited to those aged 25 to 29 in the current period. On the other hand, the individual utility function U_{jt} and the indirect utility function V_{jt} show that the utility obtained by an individual's residence in region j is defined by various economic variables at time t. Given that individuals make a region selection only on this one occasion due to the settings of region selection, the young population at time t in region m derives the (logarithmic) indirect utility v_{jt}^{e} of residing in the region in a myopic fashion. In other words, in this paper, we assume that individuals only infer the lifetime utility obtained by residing in region j using the present economic conditions in region j and do not maximize the lifetime utility based on perfect foresight.[47] This can be formulated as in Equation (46) below. In addition, since individuals observe the attributes of each region prior to migration, the observed economic characteristics of region j are the already-realized economic characteristics of the previous period. Therefore, we assume that the indirect utility v_{jt}^{e} corresponding to Equation (46) is as given in Equation (47) below.

[46] For the purposes of discussion, (1) the individual utility function and (2) the indirect utility function are restated below.

$$U_{jt} = C_{jt}^{\alpha} n_{jt}^{\beta} H_{jt}^{1-\alpha-\beta} \cdot \exp\left(X_{jt}^{\gamma}\right) \tag{1}$$

$$V_{jt} = \alpha^{\alpha}\beta^{\beta}(1-\alpha-\beta)^{1-\alpha-\beta} w_{jt}\left\{f\left(D_{jt}\right)\right\}^{-\beta} h_{jt}^{-(1-\alpha-\beta)} \cdot \exp\left(X_{jt}^{\gamma}\right) \tag{2}$$

[47] This idea is also observed in Kennan and Walker (2011), a more detailed formulation of migration. They are considering the lifetime income maximization behavior of individuals dynamically, but they make some simplifications on the assumption of verification and replacing this dynamic optimization problem with an expected lifetime income maximization problem.

$$\pi_{ijt} = \frac{\exp\left(v_{jt}^{e} - c_{ijt}\right)}{\sum_{m=1}^{M}\exp(v_{mt}^{e} - c_{imt})}. \tag{46}$$

$$V_{jt}^{e} = \alpha^{\alpha}\beta^{\beta}(1-\alpha-\beta)^{1-\alpha-\beta}w_{jt-1}\left\{f\left(D_{jt-1}\right)\right\}^{-\beta}h_{jt-1}^{-(1-\alpha-\beta)}\cdot\exp\left(X_{jt-1}^{\gamma}\right). \tag{47}$$

Additional assumptions about individual birth behavior are also needed. Based on formulation of the utility function, it is interpreted that an individual who selects region j determines the lifetime births n_{jt} between the ages of 25 and 29 based on economic variables at this time (time t). On the other hand, the subsidiary estimation (1) in Section 4.2 defines "births by prefecture" for the working generation population and estimates the function of child-rearing costs $f(D_{jt})$. This is in accordance with the general assumption that households raise children while working. Demographically strong assumptions are made to ensure consistency between these settings. That is, we assume that myopic individuals determine their lifetime births n_{jt} between the ages of 25 and 29[48] and actually raise children during this period. Under this assumption, the equation for fertility n_{jt} (Equation (45)) in the projection is modified as follows:

$$\hat{n}_{jt} = \frac{1}{CR_{jt}}\exp(-3.5982)\hat{\beta}\,D_{jt}^{-0.0136}w_{jt}. \tag{48}$$

CR_{jt} here shows the ratio of the number of births by parents aged 25 to 29 to the total number of births in region j at time t, and the adjusted birthrate $\hat{n}_{jt}$ shows the birthrate when all births in the region are given by the young population.

The discussion in this section suggests that the following assumptions need to be made to ensure consistency between the spatial equilibrium model in this paper and population dynamics.

1. Individuals in this paper make migration choices based on regional utility only once in their lifetimes. Individuals are myopic, and individuals who choose to migrate at time t derive their utility and infer their lifetime utility using only the economic variables for region j at time $t-1$.
2. Individuals in this paper give birth and rear children only in the young period (25–29 years).

[48] As the lifetime births are determined based on the economic variables (population density and wage rate) of each region at the time point, the formulation of the estimation continues to be valid. In addition, since the definition range of the fertility rate of women, which is summarized in the "Demographic Health Centers and Municipalities" (described later), is 15 to 49 years old, it is necessary to further divide the ratio of the population aged 25 to 75 years old and the population aged 15 to 49 years old in the actual calculation.$n_{jt}tD_{jt}w_{jt}$

These are relatively strong assumptions that have been made to ensure consistency between the spatial equilibrium model and population dynamics. As such, these need to be taken into account in making estimates and interpreting the results of estimates.

5.2 Population Forecast by Prefecture Based on the Quantitative Spatial Model

The definitions of population dynamics in the previous section make it possible to describe economic and population dynamics by prefecture in Japan. In this section, we use projections to forecast the economic and demographic trends in each prefecture, starting in 2015.

Various statistical and additional assumptions used in forecasting First, we explain various statistical indices used in the forecasting. In making forecasts, it is necessary to first describe the current state of the economy and population in the prefecture at the start point (2015). For the sources of data used to derive regional economies, we continue to use the statistics used in Structural Estimate B in Section 4.2. Migration can be restored and predicted from the interregional migration costs c_{ijt} estimated in Section 4.1 and the indirect utility v_{jt} derived from the spatial economic model.

Therefore, what is additionally required for forecasting is a detailed population structure and population dynamics statistics. For the population of each prefecture by five-year age group in 2015, we used "population by sex, prefecture, municipality, and age (five-year age groups)" listed in "Population Dynamics by Health Center and Municipality" (MHLW). The "total fertility rate and birthrate by maternal age group" in the same statistics are used to calculate the fertility rate. This birthrate by age group is multiplied by the population by age group to obtain the share of the number of births by age group and derive the modified birth rate $\hat{n}_{jt}$. The 2015 Life Table by Prefecture was used to derive the mortality rate d_{mt} of the elderly generation.

Several further assumptions are made about the additional matters required when using the quantitative spatial model for forecasting. Looking at the quantitative spatial model described in Figure 4 and its equilibrium conditions described in Chapter 3, regional productivity $\{A_{it}\}_{i\in M}$, the young generation population $\{N_{y,it}\}_{i\in M}$ at the beginning of the period, regional characteristics $\{X_{it}\}_{i\in M}$, and migration costs $\{c_{ijt}\}_{i,j\in M}$ should be given at the start of time t. Of these, since the young generation population is given by population dynamics, other economic variables are assumed as follows.

1. Regional characteristics $\{X_{it}\}_{i\in M}$ and migration costs $\{c_{ijt}\}_{i,j\in M}$ are constant across time.
2. Regional productivity $\{A_{it}\}_{i\in M}$ will grow at the average growth rate for each region from 1975 to 2015.

Prefectural population forecast by projection In the forecast, we use the spatial model shown in Figure 4 and the model with various assumptions in this chapter added to the population dynamics shown in Figure 5. For the parameters, we used a combination consistent with assumptions from the estimation results of structural estimates in Chapter 4.[49] As a result, the following prediction results were obtained. Figure 6 uses the parameters in the results for "amenity 1" and Figure 7 uses the parameters in the results for "social capital" with social capital as the regional characteristics. The left-hand side of each chart shows the "total Japan population forecast" aggregating the populations of each prefecture, while the right-hand side of the charts shows the population growth rates for each prefecture in the estimation period (2015–2125). For each parameter combination, population dynamics in the absence of demographic migration[50] are derived as a reference estimate and are appended to the chart. These values are expressed as the total using dashed lines in the total population forecasts and as the growth rate using horizontal axis values in population growth rate comparisons. It should be noted that the estimates are higher than the level in the Future Estimated Population (Cabinet Office), reflecting the fact that deaths up to the age of 75 are not considered and that the effects of the age structure on birth are simplified.

[Figure 6]

[Figure 7]

In common with the results of Figure 6 and Figure 7, it can be pointed out that the total population in the presence of migration is smaller than in the absence of migration, and that when migration is formulated, the population decline in urban areas with a large initial population accelerates, whereas the population decline rate in regions with a small population is greatly eased. These conclusions can be regarded as robust results for the formulation of regional characteristics, as they are derived from entirely different regional characteristics.

The latter result regarding population growth is consistent with intuition. In the endogenous population model, individuals yield more utility than the number of children, so in the model in this chapter, where child-rearing costs correlate with population density, urban utility is relatively low. In addition, an increase in housing costs due to an increase in the size of the working population also causes a decline in the utility of urban areas.

It is difficult to interpret the former result, in which the total population becomes low as a result of migration to regions where it is easy to raise children, against the result of the population growth rate,

---

[49] See Chapter 4 for details. "Amenity 1" and "social capital" in the results based on the National Census Migration Tabulation (Structural Estimation B) are applicable.

[50] This refers to the situation in which region selection based on utility at ages 25–29 is not made and the population born in each prefecture resides in the prefecture until death.

for which the utility function is easily formulated and interpreted. To deepen understanding of the results, we forecasted the population using the estimation results for "amenity 1" under the assumption that regional productivity $\{A_{it}\}_{i\in M}$ would grow at the average growth rate for all regions over the entire period from 1975 to 2015. This assumption eliminates differences in productivity growth rates among regions and states that productivity in all regions grows uniformly. The results under this assumption are shown in Figure 8.

[Figure 8]

Figure 8 shows that while the relationship of population growth rates to base estimates in Figures 6 and 7 was maintained, the decline in the total population was relatively restrained, and the difference in personal income due to differences in productivity accelerated population decline. In other words, in urban areas, the cost of raising children is high, but the expected income is high, whereas in rural areas, the cost of raising children is low, but the income, which is the source of funding for raising children, is also low. As a result, the population growth rate in the country as a whole is lower compared to intuition.

Finally, we describe the differences in the results due to the selection of regional characteristics. Appendix 3 discusses the levels of "amenity" and "social capital" used in estimates and forecasts. The "amenity" level is low in urban areas and high in rural areas, consistent with intuition, but there is a similar trend in capital. This is reflected in the "comparison of population growth rates" in Figures 6 and 7. Excluding prefectures with relatively little change in growth rates compared to the base estimates, the signs and degree of change in growth rates do not differ significantly between Figures 6 and 7. This may be due to the application of a method to select "social capital" related to migration. Since it is assumed that there is a large correlation between social capital stock by prefecture and total factor productivity by prefecture, it is possible to derive estimation results completely different to Figure 7 through the refinement of the formula for the regional labor market (corporate sector) and the structural estimation using social capital indicators. However, the fact that the results did not differ significantly between the assumption of "amenity" and of "social capital" for regional characteristics also shows that the results are robust regarding the choice of regional characteristics and can be interpreted positively.

Through the forecast by projection, the dynamics of the population by five-year age groups in each prefecture were obtained. However, since it is difficult to present all the results, this paper shows in Figure 9 the juvenile generation population (0–24 years old: dark color), the working generation population (25–74 years old: light color), and the elderly generation population (75 years and older: middle color) in the prefectures of Tokyo and Tottori, where the largest divergence in population growth rate was observed in the standard estimation and each simulation of Figures 6–8.

[Figure 9]

In the case of basic estimation, the population of Tokyo declined moderately after reaching its maximum in around 2040, still remaining above 10 million at the end of the estimate.[51] In the case of Tottori Prefecture, on the other hand, the population is predicted to decline sharply from 570,000, and at the end of the estimate, it is almost halved to less than 300,000. By contrast, when region selection is introduced, the young population (aged 25–29) leaves Tokyo, where child-rearing costs are high, and flows into Tottori, where these costs are low. As a result, while the working-generation population in Tokyo decreases, the rapid decline in this population in Tottori that was observed in the standard estimate is restrained. As a result, the Tokyo population declines sharply from around 2040, reaching 8 million to 9 million at the end of the estimate. In Tottori, the population increased until around 2060 through the inflow of the working generation population and the addition of their children to the juvenile generation population, and as of the end, the population remained at 500,000. Comparing the results between the simulations, it can be seen that the effect on the population structure of Tokyo is greater when "social capital" is defined as a regional characteristic (the case of Figure 7) than when "amenity" is defined as a regional characteristic (the case of Figure 6). In addition, when the productivity growth rate is assumed uniformly (the case in Figure 8), the population growth rate in Tokyo is further restrained compared to the case when "amenity" is taken as a regional characteristic. This fact reaffirmed that in the formulation of the quantitative spatial model in this paper, the population growth rate greatly depends on the productivity growth rate.

In this section, we predicted the population envisaged for a century in the future by implementing a projection combining a quantitative spatial model and population dynamics. Counterintuitively, it has been found that the introduction of region selection according to the quantitative space model results in a tendency to select areas where child rearing is easy, while conversely accelerating the decline in the total population of Japan as a whole. The ideal way for policies to curb the decline in Japan's total population based on the above is discussed below. We would also like to analyze which geographic and economic factors caused a decline in Japan's total population and a decline in regional populations. With regard to these matters for examination, this paper uses counterfactual analysis, which is a characteristic of quantitative space models, to analyze various factors that characterize the Japanese economy in the second half of the 20th century, such as (1) a decrease in birthrate, (2) regional disparities in birthrate, (3) improvement of traffic networks, and (4) sluggish productivity.

[51] As mentioned above, the total population is higher than the "Future Estimated Population" (Cabinet Office), reflecting the fact that the mortality rate before age 75 and the detailed structure of the birthrate were abandoned. This trend can also be seen in estimates in Tottori.

5.3 Analysis of population change factors and population policies by counterfactual analysis

As noted above, this section discusses Japan's demographic factors and desirable demographic policies by conducting counterfactual analysis. One feature of quantitative spatial models is counterfactual analysis using structural estimates for equilibrium conditions in multiple regions.

For example, Ahlfeldt et al. (2015) used city block data from before and after the division and reunification of Berlin and found that changes in urban structures observed in land prices and other elements occurred endogenously through the effects of physical division and the restructuring of transport networks and the accumulation and diversification of productivity and regional characteristics. This study shows endogeneity through two counterfactual analyses assuming exogenous regional characteristics and productivity. In the first counterfactual analysis, formal simulations are conducted to replace productivity, regional characteristics, population density, and more in the data after division and reunification with the data before to show that the restructuring of urban structures as a result of segmentation and integration was not brought about only by physical division. The second counterfactual analysis shows that the effects of division and reunification on urban structures were caused by the agglomeration of productivity and regional characteristics rather than by the basic economic conditions of the region, and further clarifies the details of changes in regional structures, such as the relative magnitude of the effects of externalities on productivity and housing and the effects of transport network restructuring.

Bryan and Morten (2015) and Morten and Oliveira (2016) also assessed regional economic structures based on structural estimates. First, Bryan and Morten (2015) pointed out that if Indonesia had had a transport network comparable to that of the United States in the past, its GDP would be 45% to 60% higher than at present, emphasizing the importance of developing transport infrastructure. Morten and Oliveira (2016) also quantitatively assessed the impact of the traffic network associated with the construction of the new capital by deriving and evaluating the migration costs if the traffic network were absent and if the traffic network centered on the former capital were maintained. Further, they explained the effectiveness of regional development policies as due to the improvement of transportation infrastructure, which affects the wage elasticity and trade of the local population.

Based on the analytical methods of these previous studies, we clarify how changes in various parameters in spatial economic models affect regional population dynamics and the total population. The scenarios considered address the following four matters discussed in the context of economic and demographic issues.

The first is a decrease in the birthrate. According to the 2019 Annual Totals of Monthly Reports on Demographic Statistics (approximate figures) (Ministry of Health, Labour and Welfare), the total fertility rate, which was 1.76 as of 1985, was 1.36 as of 2019. It has generally been pointed out that trends towards remaining single, marrying later, and bearing children later are factors in the decrease in the birthrate, but this discussion assumes that the difference between the "ideal number of children" and the "intended number of children" in the 15th Basic Survey on Birth Trends (National Institute of Population and Social Security Research) has improved. In other words, we analyze the case in which the adjusted birth rate $\hat{n}_{jt}$ improved due to a 13% decrease in child-rearing costs in Equation (48) and derive the population dynamics if the birthrate improved.[52] Let this scenario be "(1) increased birthrate." The second analysis deals with regional birthrate disparities. As pointed out in the 15th Basic Survey on Birth Trends, the total fertility rate is highest in Okinawa at 1.82 and lowest in Tokyo at 1.15. One often-mentioned hypothesis in this regard is that low birthrates in heavily populated urban areas and high birthrates in less-populated rural areas result in low birthrates in Japan as a whole. This work deals with scenarios in which only regional birthrate disparities are eliminated by treating variances in regional birthrates $\hat{n}_{jt}$ with mean-preserving contractions at each time point and deriving regional birthrates with halved standard deviations. This scenario is called "(2) improvement of regional disparities in birthrates." The next element we deal with is development of the traffic network. The above-mentioned studies of quantitative spatial models have formulated the development of traffic networks in detail.[53] Some also believe that the development of traffic networks has resulted in greater concentration as a result of reducing travel costs in urban and rural areas. Therefore, in this work, we examine how migration costs affect the population dynamics of prefectures and the population dynamics of regions through use of migration costs in the traffic network as of 1975. This scenario is called "(3) verification of traffic network effects." Finally, we examine the slump in productivity growth. As Hayashi and Prescott (2001) pointed out, one of the factors behind the slump in the Japanese economy since the 1990s is the decline in the growth rate of total factor productivity. Total factor productivity not only determines the calculation of the economy as a whole but also affects the number of births by determining individuals' incomes. If the decline in the income of the younger generation leads to a declining birthrate, an increase in productivity may raise the birthrate. As a final counterfactual analysis, therefore, we analyze the scenario in which the total factor productivity growth rate in each prefecture is the growth rate from 1975 to 1995, not the growth rate from 1975 to 2015. This scenario is called "(4) productivity improvement."

Figures 10–13 show projections up to 2125 under these scenarios: (1) increased birthrate; (2) improvement of regional disparities in birthrates; (3) verification of traffic network effects; and (4) productivity improvement. As in the previous section, the left-hand side of each chart shows the Japan total population forecast, which is an aggregate of the population of each prefecture, and the right-hand side shows each prefecture’s population growth rate over the estimation period (2015–2125).

---

[52] In the 15th Basic Survey on Birth Trends, the "ideal number of children" was 2.32, the "intended number of children" was 2.01, and the scope of the decline in child-rearing expenses corresponds to this. If the divergence between the two indicates an obstacle to child rearing, this scenario can be regarded as indicating the greatest theoretically expected effect of child-rearing support alone.

[53] Trading costs are the most important factor in this field because of the history of the establishment of new trade theory and new economic geography, which form the basis of the quantitative space model. See Krugman (1980), etc.

[Figure 10]

[Figure 11]

[Figure 12]

[Figure 13]

Figure 10 analyzes scenario (1) increased birthrate. When the obstacles to child rearing are eliminated, the effect is large; besides the total population increasing compared to the standard estimate, population decline is suppressed in all prefectures except Tokyo compared to the standard estimate without migration. This is because the "number of children" is reflected in the region selection by the young population; thus, the population is flowing into districts where the density is low. Although the population decline in Tokyo is larger than in the standard estimate, the extent of the decline has been reduced compared with the case where the regional characteristics shown in Figure 6 are defined as "Amenity 1," reflecting the rise in the birthrate. Next, we analyze scenario (2) improvement of regional disparities in birthrates, described in Figure 11. In this case, where the standard deviation of birthrates is assumed to be half, there is little change in the total population dynamics and population growth rate, except that the total population decline in 2125 is slightly suppressed (about 500,000) compared with the amenity shown in Figure 6. Even if regional differences in birthrates improve exogenously, as shown, there is little effect on population dynamics. Therefore, it can be confirmed that the impact of regional birthrate disparities on the total population is not very large.

Figure 12 is an analysis of the effects of traffic network development. Although the decline in the total population is approximately 1 million people smaller than in the case in Figure 6, there is little change. The major change here is the regional population growth rate; in addition to the population decline turning to growth in Yamanashi and Tottori, the population growth rate has improved in all prefectures except Tokyo compared to the base estimate. The development of traffic networks will increase the young population migrating from areas with high (expected) indirect utility to areas where it is lower by reducing travel costs. From these facts, we can confirm simultaneously that this effect has caused population concentration in urban areas and population outflow from rural areas, and that it will continue to be a cause of population decline in rural areas in the future. It should be noted, however, that one of the conclusions of new economic geography is that the development of traffic networks will promote interregional trade and improve the productivity and welfare of the economy as a whole.[54] It is also known that the development of traffic networks has little impact on the future decline in total population, and therefore caution must be exercised in assessing the negative effects

[54] The model in this paper assumes that social welfare in the absence of traffic network development will be reduced by the higher travel cost, since the net utility of residing in the region is obtained from the difference between the indirect utility and the travel cost. Social welfare is lower than in the standard estimate and the case of Figure 6 because the above-mentioned trends in the corporate sector are also conceivable.

of traffic networks development. Finally, we examine the effects of productivity growth improvement in Figure 13. Compared with Figure 6, the decline in the total population has been substantially restrained, almost to the same level as the standard estimate without an assumption of population migration. In addition, although not as much as in the case of the transport network development effect, the population growth rate in rural areas is improving, and the population growth rate is high in most prefectures, mainly in rural areas, compared to the standard estimate. This is because the number of children increased in each prefecture due to an increase in disposable income resulting from increased productivity in rural areas where the population is low and child rearing is comparatively easy. In Figure 8, the productivity growth rate was assumed to be the national average, but recalling that even in this case, the total population and the rural population growth rate showed improvement, it can be said that both the large variation in the growth rate of total factor productivity across prefectures and the sluggishness in the growth rate itself contributed to the decline in Japan's total population.

As in the previous section, we illustrate the time series of population by age group in Tokyo and Tottori, which fluctuate more than the standard estimate. Figure 14 shows the population dynamics of Tokyo and Tottori in response to the following scenarios: (1) increased birthrate; (2) improvement of regional disparities in birthrate; (3) verification of traffic network effects; and (4) productivity improvement.

[Figure 14]

In the case of the increased birthrate, the improvement in birthrate restrains the decline in the juvenile generation population, and in particular, the movement of the young population to Tokyo increases. As a result, the population growth rate improves mainly in the juvenile generation population and the working generation population, compared with Figure 6, in which the other conditions are identical. In the case of improvement of regional disparities in birthrates, the decline in Tokyo’s population has somewhat increased, while Tottori’s population dynamics remain mostly unchanged. This may be attributed to the fact that the indirect utility of individuals in regions with high child-rearing costs has decreased by compulsorily compensating for the divergence between "ideal number of children" and "intended number of children." This has a negative impact on Tokyo’s total population by preventing the inflow of the young generation population.

Looking at the case of verification of traffic network effects, it can be seen that the total population growth rates of Tokyo and Tottori have improved compared with the case of Figure 6. With regard to Tottori, the fact that the outflow of the young population to urban areas was restrained by the lack of improvement of the traffic network seems to be a factor in the improvement. On the contrary, the population decline in Tokyo can be seen to have been halted because the population flowing out is restrained by avoiding the high cost of raising children. Finally, we observe the effect of productivity improvement on regional population. In this scenario, disposable income increases in each region as the region is not affected by the decline in total factor productivity growth rate in the second half of the estimation period (1996–2014). This increases the number of children, particularly in rural areas where child-rearing costs were low, but income was also low. This effect is still exerted in Tokyo as well; therefore, the juvenile generation population is seen to be larger in both regions than in the standard estimate.

In this section, we examined the factors behind Japan's population decline and countermeasures against population decline by describing the future dynamics of Japan's total population and regional population, especially when the obstructors to economic and demographic growth seen in Japanese society since the 1990s did not exist or were ameliorated. As a result of the examination, in the range of the formulations in this paper, abandonment and delay of childbirth due to various obstructors to birth selection exerted a greater effect than regional differences in birthrate, and it was shown that improvement of the traffic network affected the population dynamics, such as the number of births, while improving economic welfare. Regarding the total factor productivity growth rate, it was also found that both the sluggish growth of the growth rate itself and regional disparities in the growth rate worsened population dynamics.

6. Conclusion

In this paper, we first formulated and estimated a quantitative spatial model and predicted prefectural populations based on the estimates, taking into account the sustainability of regional economies in recent years. In preparing the quantitative space model, we formulated migration that could affect the sustainability of regional economies and regional populations, heterogeneity of regional economies, and the relationship between regional economies and population structures. Two-stage estimation was carried out within the framework of the quantitative space model. First, migration between prefectures was estimated in reduced form using two types of statistics, and the fixed effect on each point in time of each destination prefecture was used as the expected utility of migrating to the region. Furthermore, under the assumption that each individual makes a region selection once in his or her life and maximizes his or her utility, we formulated and made structural estimates of utility using a discrete selection model. Second, in the population forecast by prefecture, the parameters on migration and utility function obtained from the first-stage estimation were applied to simple population dynamics, and the population dynamics from 2015 to 2125 were derived by projection.

In the first quantitative spatial model estimation, we were able to obtain highly significant estimation results for analysis of long-term migration in Japan and for analysis of increases in housing prices and child-rearing costs due to population inflow. This result enabled us to formulate and analyze the relationships between migration, heterogeneity of regional economies, and regional economies and demographic structures. In the second population dynamics forecast, regional population projections were made under certain assumptions concerning regional characteristics and regional productivity, and counterfactual analyses were conducted on various factors of the declining birthrate and aging population. As the result, it was shown that various obstructors of birth selection affected the population structure of regions rather than regional differences in birthrate, and that development of the traffic network affected the population dynamics such as the number of births while also improving economic welfare. Regarding productivity, it was found that both regional differences in productivity and slowing growth rates restrained the population.

The contribution of this paper is described by field. While conventional studies have formulated regional differences in population structures, such as birthrates, under spatial equilibrium models, this paper uses a quantitative spatial model to directly analyze the relationship between migration and birth. In the spatial equilibrium model, the economy is said to be moving towards equilibrium and population migration towards disappearance since the expected utilities of individuals in every region are equalized in the equilibrium, whereas in the formulation in this paper, the disparity in expected utility of individuals derived from regional economic conditions does not necessarily disappear, enabling analysis of population migration. In regional economic analyses as well, the author is not aware of any existing studies that apply to Japanese cases Allen and Arkolakis's (2014) quantification of geographic travel costs or the formulation of quantitative space models following Ahlfeldt et al. (2015) and Diamond (2016), and that significant results on the relation between transportation networks and demographics were obtained is also a contribution of this paper. In addition, in the population estimation, it is considered to be an achievement that an economic foundation was given to the social and natural changes in population.

After pointing out the elements that could not be dealt with completely in this paper, we describe research tasks for the future. The first is the dynamization of the problem and the introduction of an overlapping generation structure. In this paper, we assume that individuals residing in different regions have made region selections by migration at the same age, once in their lifetime. In the real world, however, individuals make multiple migration choices at different times. Because Kennan and Walker (2011) formulated migration by young groups based on dynamic optimization issues, while making assumptions about region selection, it may be helpful to extend the model on the basis of these studies. Many of the long-term financial analyses, including Auerbach and Kotlikoff (1987), have an overlapping generation structure. In analyzing the fiscal sustainability of regional economies and the interrelationship between fiscal and population structure, although it is desirable to introduce an overlapping generation structure, there are many issues such as how to set regional capital levels, savings levels, interest rates, and other requirements.

Second, a detailed formulation of the industrial sector is also an issue to be considered. In an advanced study of the quantitative spatial model of Ahlfeldt et al. (2015), industrial agglomerations were analyzed following new economic geography. Instead of considering these agglomerations, this paper derived the productivity growth rate of a region. On the other hand, it can be said that more elaborate general equilibrium analyses open up the possibility of deeper discussion of the relationship between productivity and population, as well as analyses of economic growth and welfare, such as the balance between accumulation and dispersion effects. Finally, a refinement of population dynamics estimates can be raised. In the formulation of this paper, we discard the mortality rate up to the working generation and the age structure of the birthrate for population dynamics as well as for population migration. In addition to refining the optimization problem, it is also envisioned that these derivations will allow for more elaborate demographic projections.

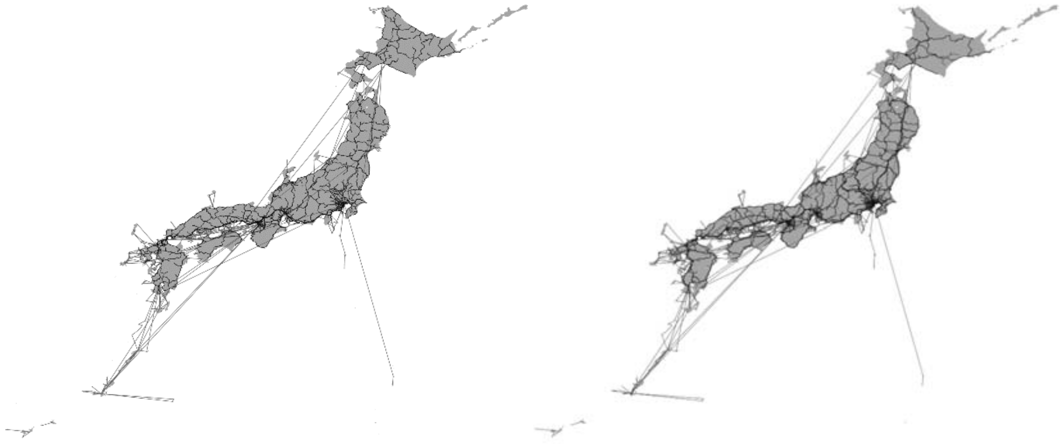
References

1. Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015) "The Economics of Density: Evidence from the Berlin Wall," Econometrica, Vol.83, No. 6, pp.2127-2189.
2. Albouy, D. (2013) "Are Big Cities Bad Place to Live? Estimating Quality of Life across Metropolitan Areas," NBER Working Paper w14472, National Bureau of Economic Research, Cambridge, MA.
3. Albouy, D. (2016) "What Are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities," Review of Economics and Statistics, Vol. 98, Issue. 3, pp.447-487.
4. Allen, T. and Arkolakis, C. (2014) "Trade and the Topography of the Spatial Economy," Quarterly Journal of Economics, Vol. 129, No. 3, pp.1085-1140.
5. Auerbach, A. J., and Kotlikoff, L. J. (1987) "Dynamic Fiscal Policy," Cambridge University Press, Cambridge.
6. Baldwin, R. E., Martin, P, Ottaviano, G. and Robert-Nicoud, F. (2003) "Economic Geography and Public Policy," Princeton University Press, Princeton, NJ.
7. Bartik, T. J. (1991) "Who Benefits from State and Local Development Policies?" Upjohn Press.
8. Börsch-Supan, A., Ludwig, A. and Winter, J. (2006) "Ageing, Pension Reform and Capital Flows: A Multi-Country Simulation Model," Economica, Vol. 73, No. 292, p.625-658.
9. Braun, R. A, and Joines, D. H. (2015) "The implications of a graying Japan for government policy," Journal of Economic Dynamics and Control, Vol. 52, pp.1-23.
10. Brueckner, J. K. and Neumark, D. (2014) "Beaches, Sunshine, and Public Sector Pay: Theory and Evidence on Amenities and Rent Extraction by Government Workers," American Economic Journal: Economic Policy, Vol. 6, No.2, pp.192-230.
11. Bryan, G. and Morten, M. (2015) "Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia," Working paper.
12. De La Croix, D. and Gobbi, E. P. (2017) "Population Density, Fertility and Demographic Convergence in Developing Countries," Journal of Development Economics, Vol. 127, pp.13-24.
13. Diamond, R. (2016) "The Determinants and Welfare Implications of US Worker's Diverging Location Choices by Skill," American Economic Review, Vol. 106, No. 3, pp.479-524.
14. Diamond, R. (2017) "Housing Supply Elasticity and Rent Extraction by State and Local Governments," American Economic Journal: Economic Policy, Vol. 9, No.1, pp.74-111.
15. Kroon, D.J. (2009) "Accurate Fast Marching," (https://www.mathworks.com/matlabcentral/fileexchange/24531-accurate-fast-marching), MATLAB Central File Exchange.
16. Eaton, J. and Kortum, S. (2002) "Technology, Geography, and Trade," Econometrica, Vol. 70, No. 5, pp.1741-1779.
17. Fougère, M., Harvey, S., Mercenier, J. and Mérette, M. (2009) "Population Ageing, Time Allocation, and Human Capital: A general equilibrium analysis for Canada," Economic Modeling, Vol. 26, pp 30-39.
18. Fujita, M., Krugman, P., and Venables, A. J. (1999) "The Spatial Economy. Cities, Regions and International Trade", MIT Press, Cambridge, MA.

19. Hayashi, F. and Prescott, E. C. (2002) "The 1990s in Japan: A Lost Decade," Review of Economic Dynamics, Vol.5, pp.206-235.
20. Ishida, R., Oguro, K., and Yasuoka, M. (2018) "Population density, fertility and childcare services from the perspective of a two-region overlapping generations model," Economic Analysis and Policy, Vol.59, pp.29-39.
21. Kato, H. (2017) "Gap in Birthrate by Municipality and Analysis for the Underlying Factors," Public Policy Review, Vol.14, No. 1, pp.1-24.
22. Kennan, J. and Walker, J.R. (2011) "The Effect of Expected Income on Individual Migration Decisions," Econometrica, Vol.79, No.1, pp.211-251.
23. Krugman, P. (1980) "Scale Economies, Product Differentiation, and the Pattern of Trade," The American Economic Review, Vol.70, No.5, pp.950-959.
24. Krugman, P. (1991) "Increasing Returns and Economic Geography," Journal of Political Economy, Vol.99, No.3, pp. 483-499.
25. McFadden, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior," in Zarembka, P. (Ed.), Frontiers in Econometrics, Academic Press, New York, pp.105-142.
26. Minamimura, K. and Goto, H. (2018) "Geography and Demography: New Economic Geography with Endogenous Fertility," The Japanese Economic Review, forthcoming.
27. Moretti, E. (2011) "Local Labor Markets," in Card, D. and Ashenfelter, O. (Ed.), Handbook of Labor Economics, Vol.4b, North-Holland, Amsterdam.
28. Morten, M. and Oliveira, J. (2016) "Paving the Way to Development: Costly Migration and Labor Market Integration," NBER Working Paper 22158, National Bureau of Economic Research, Cambridge, MA.
29. Morita, T. and Yamamoto, K. (2018) "Interregional Fertility Differentials and Agglomeration," The Japanese Economic Review, Vol. 69, No. 2, pp.171-188.
30. Redding, S. J. and Rossi-Hansberg, E. (2017) "Quantitative Spatial Economics," Annual Review of Economics, Vol. 9, pp.21-58.
31. Roback, J. (1982) "Wages, Rents and the Quality of Life," Journal of Political Economy, Vol. 90, pp.1257-1278.
32. Rosen, S. (1979) "Wages-based Indexes of Urban Quality of Life," in Miezskowski, P and Straszheim, M. (Ed.), Current Issues in Urban Economics, John Hopkins University Press.
33. Saiz, A. (2010) "The Geographic Determinants of Housing Supply," The Quarterly Journal of Economics, Vol. 125, No. 3, pp.1253-1296.
34. Sato, Y. and Yamamoto, K. (2005) "Population Concentration, Urbanization, and Demographic Transition," Journal of Urban Economics, Vol. 58, pp.45-61.
35. Sato, Y. (2007) "Economic Geography, Fertility, and Migration," Journal of Urban Economics, Vol. 61, No. 2, pp.372-387.
36. Tokui, J., Makino, T., Fukao, K., Miyagawa, T., Arai, N., Arai K., Inui, T., Kawasaki, K., Kodama, N., and Noguchi, Y. (2013) "Compilation of the Regional-Level Japan Industrial Productivity Database(R-JIP) and Analyses of Productivity Differences across Prefectures, *RIETI Discussion Paper Series* 13-J-03, Research Institute of Economy, Trade and Industry.

Charts

Figure.1: Japan's transportation network in 1975 and 2015



(a) In 1975 **(b) In 2015**

NOTE) Created by the author using ArcGIS10 and QGIS3 using "Numerical Information on Land" (Ministry of Land, Infrastructure, Transport and Tourism).

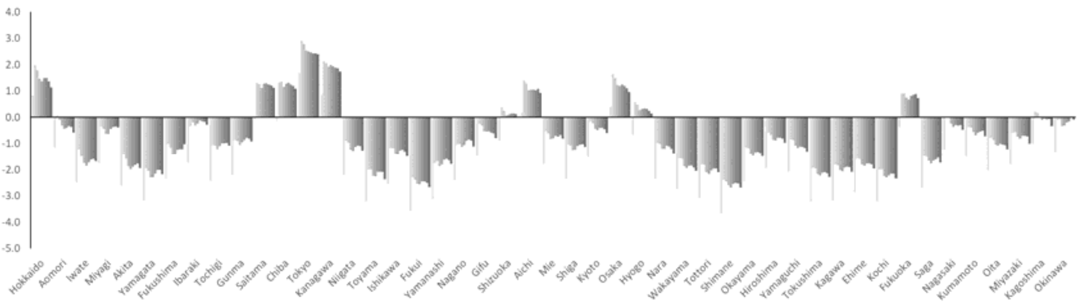
Figure.2: Outline of the transportation network improvement plan in Japan

(a) Shinkansen Development Project **(b) High Standards Trunk Road Network Project**

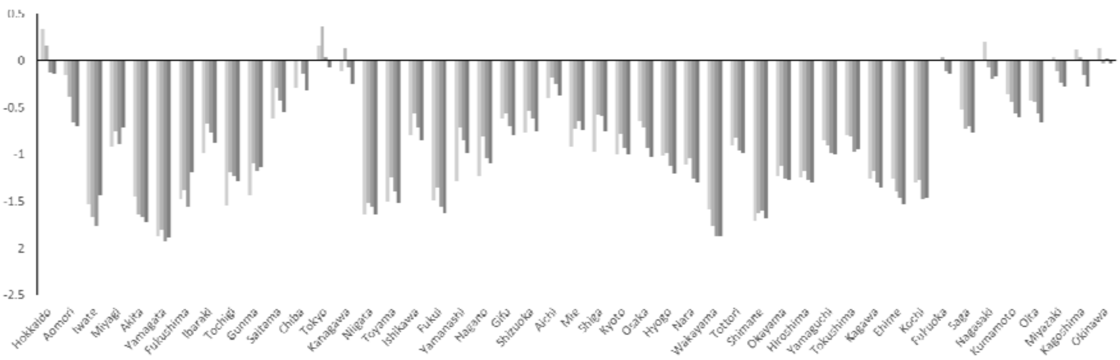
NOTE) Prepared by the author using ArcGIS10 and QGIS3 using data provided by the Ministry of Land, Infrastructure, Transport and Tourism. "Shinkansen improvement plan" is a "high standard trunk road network plan" which the author draws based on each improvement plan drawing.

Figure.3: Regional Indirect Utility Indicators (First-Stage Estimation Results)

1. Indirect Utility Indicators Based on the Basic Resident Register Population (1975-2015)

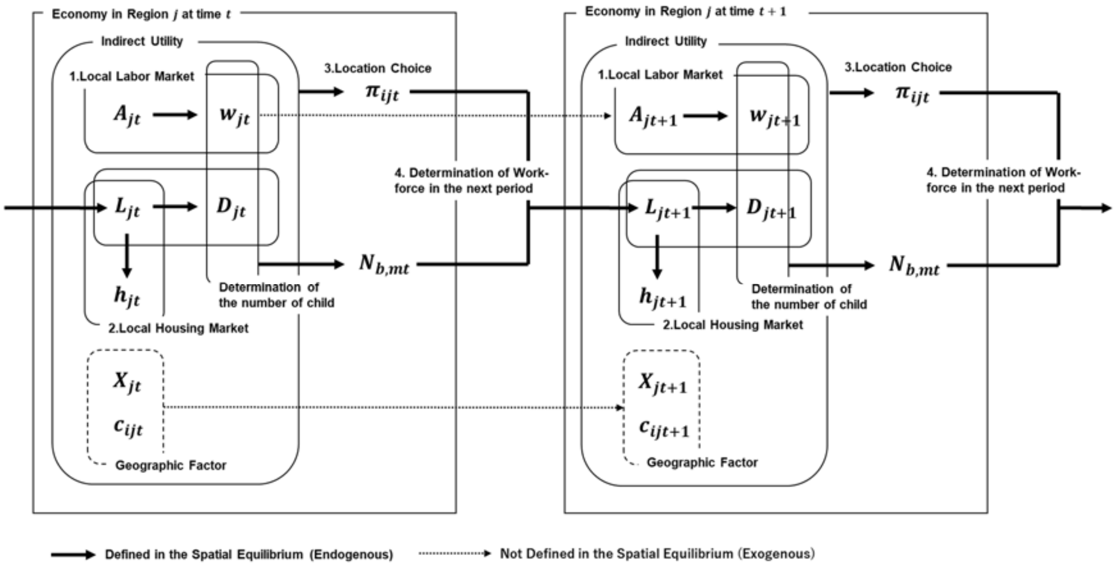


2. Indirect Utility Indicators Based on Census Population Movements (1990-2015)



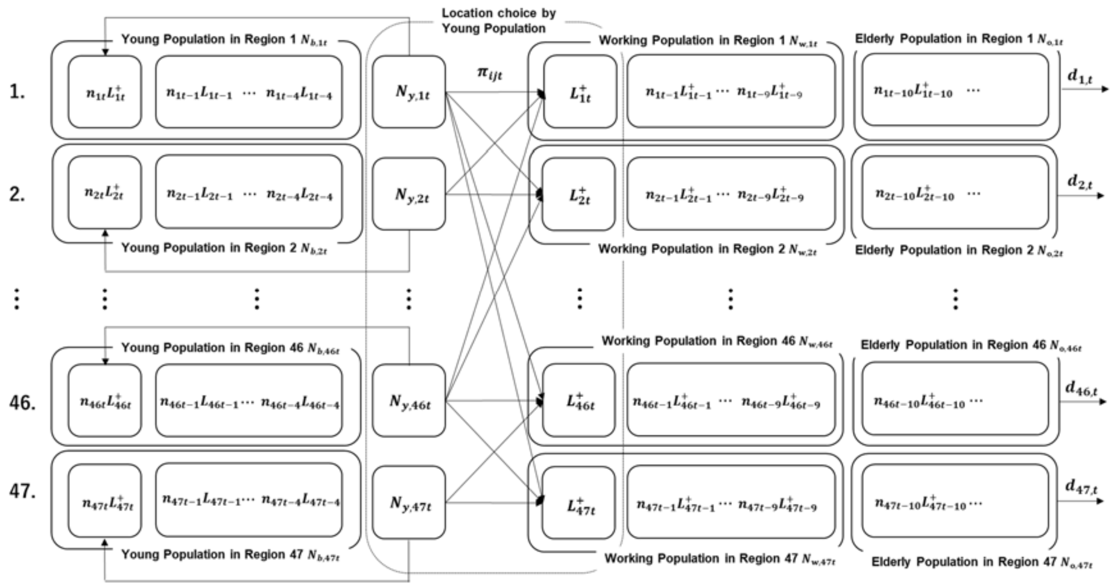
NOTE) These tables describe the results of the first-stage guidance system estimation. Regional indirect utility indicators correspond to destination fixed effects in Equation (23). Panel data estimation using the two-step least-squares method using the development plan formulated by the Japanese government as an operational variable is being implemented.

Figure.4: Structure of the spatial equilibrium model



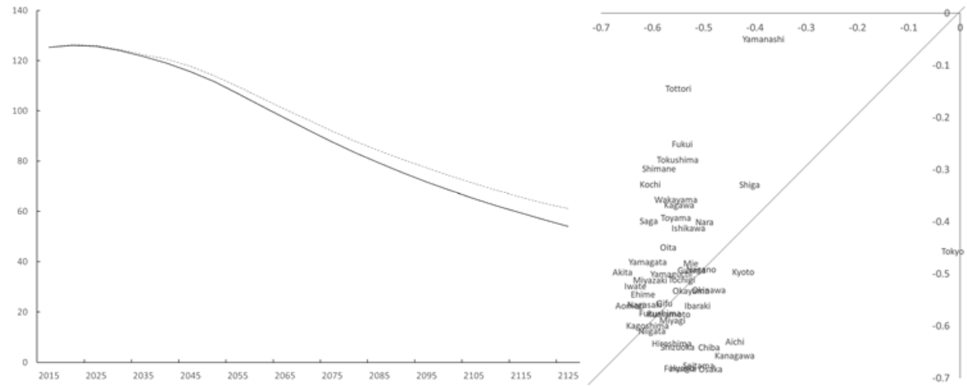
NOTE) This figure was made from the theoretical model of this paper.

Figure.5: Schematic diagram of demographic dynamics



NOTE) This figure was made from the theoretical model of this paper.

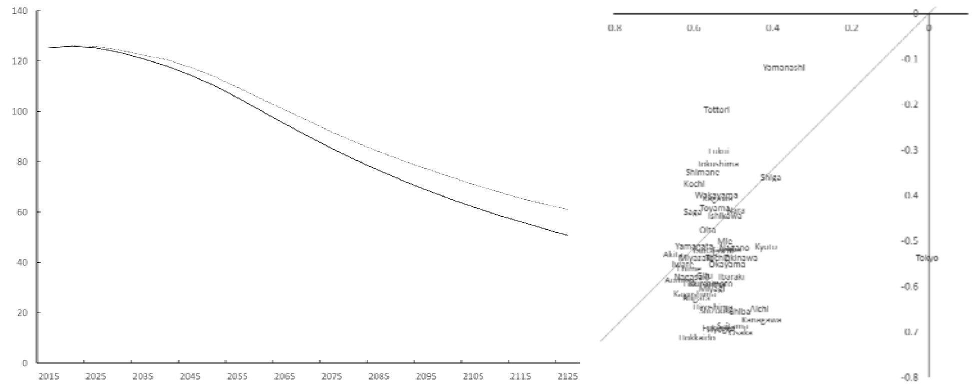
Figure.6: Simulation Results (1) Population Dynamics (Amenity)



(a) Total Japan Population Forecast (2015-2125) **(b) Comparison of population growth rate**

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

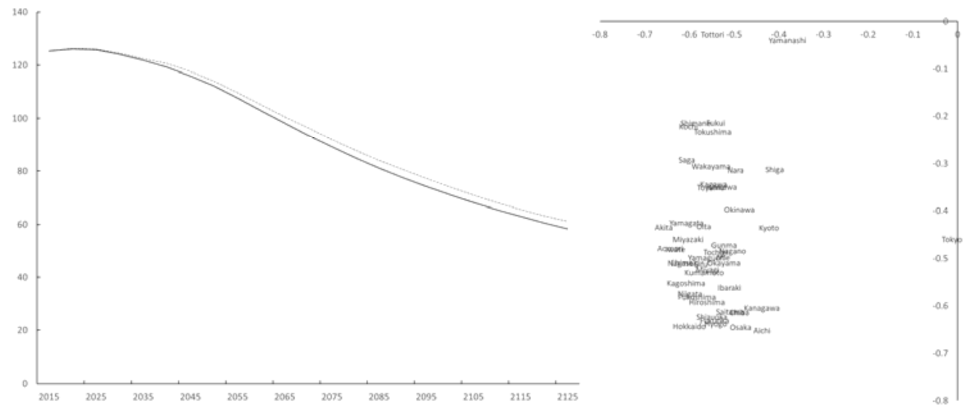
Figure.7: Simulation Results (2) Population Dynamics (Social Infrastructure)



(a) Total Japan Population Forecast (2015-2125) **(b) Comparison of population growth rate**

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

Figure.8: Simulation Results (3) When Productivity Growth Rates Are Uniform (Amenity)

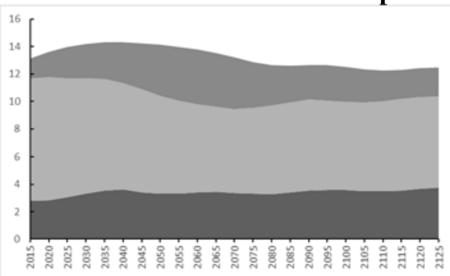


(a) Total Japan Population Forecast (2015-2125) **(b) Comparison of population growth rate**

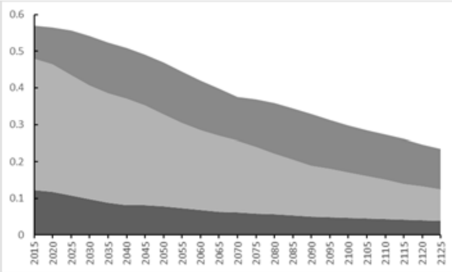
NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

Figure.9: Simulation Results (1-3) Comparison of Regional Populations (Tokyo and Tottori Prefectures)

1. Baseline Estimate-Without Population Movements

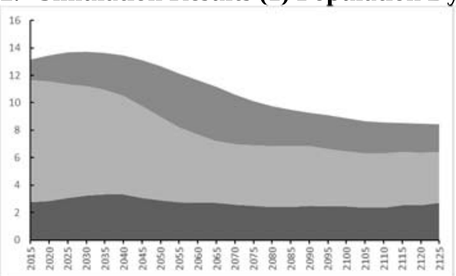


(a) Tokyo

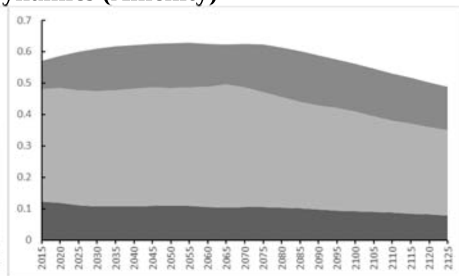


(b) Tottori Prefecture

2. Simulation Results (1) Population Dynamics (Amenity)

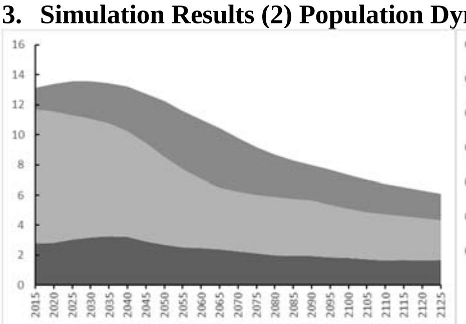


(b) Tokyo

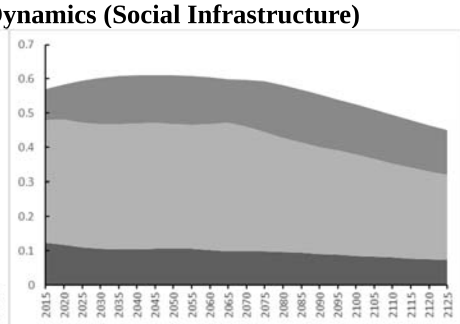


(b) Tottori Prefecture

3. Simulation Results (2) Population Dynamics (Social Infrastructure)

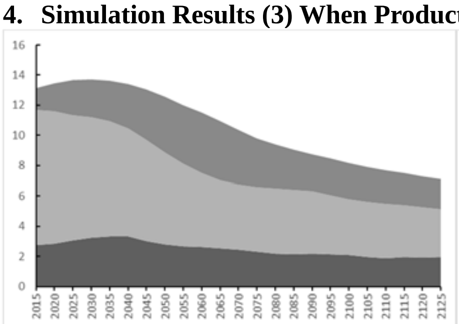


(a) Tokyo

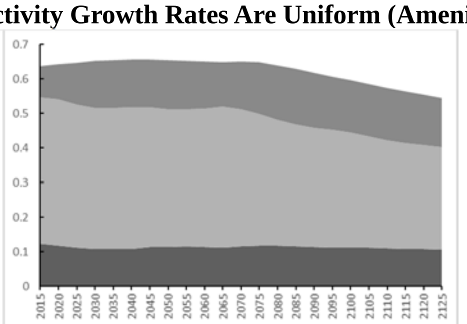


(b) Tottori Prefecture

4. Simulation Results (3) When Productivity Growth Rates Are Uniform (Amenity)



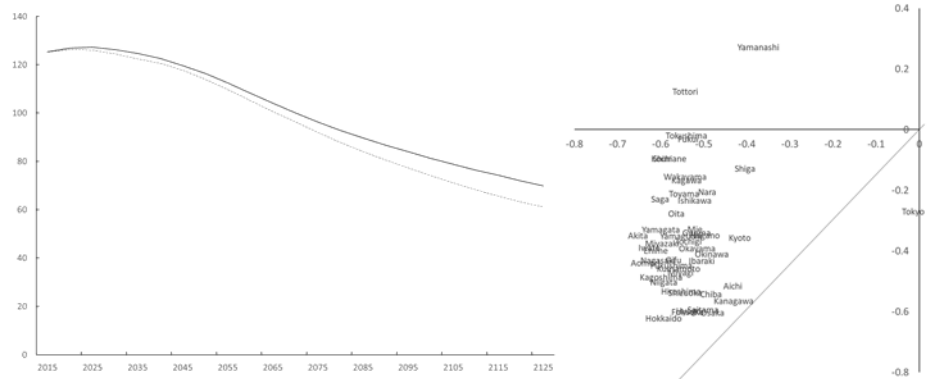
(a) Tokyo



(b) Tottori Prefecture

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

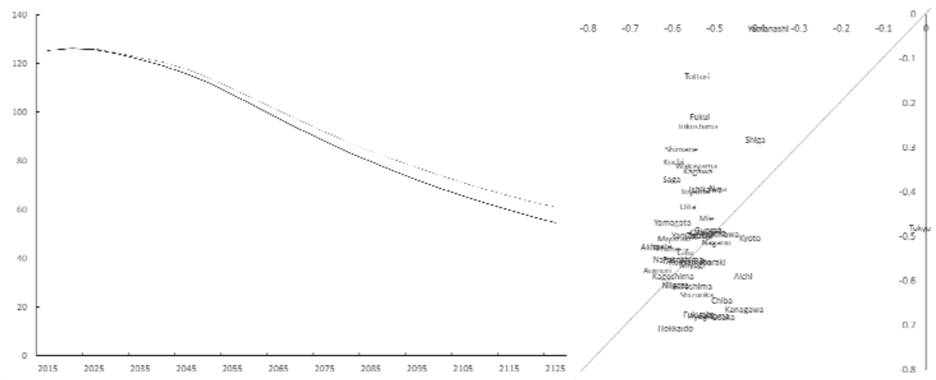
Figure.10: Simulation Results (4) Counterfactual Analysis (1): Increased Fertility Rate



(a) Total Japan Population Forecast (2015-2125) (b) Comparison of population growth rate

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

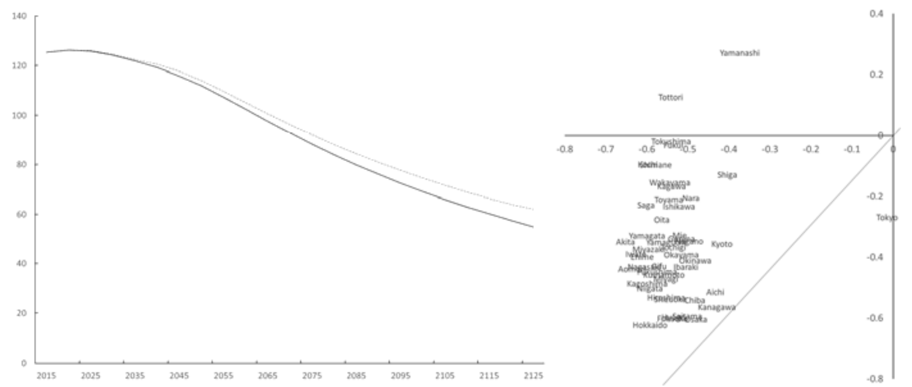
Figure.11: Simulation Results (5) Counterfactual Analysis (2):
Improvement of regional disparities in fertility rates



(a) Total Japan Population Forecast (2015-2125) (b) Comparison of population growth rate

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

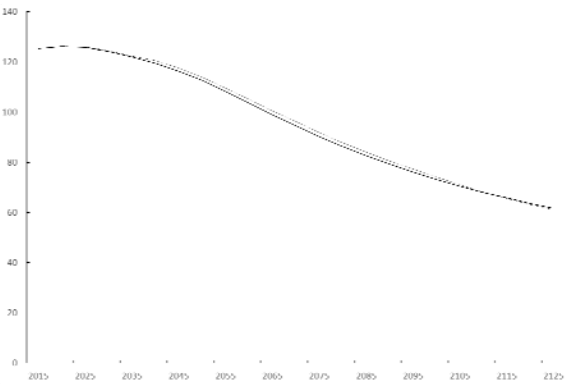
Figure.12: Simulation Results (6) Counterfactual Analysis (3):
Traffic Network Effectiveness Verification

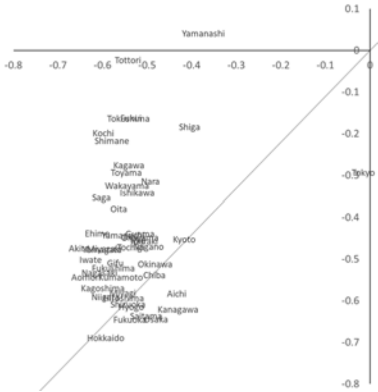


(a) Total Japan Population Forecast (2015-2125) (b) Comparison of population growth rate

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

Figure. 13: Simulation Results (7) Counterfactual Analysis (4): Productivity Improvement



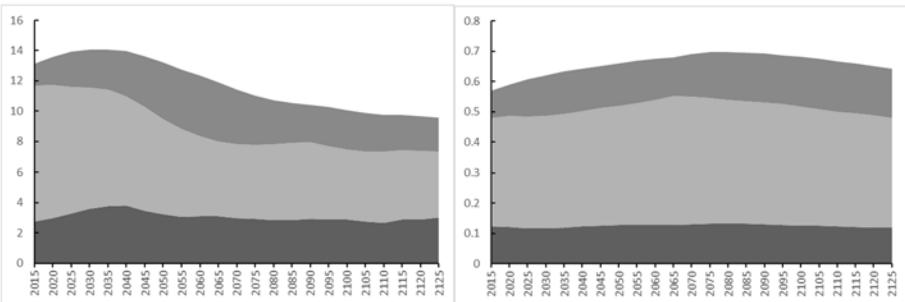


(a) Total Japan Population Forecast (2015-2125) **(b) Comparison of population growth rate**

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

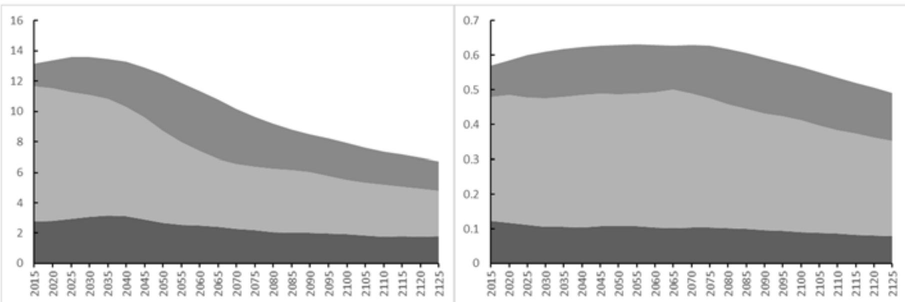
Figure. 14-1: Simulation Results (4-7) Comparison of Regional Population (1): (Tokyo and Tottori Prefectures)

1. Simulation Results (4) Counterfactual Analysis (1): Increased Fertility Rate



(a) Tokyo **(b) Tottori Prefecture**

2. Simulation Results (5) Counterfactual Analysis (2): Improvement of regional disparities in fertility rates

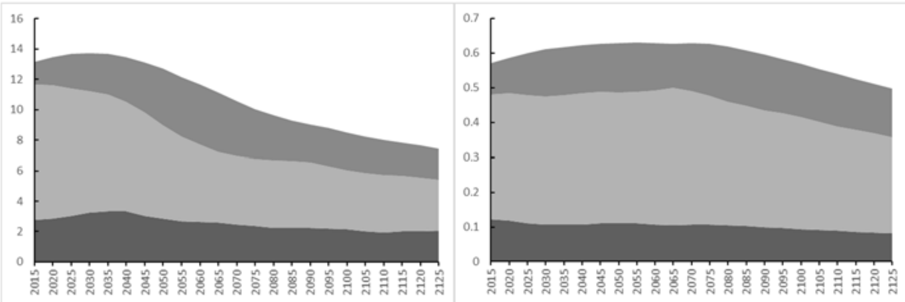


(a) Tokyo **(b) Tottori Prefecture**

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

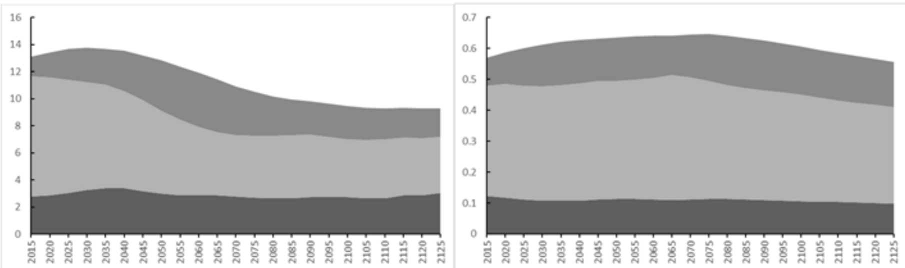
Figure. 14-2: Simulation Results (4-7) Comparison of Regional Population (2) (Tokyo and Tottori Prefectures)

3. Simulation Results (6) Counterfactual Analysis (3): Traffic Network Effectiveness Verification



(a) Tokyo **(b) Tottori Prefecture**

4. Simulation Results (7) Counterfactual Analysis (4): Productivity Improvement



(a) Tokyo **(b) Tottori Prefecture**

NOTE) The author's work is based on the spatial economic dynamics and demographic dynamics of this paper.

Table. 1: Descriptive statistic for each variable in the first-stage estimation

Estimation of "Basic Resident Register Population"

	Number of movements	Transfer probability	Distance	Travel time	Shinkansen Project	Highway plan
Max	**11264664**	**0.9552**	**1.0000**	**1.0000**	**1.2318**	**0.7479**
Min	**19**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
Mean	**55469**	**0.0213**	**0.2400**	**0.1345**	**0.1714**	**0.1508**
Median	**1576**	**0.0008**	**0.2060**	**0.1020**	**0.1276**	**0.1257**
S. Deviation	**442383**	**0.1282**	**0.1669**	**0.1301**	**0.1733**	**0.1153**

Estimation of "Census Population Movement Aggregation"

	Number of movements	Transfer probability	Travel distance	Travel time
Max	**12607447**	**0.9757**	**1.0000**	**0.8996**
Min	**14**	**0.0001**	**0.0000**	**0.0000**
Mean	**57231**	**0.0213**	**0.2400**	**0.1290**
Median	**828**	**0.0004**	**0.2060**	**0.1276**
S. Deviation	**502962**	**0.1362**	**0.1669**	**0.1257**

NOTE) The "number of movements" indicates the number of inhabitants transferred from one prefecture to another. The "movement probability" indicates the movement probability of the inhabitant to each prefecture when the prefecture of the transfer source is given. In the "Basic Resident Register Population Estimate", these values are derived using the 1975-2015 "Basic Resident Register Population Movement Report Annual Report" (Ministry of Internal Affairs and Communications). In the "Census Population Migration Aggregation", these values are derived using the "Census: Population Migration Aggregation" for 1990, 2000, 2010 and 2015 (MIC). "Distance" is the Euclidean distance between prefectural capitals, and "travel time" is the estimated travel time calculated every five years by Fast Marching Method (FMMs) using various data from 1975-2015 in the National Land Numerical Information (MLIT). "Shinkansen Project" and "Expressway Project" are calculated by FMM in the same way as "Travel Time" using the geographic data of the traffic network planned by "National Shinkansen Railway Improvement Act" (1970-1973) and "National Land Development Trunk Expressway Construction Act" (1962-1987).

Table. 2: Results of the first-stage estimation

Estimation results based on "Basic Resident Register Population" (1975-2015)

	OLS	IV-1	IV-2	IV-3	Panel	PanelIV-1	PanelIV-2	PanelIV-3
Movable fixed cost dummy	**-5.8824 *****	**-5.5013 *****	**-5.5932 *****	**-5.7640 *****	**-5.8824 *****	**-3.8356 *****	**-4.4133**	**-3.8356 *****
	(0.1087)	**(0.0316)**	**(0.0312)**	**(0.0309)**	**(0.1087)**	**(0.0821)**	**(79.787)**	**(0.0821)**
Same local dummy	**2.4735 *****	**1.0943 *****	**1.1496 *****	**1.2523 *****	**2.4735 *****	**0.4258 *****	**1.0037**	**0.4258 *****
	(0.1341)	**(0.0160)**	**(0.0157)**	**(0.0156)**	**(0.1341)**	**(0.1020)**	**(79.824)**	**(0.1020)**
Trend	**0.0322 *****	**1.1888**	**1.1260 *****	**1.0092 *****	**0.0149**	**-0.1210 ****	**-0.0826**	**-0.1210 ****
	(0.0072)	**(0.0164)**	**(0.0161)**	**(0.0159)**	**(0.0487)**	**(0.0491)**	**(5.2985)**	**(0.0491)**
Travel time	**0.9344 *****	**-11.4353 *****	**-10.8094 *****	**-9.6459 *****	**0.9344 *****	**-3.4402 *****	**-2.2055**	**-3.4402 *****
	(0.1998)	**(0.0908)**	**(0.0870)**	**(0.0845)**	**(0.1998)**	**(0.1012)**	**(170.56)**	**(0.1012)**
Constant term					**0.1555**	**1.1260 *****	**0.8519**	**1.1259 *****
					(0.4173)	**(0.4221)**	**(37.828)**	**(0.4221)**
Origin-Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Destination-Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Origin-Destination FE	**Yes**	**No**	**No**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
Observation	19881	19881	19881	19881	19881	19881	19881	19881
F-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R-Squared	0.9995	0.8846	0.8860	0.8865	0.9919	0.9917	0.9918	0.9917

Estimation Results by "Population Migration Aggregation in National Census" (1990-2015)

	OLS	IV-1	IV-2	IV-3	Panel	PanelIV-1	PanelIV-2	PanelIV-3
Movable fixed cost dummy	**-6.8194 *****	**-6.2359 *****	**-6.3291 *****	**-6.4959 *****	**-6.8194 *****	**-7.5394**	**-4.8156 *****	**14.3279**
	(0.3167)	**(0.0447)**	**(0.0442)**	**(0.0438)**	**(0.3017)**	**(399.7242)**	**(0.1153)**	**(423.74)**
Same local dummy	**2.3295 *****	**1.1098 *****	**1.1663 *****	**1.2674 *****	**2.3295 *****	**3.0497**	**0.3249 ****	**-13.0277**
	(0.3369)	**(0.0226)**	**(0.0223)**	**(0.0221)**	**(0.3210)**	**(399.8807)**	**(0.1417)**	**(196.41)**
Trend	**0.0524 ***	**2.7219 *****	**2.5716 *****	**2.3025 *****	**0.1325 *****	**0.1808**	**-0.0018**	**8.5402**
	(0.0316)	**(0.0426)**	**(0.0415)**	**(0.0409)**	**(0.0490)**	**(26.7953)**	**(0.0453)**	**(196.48)**
Travel time	**0.9335**	**-11.8149 *****	**-11.1537 *****	**-9.9700 *****	**0.9335**	**2.4867**	**-3.3897 *****	**0.5487**
	(0.6503)	**(0.1333)**	**(0.1276)**	**(0.1239)**	**(0.6292)**	**(862.3998)**	**(0.1468)**	**(13.166)**
Constant term					**-0.3204 ***	**-0.4521**	**0.0461**	**-1.4559**
					(0.1838)	**(73.1103)**	**(0.1783)**	**(35.924)**
Origin-Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Destination-Year FE	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Origin-Destination FE	**Yes**	**No**	**No**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
Observation	19881	19881	19881	19881	19881	19881	19881	19881
F-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R-Squared	0.9997	0.9019	0.9031	0.9038	0.9944	0.9944	0.9944	0.9940

NOTE) Standard errors are shown in parentheses. Dependent variables are migration probabilities (upper) calculated from the 1975-2015 Basic Resident Register Population Movement Report, and migration probabilities (lower) calculated from the 1990-2010 Census: Population Movement Aggregate. See Section 4.1 for an explanation of "fixed moving costs" and "the same local dummy." All estimation equations include year-destination dummies as an indicator of year-by-year regional fixed effects, but the values are omitted. In IV and PanelIV, a two-step least-squares method is used. In the first step, the "travel time" is regressed to the "distance" and other instrumental variables. In IV-1 and PanelIV-1, only "distance" is regressed. "Distance" and "Shinkansen Project" are used in IV-2 and PanelIV-2, while "Distance" and "Expressway Project" are used in IV-3 and PanelIV-3. Panel and PanelIV also use random-effects modeling to introduce fixed-effect dummies.

Table. 3: Estimation Results of Auxiliary Estimation Equations (1: Birth Rate Model)

	OLS				XT			
	Basic estimation		Extended estimation		Basic estimation		Extended estimation	
Log(density)	**0.0050**	***	**-0.0136**	***	**0.0146**	***	**-0.0045**	
	(0.0017)		**(0.0030)**		**(0.0052)**		**(0.0096)**	
Marriage			**0.0457**	***			**0.0861**	***
			(0.0054)				**(0.0129**	
Log(PDP)			**-0.0478**	***			**-0.3086**	***
			(0.0004)				**(0.0515)**	
Constant	**-3.5118**	***	**-3.5982**	***	**-3.5671**	***	**-3.5856**	***
	(0.0132)		**(0.0181)**		**(0.0277)**		**(0.0823)**	
Observation	1269		1269		1269		1269	
F-value	0.0000		0.0000		0.0000		0.0000	
R-Squared	0.5618		0.6050		0.7213		0.8125	

NOTE) Standard errors are shown in parentheses. The dependent variables are the birth rate indicators derived from "births by prefecture" in the "Vital Survey" (Ministry of Health, Labour and Welfare) and "birthrate indicators derived from "population, demographics and households based on the basic resident register" (Ministry of Internal Affairs and Communications) for "population aged 25-71". The explanatory variables are "population density" and "annual marriage rate by prefecture (per 1,000 population)" in the "Vital Survey" and "gross prefectural production per capita" in the "Prefectural Economic Accounts" (Cabinet Office).

Table. 4: Estimation Results of Auxiliary Estimation Equations (2: Growth accounting)

	Level				Differential			
	OLS		XT		OLS		XT	
Labor	**0.4951**	***	**0.3744**	***	**0.2953**	**	**0.2997**	***
	(0.0057)		**(0.0902)**		**(0.1191)**		**(0.0715)**	
Capital	**0.5321**	***	**0.5396**	***	**0.3616**	***	**0.3577**	***
	(0.0048)		**(0.0150)**		**(0.0269)**		**(0.0313)**	
Observation	1786		1786		1739		1739	
F-value	0.0000		0.0000		0.0000		0.0000	
R-Squared	1.0000		0.9393		0.2209		0.0725	

NOTE) Standard errors are shown in parentheses. "Prefectural GDP" from "Prefectural accounts" (Cabinet Office) is used as the dependent variable. Statistics on workforce and capital are based on the Prefectural Resident Economic Calculations (cabinet) and Tokui (2013). Data from 1975-2012 were used in the calculation of capital intensity, and data from 1975-2014 were used in the derivation of TFP by region.

Table. 5 (a): Descriptive Statistics for Variables in the Second-Stage Estimation (Basic Resident Register)

	Utility	Wage rate	Regional productivity	Working population	Population density	Residential expenses (rent)
Max	**2.913**	**513.369**	**4.8179**	**8847048**	**9528.533**	**9230.000**
Min	**-3.672**	**139.674**	**2.0068**	**347643**	**67.988**	**987.000**
Average	**-0.825**	**303.817**	**2.6805**	**1637821**	**862.619**	**3476.000**
S. Deviation	**1.272**	**71.862**	**0.3622**	**1579977**	**1360.064**	**1346.000**
	Amenity 1	**Amenity 2**	**Social capital**	**Bartik manipulated variables**	**Labour Market Indicators**	**Percentage of residential area**
Max	**3.884**	**4.580**	**11.298**	**0.091**	**5.544**	**0.698**
Min	**-3.583**	**-4.235**	**-2.499**	**-0.134**	**1.442**	**0.163**
Average	**0.000**	**0.000**	**0.000**	**-0.029**	**5.119**	**0.368**
S. Deviation	**1.414**	**1.718**	**2.208**	**0.046**	**1.022**	**0.145**

NOTE) All variables are logarithmically transformed into prefectural data, with differences between 1975 and 2015. "Utility" is the collective indirect utility of each prefecture derived as a fiscal year regional fixed effect by prefecture. "Wage rate" is based on "Average Monthly Cash Salary" (Ministry of Health, Labour and Welfare). "Regional productivity" is derived by applying growth accounting to "gross prefectural product" in "prefectural accounts" (Cabinet Office). The "number of workers" refers to the labor force population of each prefecture (25-71 years old) derived based on the Census. Population density was derived from "Prefectural area" (MLIT). "House rent" refers to "rent per 3m per month and 3. [2]" in the Annual Report of the Retail Price Statistics Survey (Ministry of Internal Affairs and Communications). "Amenity" is a figure obtained by dealing with various indices listed in the "Statistics of Prefectural Sugata" (Ministry of Internal Affairs and Communications) by principal component analysis (Principal Component Analysis) based on the method of Diamond (2016). "Social capital" is the figure obtained by applying the same method as "amenity" in the "Net Capital Stock" in the "Social Capital Stock Estimation Data" (Cabinet Office). "Bartik instruments" refer to instrumental variables derived using the "number of employees and employees by major industry classification" (MHLW) based on the method of Bartik (1999). "Labour market index" is an integrated indicator of the labor market introduced by Morten and Oliveira (2016), "habitable area ratio" is an index calculated by "prefectural area" and "habitable area ratio" of "prefectural area" (Ministry of Land, Infrastructure, Transport and Tourism)," and is used in various spatial economic analyses, including Diamond (2017).

Table. 5 (b): Descriptive Statistics for Variables in the Second-Stage Estimation (Census)

	Utility	Wage rate	Regional productivity	Working population	Population density	Residential expenses (rent)
Max	**0.3717**	**387.927**	**4.0189**	**7302654**	**5470.63**	**5397.0000**
Min	**-1.9303**	**139.674**	**2.0068**	**350268**	**67.99**	**987.0000**
Average	**-0.8409**	**226.3039**	**2.5334**	**1481995**	**588.97**	**2195.4184**
S. Deviation	**0.5681**	**56.9582**	**0.3026**	**1351430**	**1042.88**	**782.2333**
	Amenity 1	**Amenity 2**	**Social capital**	**Bartik manipulated variables**	**Labour market indicators.**	**Percentage of residential area**
Max	**3.1313**	**3.7595**	**1.8386**	**-0.0063**	**5.5441**	**0.6984**
Min	**-3.4316**	**-4.9290**	**-10.9600**	**-0.1343**	**4.2713**	**0.1633**
Average	**0.0000**	**0.0000**	**0.0000**	**-0.0629**	**4.7984**	**0.3716**
S. Deviation	**1.3595**	**1.8097**	**2.0115**	**0.0395**	**0.2182**	**0.1476**

NOTE) All variables are logarithmically transformed into prefectural data, with differences between 1990-2000 and 2000-2010. "Utility" is the collective indirect utility of each prefecture derived as a fiscal year regional fixed effect by prefecture. "Wage rate" is based on "Average Monthly Cash Salary" (Ministry of Health, Labour and Welfare). "Regional productivity" is derived by applying growth accounting to "gross prefectural product" in "prefectural accounts" (Cabinet Office). "Labor population" refers to the labor force population of each prefecture (25-71 years old) derived based on the Census. Population density was derived from "Prefectural area" (MLIT). "Rent" is the "Rent per month/3. 3m2" published in the "Annual Report of the Retail Price Survey" (Ministry of Internal Affairs and Communications). "Amenity" is a figure obtained by dealing with various indices listed in the "Statistics of Prefectural Sugata" (Ministry of Internal Affairs and Communications) by principal component analysis (Principal Component Analysis) based on the method of Diamond (2016). "Social capital" is the figure obtained by applying the same method as "amenity" in the "Net Capital Stock" in the "Social Capital Stock Estimation Data" (Cabinet Office). "Bartik maneuvering variables" refers to the maneuvering variables derived using the "number of employers by industry class" (MHLW) based on Bartik (1999) method. "Labour market index" is an integrated labor market index introduced by Morten and Oliveira (2016), "habitable area ratio" is an index calculated by "prefectural area" and "habitable area ratio" of "prefectural area" (Ministry of Land, Infrastructure, Transport and Tourism)," and is used in various spatial economic analyses, including Diamond (2017).

Table. 6: Estimation Results of Phase II Estimation (Basic Resident Register and Population Census)

	(Structural Estimate A) Basic Resident Register Population Movement Report (1975-2015)				(Structural Estimate B) Census Population Movements (1990-2010)			
	Standard estimation	Amenity 1	Amenity 2	Social capital	Standard estimation	Amenity 1	Amenity 2	Social capital
Wage elasticity	0.9983 ***	0.9983 ***	0.9983 ***	0.9983 ***	1.0005 ***	0.9992 ***	0.9992 ***	0.9992 ***
	(0.0035)	(0.0035)	(0.0035)	(0.0035)	(0.0064)	(0.0065)	(0.0065)	(0.0065)
Residential Demand Reverse Elasticity	0.9482 ***	0.9486 ***	0.9486 ***	0.9486 ***	0.9592 ***	0.5248 ***	0.5291 ***	0.5363 ***
	(0.0268)	(0.0268)	(0.0268)	(0.0269)	(0.0370)	(0.1143)	(0.1135)	(0.1121)
Child preferences	0.8216 ***	0.7921 ***	0.7907 ***	0.8084 ***	0.6086 ***	0.6186 ***	0.6151 ***	0.6009 ***
	(0.1123)	(0.1456)	(0.1484)	(0.1210)	(0.1892)	(0.2019)	(0.2046)	(0.1999)
Dwelling preference	0.3530 ***	0.3684 ***	0.3690 ***	0.3606 ***	0.4270 ***	0.2532 ***	0.2667 ***	0.2903 ***
	(0.0543)	(0.0734)	(0.0751)	(0.0596)	(0.0944)	(0.0961)	(0.1002)	(0.0979)
Regional characteristics		0.1068	0.0760	0.4766		0.3229	1.0584	0.6445
Annual fixed effect	No	No	No	No	Yes	Yes	Yes	Yes
Number of observations	423	423	423	423	141	141	141	141
Over-discrimination test value	0.6977	0.6075	0.6075	0.6075	4.2330	0.0634	0.0625	0.0609
<u>Model parameter restoration value</u>								
α	-0.1746	-0.1605	-0.1597	-0.1690	-0.0356	0.1282	0.1182	0.1088
β	0.8216	0.7921	0.7907	0.8084	0.6086	0.6186	0.6151	0.6009
γ	-	0.1068	0.0760	0.4766	-	0.3229	1.0584	0.6445

NOTE) Standard errors are shown in parentheses. Equations (29)-(31) in Section 4.2 are used in all estimates. The "Basic Resident Register Population Movement Report" and the "Standard Estimation of Indirect Utility Indicators for the Census Population Movement Aggregate" that do not take into account regional characteristics, and the results of descriptions of "Amenity 1", "Amenity 2", and "Social Infrastructure" that take into account various regional characteristics (GMM Estimation Results) are presented. In addition, the parameters in Equation (31) are restored according to the estimated values.

Appendix

1. Derivation of equations

1.1. Derivation of the indirect utility function (15)

In deriving the indirect utility (expected utility) obtained by the agent living in the region j, the agent solves the following utility maximization problem with the utility function (10) constrained by the budget constraint (11).

$$\max_{\{C_{jt},n_{jt},H_{jt}\}} U_{jt} = C_{jt}^{\alpha} n_{jt}^{\beta} H_{jt}^{1-\alpha-\beta} \cdot \exp\left(X_{jt}^{\gamma}\right), \qquad s.t. \quad w_{jt} - h_{jt}H_{jt} = f\left(D_{jt}\right)n_{jt} + C_{jt}.$$

The Lagrangian $\mathcal{L}$ corresponding to this maximization problem can be defined as follows.

$$\mathcal{L} = C_{jt}^{\alpha} n_{jt}^{\beta} H_{jt}^{1-\alpha-\beta} \cdot \exp\left(X_{jt}^{\gamma}\right) + \lambda \cdot \left\{w_{jt} - h_{jt}H_{jt} - f\left(D_{jt}\right)n_{jt} - C_{jt}\right\}.$$

The first-order conditions for the problem are then given by

$$\frac{\partial \mathcal{L}}{\partial C_{jt}} = \alpha C_{jt}^{\alpha-1} n_{jt}^{\beta} H_{jt}^{1-\alpha-\beta} \cdot \exp\left(X_{jt}^{\gamma}\right) - \lambda = 0. \tag{A1}$$

$$\frac{\partial \mathcal{L}}{\partial n_{jt}} = \beta C_{jt}^{\alpha} n_{jt}^{\beta-1} H_{jt}^{1-\alpha-\beta} \cdot \exp\left(X_{jt}^{\gamma}\right) - \lambda f\left(D_{jt}\right) = 0. \tag{A2}$$

$$\frac{\partial \mathcal{L}}{\partial H_{jt}} = (1-\alpha-\beta) C_{jt}^{\alpha} n_{jt}^{\beta} H_{jt}^{-\alpha-\beta} \cdot \exp\left(X_{jt}^{\gamma}\right) - \lambda h_{jt} = 0. \tag{A3}$$

By combining (A1) and (A2), and (A3) and (A1), the number of children n_{jt} and the housing demand H_{jt} in terms of the consumption C_{jt} are derived.

$$\frac{\alpha}{\beta} \cdot \frac{n_{jt}}{C_{jt}} = \frac{1}{f\left(D_{jt}\right)} \quad \Leftrightarrow \quad n_{jt} = \frac{\beta}{\alpha} \frac{1}{f\left(D_{jt}\right)} C_{jt}. \tag{A4}$$

$$\frac{\alpha}{1-\alpha-\beta} \cdot \frac{H_{jt}}{C_{jt}} = \frac{1}{h_{jt}} \quad \Leftrightarrow \quad H_{jt} = \frac{1-\alpha-\beta}{\alpha} \frac{1}{h_{jt}} C_{jt}. \tag{A5}$$

By substituting (A4) and (A5) into the budget constraint (11), the consumption level for C_{jt} for income w_{jt} can be defined.

$$w_{jt} - \frac{1-\alpha-\beta}{\alpha} C_{jt}. - \frac{\beta}{\alpha} C_{jt} - C_{jt} = 0. \quad \Leftrightarrow \quad C_{jt} = \alpha w_{jt}. \tag{A6}$$

Substituting this equation (A6) back to (A4) and (A5), we yield equations (12)-(14). The indirect utility function (15) can be derived by substituting (12)-(14) into the utility function (10).

1.2. Derivation of the migration function (3)

Following the discrete choice model formulation by McFadden (1974)[55], the migration equation of an agent born in a specific region is derived from its regional fitness and indirect utility. In addition to this approach, we discuss the derivation method in the formalized quantitative spatial model of Eaton and Kortum (2002), which can be granted as a technical supplement to this paper.

Discrete choice model by McFadden (1974) Suppose an agent in a group has the characteristic $s \in S$ and chooses x_j from the option $B = \{x_1, \cdots, x_J\}$. Also, the individual utility function can be described in the following format.

$$U = V(s, x) + \varepsilon(s, x). \tag{A7}$$

V describes the non-stochastic, representative preference of the population, and ε represents the probabilistic, idiosyncrasies of individual preference for x. The probability P_i that an individual, who is randomly selected from the population corresponding to the characteristic s and the option B, chooses the option x_i can be defined as follows.

$$\begin{aligned} P_i &= P(x_i | s, B) \\ &= P\left[\varepsilon(s, x_j) - \varepsilon(s, x_i) < V(s, x_i) - V(s, x_j) \text{ for } \forall j \neq i\right]. \end{aligned} \tag{A8}$$

We define further the simultaneous cumulative distribution function $F(\varepsilon_1, \cdots, \varepsilon_J)$ for the values[56] of $\varepsilon(s, x_j)$ for $j = 1, \cdots, J$, and its partial derivative F_i with respect to i (simultaneous probability density function). For $V_i = V(s, x_i)$, the equation (8) can be written as follows.

$$P_i = \int_{\varepsilon=-\infty}^{\infty} F_i\left(\varepsilon + V_i - V_1, \cdots, \varepsilon + V_i - V_J\right) d\varepsilon. \tag{A9}$$

The members of the group have the utility $U(s, x) = V(s, x) + \varepsilon(s, x)$ for the possible choices in option $B = \{x_1, \cdots, x_J\}$. It is assumed that $\varepsilon(s, x)$ corresponding to individual preference's peculiarity

---

[55] The definition of the utility function in McFadden (1974) is different from this paper, but the derived migration equations are the same ones.

[56] In deriving equation (A12) from the definitions of equations (A7) and (A9), McFadden (1974) defines three principles for discrete choice models. They constst of (1) Independence of Irrelevant Alternatives, (2) Posititvity, (3) Irrelevance of Alternative Set Effect. Here, all are satisfied by the formulation of the utility function, the specific term ε and the assumptions about the nature of option B.

is an independent and identical distribution (i.i.d) in the standard Gumbel distribution[57]. That is, $\varepsilon(s,x)$ satisfies the following equation.

$$P\left(\varepsilon\left(s,x_j\right)\leq\varepsilon\right)=\exp\{-\exp(-\varepsilon)\}. \tag{A10}$$

If we set $V_i=V(s,x_i)$ here, we can define the simultaneous probability density function F_i for i.

$$\begin{aligned}F_i\left(\varepsilon+V_i-V_1,\cdots,\varepsilon+V_i-V_J\right)&=\exp(-\varepsilon)\prod_{j=1}^{J}\exp\left(-\exp\left(\varepsilon+V_i-V_j\right)\right)\\&=\exp(-\varepsilon)\exp\left\{-[\exp(-\varepsilon)]\left[\sum_{j=1}^{J}\exp\left(V_j-V_i\right)\right]\right\}. \qquad \text{(A11)}\end{aligned}$$

Replacing this with equation (A9) and rearranging[58] it, the option x_i's selection probability P_i is given as follows.

$$\begin{aligned}P_i&=\int_{\varepsilon=-\infty}^{\infty}\exp(-\varepsilon)\exp\left\{-[\exp(-\varepsilon)]\left[\sum_{j=1}^{J}\exp\left(V_j-V_i\right)\right]\right\}d\varepsilon\\&=-\int_{x=0}^{\infty}\exp\left\{-\exp(-V_i)\left[\sum_{j=1}^{J}\exp\left(V_j\right)\right]x\right\}dx\\&=-\exp(V_i)\left[\sum_{j=1}^{J}\exp\left(V_j\right)\right]^{-1}\cdot\int_{x=0}^{\infty}\left\{-\exp(-V_i)\left[\sum_{j=1}^{J}\exp\left(V_j\right)\right]\right\}\cdot\exp\left\{-\exp(-V_i)\left[\sum_{j=1}^{J}\exp\left(V_j\right)\right]x\right\}dx\\&=-\exp(V_i)\left[\sum_{j=1}^{J}\exp\left(V_j\right)\right]^{-1}\cdot\left|\exp\left\{-\exp(-V_i)\left[\sum_{j=1}^{J}\exp\left(V_j\right)\right]x\right\}\right|_0^{\infty}\\&=\exp(V_i)\left[\sum_{j=1}^{J}\exp\left(V_j\right)\right]^{-1}=\frac{\exp(V_i)}{\sum_{j=1}^{J}\exp\left(V_j\right)}. \qquad \text{(A12)}\end{aligned}$$

The discussion in this paper can be described easily by applying the equation (A12). The group

---

[57] In the original paper, it is called the Weibul distribution, but here it is calles the standard Gumbel distibution according to the convention of the quantitative spatial model. The Type I Extreme Value distribution is the Gumbel distribution, and its density function and cumulative density function can be described as

$$f(\varepsilon)=\frac{1}{\theta}\exp\left\{-\left(\frac{\varepsilon-\mu}{\theta}\right)\right\}\exp\left[-\exp\left\{-\left(\frac{\varepsilon-\mu}{\theta}\right)\right\}\right],F(\varepsilon)=\exp\left[-\exp\left\{-\left(\frac{\varepsilon-\mu}{\theta}\right)\right\}\right].$$

The standard Gumbel distribution is defined as the Gumbel distribution with $\theta=1,\mu=0$.

[58] In the 1st and 2nd columns, the integration by substitution is performed with $x=\exp(-\varepsilon)$, and the first $\exp(-\varepsilon)$ is removed from $d\varepsilon/dx=-1/x$. In the 2nd - 4th columns, the value of P_i is derived based on the integration formula. Since the internal value of $|\cdot|$ is 0-1=-1, the result in the 5th column is derived.

residing in the region i at the start of the time t selects the desired region for residence from the regions $j = 1, \cdots, J$. The indirect utilities of the region at that time are $N_t = \{\nu_{1t}, \cdots, \nu_{Jt}\}$ but since the travel costs $C_{it} = \{c_{i1t}, \cdots, c_{iJt}\}$ must be borne in selecting the region, the options for the individuals in this group are $B_{it} = \{\nu_{1t} - c_{i1t}, \cdots, \nu_{Jt} - c_{iJt}\}$. The options are the expected utilities of the regions, and the individual l additionally observes its preferences $\{\varepsilon_{l1t}, \cdots, \varepsilon_{lJt}\}$. It is assumed that these preferences ε_{ljt} follows the standard Gumbel distribution, and the utility when individual l selects the region j is given by $\nu_{jt} - c_{ijt} + \varepsilon_{ljt}$. The individual utility maximization problem, in this case, is the discrete choice problem of McFadden (1974). According to the equation (A12), the following probability is derived.

$$P_{ijt} = \frac{\exp(\nu_{jt} - c_{ijt})}{\sum_{m=1}^{M} \exp(\nu_{mt} - c_{imt})}.$$

P_{ijt} is precisely the same as π_{ijt} in the equation (13).

Derivation method in quantitative spatial models In the formulation of McFadden (1974), the simultaneous cumulative distribution function $F(\varepsilon_1, \cdots, \varepsilon_J)$ for all options was formulated but in quantitative spaces such as Eaton and Kortum (2002) and Ahlfeldt et al. (2015). The model takes different approaches based on ordered statistics and extreme value theorem. Since this paper is a quantitative spatial model paper, we describe this approach and conclude the supplementary discussion on the derivation process.

The probability P_i that an individual with the characteristic s, who is randomly selected from the group, selects the option x_i from the set B of the options can also be defined by the following equation (A13).

$$P_i = \Pr\left[U_i > \max_{j \in J/\{i\}}(U_j)\right]. \tag{A13}$$

We continue to assume that the error term $\varepsilon(s, x_i) = \varepsilon_i$ the standard Gumbel distribution is an extreme value distribution (the Standard Type I extreme value distribution); it has the maximum stability. Maximum stability means that the maximum value of a random variable sequence that follows an extreme value distribution also follows the same extreme value distribution. Therefore, the following relationship is established for the cumulative distribution function of the standard Gumbel distribution $F(x_i)$ and the cumulative distribution function of its maximum value $F_1(x_i)$.

$$F(x_i) = \Pr[\varepsilon_i \leq x_i] = \exp\{-\exp(-x_i)\}. \tag{A14}$$

$$F_1(x_i) = \Pr\left[\max_{k\in K}(\varepsilon_k) \le x_i\right] = \exp\left\{-\exp\left(-\sum_{k=1}^{K} x_k\right)\right\}. \tag{A15}$$

The same result as the first equation of (A12) can be obtained by performing the equation conversion[59] for the above equation (A13) of the selection probability P_i. Then, the migration probability (3) can be derived by following the same transformation as the equation (A12).

$$\begin{aligned}
P_i &= \Pr\left[V_i + \varepsilon_i > \max_{j\in J/\{i\}}\left(V_j + \varepsilon_j\right)\right] \\
&= \Pr\left[\varepsilon_i + V_i - V_j > \max_{j\in J/\{i\}}\left(\varepsilon_j\right)\right] \\
&= \int_{-\infty}^{\infty} f(\varepsilon_i) F_1\left(\varepsilon_j + V_i - V_j\right) d\varepsilon_i \\
&= \int_{-\infty}^{\infty} \exp(-\varepsilon_i)\exp\{-\exp(-\varepsilon_i)\}\exp\left\{-\exp\left(-\sum_{j=1}^{j\in J/\{i\}} \left(\varepsilon_j + V_i - V_j\right)\right)\right\} d\varepsilon \\
&= \int_{-\infty}^{\infty} \exp(-\varepsilon)\exp\left\{-\exp(-\varepsilon)\left(-\sum_{j=1}^{J}\exp\left(V_i - V_j\right)\right)\right\} d\varepsilon.
\end{aligned} \tag{A16}$$

This method of transforming optimization problems for economic variables (utility maximization problems and cost minimization problems) into a conditional logit model using the extreme value distribution properties has been widely applied. For example[60], with this method, Eaton and Kortum (2002) formulate the selling price under free trade with multiple countries, whereas Ahlfeldt et al. (2015) analyze employment and residence choice among multiple districts in the city.

2. Fast Marching Method (FMM)

This section outlines the Fast Marching Method (FMM), a method for deriving the travel cost index used as the travel time and instrumental variables in estimating interregional population migration,

[59] Lines 1-2 are operations within the inequality sign. In the 2-3 lines, since ε_i is a random variable, it is defined by the probability density function $f(\varepsilon_i)$. Furthermore, in lines 4-5, $V_i - V_i = 0$ holds, and the terms related to the random variable ε_i outside the brankets and ε_j inside the brackets can be summarized. Also, since the random variables ε_i and ε_j are randomly selected from the same independent and identical distribution, the subscripts are eliminated.

[60] The selling price of Eaton and Kortum (2002) is the problem of minimizing the cost of purchasesers in the target country, and the choice of employment/residence of Alhfeldt et al. (2015) is the problem of maximizing the utility of residents. In addition, the derivation of equilibrium conditions without using extreme value theory was limited to three countries at most, and the description was complicated.

and details the derivation process.

2.1. Explanation of Fast Marching Method

Originally, FMM is an analysis method for image analysis for analyzing the shortest distance problem on a solid surface. Allen and Arkolakis (2014) applied this to a two-dimensional transportation network and applied the time required for the shortest path to the derivation. Since then, it has been widely used as the primary analysis method in quantitative spatial analysis.

Let $S \in \mathbb{R}^2$ be a coordinate system on the plane. Suppose we define the time[61] required for the function $\tau: S \to \mathbb{R}_+$ to pass a given point $i \in S$ on the plane. Define the solution $t(i,j)$ of the travel cost minimization problem from point i to point j. Here, $g: [0,1] \longrightarrow S$ is the path and $\Gamma(i,j) \equiv \{g \in C^1 | g(0) = i, g(1) = j\}$ is a set of paths from point i to point j, which is continuous and can be differentiated once.

$$t(i,j) = \inf_{g \in \Gamma(i,j)} \int_0^1 \tau\big(g(t)\big) \left\| \frac{dg(t)}{dt} \right\| dt. \tag{A17}$$

For those paths, a geographic travel cost $T(i,j) = f\big(t(i,j)\big)$ is defined[62] for all regional combinations. This geographic travel cost is known to have two properties. One is symmetry, where $T(i,j) = T(j,i)$ holds for all pairs i,j of points. The other is the continuity[63] of travel costs due to the continuity of the terrain.

The solution to this geographical movement cost $T(i,j)$ can be defined by the following formula. The gradient $\nabla t(i,j)$ is defined for the point j.

$$\|\nabla t(i,j)\| = \tau(j). \tag{A18}$$

Utilizing this property, the transfer cost $T(i,j)$ can be defined by the following process. Here, the same cost curve $\{j | t(i,j) = C\}$ that can be moved from the point i at the same cost (time) C is used as the derivation standard.

1. Define the equal-cost curve $\{j | t(i,j) = C\}$ corresponding the cost C based on the travel cost $\tau(j)$.
2. Expand the equal-cost curve in the form corresponding to the cost $C + \varepsilon$ and define the equal-cost curve $\{j' | t(i,j') = C + \varepsilon\}$. The expansion width is given by $\varepsilon / \tau(j)$ according

[61] It is interpreted as transpottation costs in the new trade theory.
[62] The function $f: \mathbb{R}_+ \longrightarrow [1, \infty)$ must be monotonically increasing ans satisfy $f(0) = 1$.
[63] The formulation is as follows.

$$\lim_{S \to i} T(s,j) = T(i,j)$$

to the expansion destination j.

3. Repeat steps 1 and 2 until the equal-cost curve reaches all points in $S \in \mathbb{R}^2$.

In the next section, the travel cost index $T(i,j)$, which is derived by applying the above process to the travel costs defined for each point.

2.2. Derivation method of travel time index

In this paper, following Allen and Arkolakis (2014), the travel cost index $T_t(i,j)$ was derived every five years from 1975 to 2015. In defining the travel cost $\tau_t(j)$ at the point j at the time t, the time-series geographical data on the transportation means described in "Numerical Land Information" (Ministry of Land, Infrastructure, Transport, and Tourism) was used. As means of transportation, "Shinkansen (bullet train)," "Railway," "Expressway," "National Road," "Regular Sea Route," "Irregular Sea Route," "Land," and "Sea" in order of expected movement speed are considered[64].

For each means of transportation, moving speed is assumed, and then the reciprocal of the relative moving speed is used as the moving cost of the means of transportation. Let the travel cost of the fastest means of transportation at point j be the travel cost $\tau_t(j)$. Here, each pixel on the geographically referenced image of the Japanese map is designed as point j, and pixel j is colored with a shade color corresponding to the fastest means of transportation existing at point j recorded on "National Land Numerical Information" at time t. Note that the coloring is performed ib a grayscale corresponding to the value $c \in [0,255]$, and the darker the color, the faster transportation means exist. The following table (Table A1) shows the code/series name, estimated travel speed, transportation cost τ, and coloring value c on the "National Land Numerical Information" of each means of transportation.

[Table A1]

Of the series of data, "Shinkansen," "Railway," and "Expressway" are one-year time series data, "National Roads," "Regular Sea Routes," "Irregular Sea Routes," "Land," and "Sea" are the data at the time of creation. The range defines the coloring value for the convenience of output by ArcGis and MATLAB. The ArcGIS output images for 1975 and 2015 are shown in Figure 1 and Figure 2.

Next, apply FMM to the nine images corresponding to the transportation network from 1975 to 2015, and derive the travel cost $T(i,j)$ between the prefectural capitals of each prefecture. In deriving, the MATLAB code of Kroon (2009) used by Allen and Arkolakis (2014) was modified as necessary. The structure of the code is described below.

---

[64] Each sea route is required when moving from each prefecture other than Honshu to the prefectures of Honshu. In addition, the cost of movinf land is ser to guarantee the existence of a solution when there is no means of transportation at a specific point. Further, as will be described later, it is impossible to pass on the sea without a route.

Step1. Image capture into MATLAB and identification of primary elements.

- **(a) Capture prefectural office location coordinates (latitude/longitude) and image.**
- **(b) Define the colored range of transportation.**
- **(c) Specify the coordinates of the four corners of the image.**
- **(d) Convert between departure point coordinates and destination coordinates.**

Step.2 Allocation of travel speed and estimation of travel time from all departure points to destinations.

- **(a) Create a combination of departure point and destination (excluding repetition).**
- **(b) Create transportation cost τ in all image coordinates.**
- **(c) Create movement cost index $T(i,j,k)$ from each departure point k to all image coordinates.**
- **(d) Extract travel cost index $T(i,j,k)$ from each departure point k to all destinations.**

By definition, there are only one means of transportation in each image coordinate, so the transportation cost τ is the sum of the defining functions of all means of transportation multiplied by the cost. In this paper, the travel cost index for all combinations derived in this process is used in the first-stage estimation.

3. Regional Characteristics

Here, we will explain the method of deriving the regional characteristics used in the previous studies survey that played an essential role in the first-stage estimation of this paper and the actual analysis and their properties.

3.1. Discussion

While local amenities play an important role in spatial equilibrium, their concepts differ from study to study. Here, we will supplement the discussion on regional characteristics in this paper by organizing previous studies that mention regional characteristics.

According to Rosen (1976) and Roback (1982), which are early studies of the spatial equilibrium model, workers benefit from nominal wages, housing costs, and regional characteristics. The labor market is region-specific, with shocks to labor demand reflected in nominal wages, while shocks to labor supply are reflected in housing costs through population fluctuations. While workers move freely, the land is a fixed element, and regional characteristics are formulated as an evaluation index for workers for this fixed element.

Redding and Sturm (2008) also mention regional characteristics. This study is an extension of

Helpman (1998) to multiple regions, and regional characteristics are defined as an extension of the non-tradable goods (housing) in his economic geography model. In the new economic geography of Krugman (1991), these non-tradable fixed elements (or endowed resources) are thought to act as a factor in the dispersion force against the agglomeration forces. The effects of a diversified economy include intensifying price competition with neighboring companies due to the choice of business location (market congestion effect) and rising living costs due to rising fixed factor prices due to population growth (overcrowding effect). However, unlike the spatial equilibrium model of Roback (1982), in new economic geography, lands that do not have the same characteristics in advance are formulated to produce characteristics as a result of economic activity.

Redding and Rossi-Hansberg (2017) discuss this difference. They set the regional characteristics as "any characteristic that makes a location more or less desirable place of residence" and gives a general definition. In particular, they classify these characteristics into exogenous agglomerations that consider climate change, water resources, and physical geographic factors, and endogenous agglomerations that respond to endogenous shocks to the local economy, such as crime rates.

Based on the above theoretical treatment of regional characteristics, what variables are used as regional characteristics in previous studies is summarized (Table A2).

[Table A2]

In Roback (1982) and early study of the spatial equilibrium model, both exogenous habitability such as weather and endogenous habitability such as crime rate and unemployment rate are used without distinction. This tendency is also seen in Albouy (2013,2016), who has developed the spatial equilibrium model of Roback (1982) and analyzed the quality of life (QOL).

The variable selection of regional characteristics in other studies depends on the study's interests and the formulation of the model. Brueckner and Neumark (2014) and Diamond (2017) formulate natural factors to be reflected in government taxation behavior, considering only exogenous regional characteristics. Since the goal of Ahlfeldt et al. (2015) is to estimate residential externalities and labor externalities using the natural experiment of Berlin's division and variables related to exogenous regional characteristics and transportation, variables related to the post-unification government (endogenous regional characteristics) are adopted. Diamond (2016) also argues that widening regional disparities in worker quality in the United States is due to the accumulation of regional characteristics associated with highly skilled workers, with variables relating exclusively to endogenous regional characteristics being selected in her analysis.

What is noteworthy as a tendency seen in all previous studies is that there is not much detail about how individual regional characteristics affect utility. Within each study, indicators that positively affect utility (green areas, sunny days) and indicators that have a negative effect (noise, crime rate) are often evaluated in the same framework. This tendency may be because regional characteristics are instrumental variables[65] for other variables formulated in the estimation. To address this issue, Diamond (2016) divided 13 regional characteristics into six groups and performed principal component analysis (PCA) in each group for all groups. One virtual regional characteristic is created by performing the principal component analysis again. This procedure allows Diamond (2016) to analyze the nature of the regional characteristics themselves.

This section has developed a survey on the economic definition and concrete indicators regional characteristics (Amenity) in the spatial equilibrium/quantitative spatial model. In particular, the method that uses the principal component analysis of Diamond (2016) step by step is a technique that should be noted when dealing with regional characteristics.

3.2. Derivation process of regional characteristics/ social capital levels

This section describes the process of deriving the regional characteristics “Amenity1,” “Amenity2”, and “Social Capital” used in Section.4 of this paper. As described in the previous chapter, the method using principal component analysis by Diamond (2016) was applied in the derivation.

Candidates for regional variables used for each “Amenity” are extracted and processed according to data development status based on ”Social Demographic System; Statistics of Prefectures and Municipalities.”(Ministry of Internal Affairs and Communications). The variable candidates used for “Social Capital” were “Net Capital Stock by 16 sectors.”(Cabinet Office). The following table (Table A3) summarizes the details of each of these variables.

[Table A3]

Estimates using the “Basic Resident Register Population Migration Report” (Ministry of Internal Affairs and Communications) are from 1975 to 2015, and estimates using the “Census Population Movement Aggregate” (Ministry of Internal Affairs and Communications) are from 1990 to 2015. Therefore, the economic variables used in this derivation, recorded in ”Social Demographic System; Statistics of Prefectures and Municipalities,” that cover the target periods are different among each estimation. As a result, the economic variables used to derive “Amenity” are different. Also, various economic variables are processed into ratios in order to align the units. Social capital uses the aggregated values[66] of 16 departments of “Net Capital Stock” in “Social Capital Stock Estimate data” (Cabinet Office) and is defined by department name.

We created variables "Amenity 1", "Amenity 2", and "Social Capital" by performing the following

[65] Albouy (2013, 2016) explicitly mentions regional characteristics as “instrumental variables”.
[66] Here, we use the one adjusted to the 2010 level using the time-series of the “Regional Consumer Price Index” (Ministry of Internal Affairs and Communications) of the “Consumer Pice index.”

procedure on each of these variable sets using Principal Component Analysis (PCA) in Diamond (2016). As the first stage of the procedure, the variables are divided into groups. Regarding the classification of Diamond (2016) and Table A1, we classified those statistical indicators by prefectures into four divisions, (1) “Environment,” (2) “Education,” (3) “Urban,” (4) “Welfare.” For social capitals, five departments were created ((1) “Transportation,” (2) “Living,” (3) “Education,” (4) “Environment,” (5) “Industry”), and variables are classified into these departments. Furthermore, as the second step, principal component analysis (PCA) was performed for each department, and only the first principal component of each department was extracted. In the third stage, the first principal component of the characteristics was extracted by performing PCA on each department's principal components. This component was designed as the regional characteristic X_{jt}. Regarding “Amenity2”, the same processing was performed after extracting only those having a significant effect on the probability of population migration between prefectures. Table A4 shows the case classification for each sector in each estimation and shows the correspondence between economic variables and sector names and the presence or absence of explanatory power (10% significant) for the probability of population migration. Although it is not used in the estimation of this paper, it is also added that there is an explanatory power for the population migration probability of each sector of “Social Capital Net Stock.”

[Table A4]

Finally, the nature of these regional characteristics is described. In Table A5, the ranking of the time-series averages of “Amenitt1” and “Social Capital” by prefecture is stated. As can be inferred from this, regional characteristics tend to be low in urban areas and high in rural areas, and it can be said that the regional characteristics in this paper show a diversification effect. Regarding the point that a diversification effect can be seen in social capital, the industrial sector's formula, which is expected to contribute to the social capital stock, is limited to a simple one using our productivity. It seems that the background is the accumulation effect could not be dealt with here.

[Table A5]

Charts of Appendix

Table. A1: Geographical reference data of transportation network in "National Land Numerical Information."

Means of Transportation	Code	Series Name	Estimated Movement Speed	Movement Cost	Coloring Value
Shinkansen (Bullet Trains)	N02	Train(Shinkansen)	240km/h	1	0~22
Railway	N02	Railway (Ordinary railway)	120km/h	2	23~50
Expressway	N06	High-speed road time series	80km/h	3	51~72
National Road	N01	Road data	60km/h	4	73~97
Regular Sea Routes	N09	Regular Passenger route data	40km/h	6	98~107
Irregular Sea Routes	C20	Route data	30km/h	8	108~155
Land	-	--	15km/h	16	156~189
Sea	-		Impassible	1000	190~

NOTE) Created by the author using "Numerical Information on Land" (Ministry of Land, Infrastructure, Transport, and Tourism).

Table. A2: Regional characteristics in previous studies

Studies	Target	Category	Regional Characteristics
Roback (1982)	Impact of regional characteristics on wages.	Spatial equilibrium model	Total Crime rate, Particulate Level, Local Unemployment Rate, Population Density, Heating Degree Days, Total Snowfall, Number of Cloudy Days, Number of Clear Days
Albouy (2013)	Impact of regional characteristics on QOL.	Spatial equilibrium model	Heating and Cooling Degree Days, Sunshine, Average Slope, Coastal Proximity, Violent Crimes, Property Crimes, Air Quality Index, Bars and Restaurants, Arts and Culture Index
Brueckner and Neumark (2014) (Diamond, 2017)	Impact of regional characteristics on public sector wages.	Spatial equilibrium model	Warm index: Deviation from the average temperature and 20 degrees Celsius. Dryness index: Negative value of average precipitation. Proximity index: Average distance from nearby coastlines and major rivers. Population density: Population density in the census tract.
Ahlfeldt et al. (2015)	The estimate of accumulation/dispersion effect for each block.	Spatial equilibrium model	Block and Land area and Centroids, Distance to the Nearest U-Bahn and S-Bahn Stations, Green areas; Public Parks, Forests and other Green Public Areas, Water Areas; Canal, River, and Lade, Schools, Noise, Land Use; Commercial, Residential, Industrial and Mixed, Second War Destruction, Listed Buildings, Urban Regeneration Policies Post- Reunification, Government Buildings Post-Reunification, Wages
Diamond (2016)	Factors and impacts of regional disparities in labor skills.	Quantitative Spatial model	Apparel Stores, Eating and Drinking Places and Movie Theaters per 1000 Residents, Busses per Capita, Public Transit Index, Average Annual Daily Traffic-Interstates, Average Annual Daily Traffic-Major Urban Roads, Poverty and Violent Crimes per 1000 Residents, EPA Air Quality Index, Government Spending on Parks and Recreation per Capita, Government Spending on K-12 Education, Pupils Enrolled in Public K-12 Education, Number of Enrolled in Public K-12 Education, Patents Per Capita, Employment Rate
Albouy (2016)	Impact of regional characteristics on quality of life, wages, and corporate productivity.	Quantitative Spatial model	Heating and Cooling Degree Days, Sunshine, Average Slope, Coastal Proximity, Latitude, Wheaton Residential Land Use Regulatory Index; from Gyokuro, Saiz, and Summers (2008)

NOTE) This table was made from the theoretical model of this paper.

Table. A3: Source and Processing Procedure of Regional Characteristics

Name of Variables	Processing Process	Statistical Series Name	Source
Amenities (Basic Resident Register Population)			
Average Temperature	-	"Average Temperature"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Highest Temperature	-	"Highest Temperature"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Lowest Temperature	-	"Lowest Temperature"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Percentage of Daylight Hours	"Daylight Hours(yearly)"/"Hours(yearly)"	"Daylight Hours(Yearly)"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Block Park Area Ratio	"Block park area" / "total area"	"Block park area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Neighborhood Park Area Ratio	"Neighborhood park area" / "total area."	"Neighborhood park area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Sports Park Area Ratio	"Athletic park area" / "total area"	"Athletic park area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Green Space Area Ratio	"Green area" / "Total area"	"Green area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Green Road Area Ratio	"Green road area" / "total area"	"Green road area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Highschool Graduation Rate	"Final education population (high school / old middle school)" / "total population"	"Final Education Population (High School / Old Middle School)" "Total population."	"Census Report (Large-scale Survey)" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
Vocational School Graduation Rate	"Final education population (junior college / technical college)" / "total population"	"Final Education Population (Junior College / Technical College)" "Total population."	"Census Report (Large-scale Survey)" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
University/Graduate School Graduation Rate	"Final education population (university / graduate school)" / "total population"	"Final Education Population (University / Graduate School)" "Total population."	"Census Report (Large-scale Survey)" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
Labor Force Population Ratio	"Labor population" / "total population"	"Labor population." "Total population."	"Labor Force Survey" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
Unemployment Rate	---	"Unemployment rate."	"Labor Force Survey" (Ministry of Internal Affairs and Communications)
Number of crimes per 1000 people	1000 x "Number of criminal offenses recognized" / "Total population"	"Number of criminal offenses recognized." "Total population."	"White Paper on Police" (National Police Agency) "Census Report" (Ministry of Internal Affairs and Communications)
Number of pregnant women instructors per 1000	1000 x "Pregnant women's health guidance staff."	"Pregnant women's health guidance staff." "Total population."	"Community Health Business Report (Health Center Management Report)" "Census Report" (Ministry of Internal Affairs and Communications)
Number of infant health instructors per 1000	/"Total population"	"Infant health guidance instructed personnel." "Total population."	"Community Health Business Report (Health Center Management Report)" "Census Report" (Ministry of Internal Affairs and Communications)
Number of hospitals per 1000 people	1000 x "Infant Health Guidance Instructed Personnel."	"Number of hospitals." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Number of clinics per 1000	/"Total population"	"Number of general clinics." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Number of hospital beds per 1000 people	1000 x "number of hospitals" / "total population"	"Number of hospital beds." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Number of clinic beds per 1000 people	1000 x "Number of general clinics" / "Total population"	"Number of beds in general clinics." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Amenities (Census)			
Lake Area Ratio	"Main lake area" / "Total area."	"Main lake area." "Total area."	"Area adjustment by prefecture, city, ward, town, and village." (Geographical Survey Institute) "Area adjustment by prefecture, city, ward, town, and village." (Geographical Survey Institute)

Forest Area Ratio	"Forest area" / "Total area"	"Forest area." "Total area."	"Area adjustment by prefecture, city, ward, town, and village." (Geographical Survey Institute) "Area adjustment by prefecture, city, ward, town, and village." (Geographical Survey Institute)
Average Temperature	-	"Average Temperature"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Highest Temperature	-	"Highest Temperature"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Lowest Temperature	-	"Lowest Temperature"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Average Temperature	-	"Average Temperature"	"Meteorological Observatory Observation Data" (Japan Meteorological Agency)
Number of Retail Stores per 1000 People	1000 x "Number of retail stores" / "Total population"	"Number of retail stores." "Total population."	"Economic Census" Business / Corporate Statistics Survey " (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
Sales Floor Area per Retail Store	"Retail sales floor area" / "Number of retail stores."	"Retail sales floor area." "Number of retail stores."	"Commercial Statistics" (Ministry of Economy, Trade, and Industry) "Economic Census" Business / Corporate Statistics Survey " (Ministry of Internal Affairs and Communications)
Road Density	"Actual road extension" / "total area"	"Actual road extension." "Total area."	"Annual Road Statistics" (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Block Park Area Ratio	"Block park area" / "total area"	"Block park area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Neighborhood Park Area Ratio	"Neighborhood park area" / "total area."	"Neighborhood park area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Sports Park Area Ratio	"Athletic park area" / "total area"	"Athletic park area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Green Space Area Ratio	"Green area" / "Total area"	"Green area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Green Road Area Ratio	"Green road area" / "total area"	"Green road area." "Total area."	"City Planning Status Survey" City Planning Annual Report " (Ministry of Land, Infrastructure, Transport, and Tourism) "Area adjustment by prefecture, city, ward, town, and village" (Geographical Survey Institute)
Highschool Graduation Rate	"Final education population (high school / old middle school)" / "total population"	"Final Education Population (High School / Old Middle School)" "Total population."	"Census Report (Large-scale Survey)" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
Vocational School Graduation Rate	"Final education population (junior college / technical college)" / "total population"	"Final Education Population (Junior College / Technical College)" "Total population."	"Census Report (Large-scale Survey)" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
University/Graduate School Graduation Rate	"Final education population (university / graduate school)" / "total population"	"Final Education Population (University / Graduate School)" "Total population."	"Census Report (Large-scale Survey)" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
Junior High School Enrollment Rate	-	"Junior high school enrollment rate"	"School Basic Survey" (Ministry of Education, Culture, Sports, Science, and Technology)
High School Enrollment Rate	-	"High school student enrollment rate"	"School Basic Survey" (Ministry of Education, Culture, Sports, Science, and Technology)
Learning Time Ratio	"Average study time" / 24	"Average time of study"	"Social Life Basic Survey" (Ministry of Internal Affairs and Communications)
Commuting / School Time Ratio	"Average time for commuting to work/school" / 24	"Average time for commuting to work / school"	"Social Life Basic Survey" (Ministry of Internal Affairs and Communications)
Labor Force Population Ratio	"Labor population" / "total population"	"Labor population." "Total population."	"Labor Force Survey" (Ministry of Internal Affairs and Communications) "Census Report" (Ministry of Internal Affairs and Communications)
Unemployment Rate	---	"Unemployment rate."	"Labor Force Survey" (Ministry of Internal Affairs and Communications)
Number of crimes per 1000 people	1000 x "Number of criminal offenses recognized" / "Total population"	"Number of criminal offenses recognized." "Total population."	"White Paper on Police" (National Police Agency) "Census Report" (Ministry of Internal Affairs and Communications)
Birth to Population Ratio	"Number of births" / "total population."	"Number of births." "Total population."	"Vital Statistics" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Number of pregnant women instructors per 1000	1000 x "Pregnant women's health guidance staff."	"Pregnant women's health guidance staff." "Total population."	"Community Health Business Report (Health Center Management Report)" "Census Report" (Ministry of Internal Affairs and Communications)
Number of infant health instructors per 1000	/"Total population"	"Infant health guidance instructed personnel." "Total population."	"Community Health Business Report (Health Center Management Report)" "Census Report" (Ministry of Internal Affairs and Communications)
Number of hospitals per 1000 people	1000 x "Infant Health Guidance Instructed Personnel."	"Number of hospitals." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare)

			"Census Report" (Ministry of Internal Affairs and Communications)
Number of clinics per 1000	/"Total population"	"Number of general clinics." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Number of hospital beds per 1000 people	1000 x "number of hospitals" / "total population"	"Number of hospital beds." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Number of clinic beds per 1000 people	1000 x "Number of general clinics" / "Total population"	"Number of beds in general clinics." "Total population."	"Medical Facility Dynamics Survey" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Number of Doctors per 1000	1000 x "number of doctors" / "total population"	"Number of doctors." "Total population."	"Physician / Dentist / Pharmacist Statistics" (Ministry of Health, Labor and Welfare) "Census Report" (Ministry of Internal Affairs and Communications)
Medical Expenses (Logarithmic)	Logarithmic conversion of "medical expenses (national medical expenses)"	“Medical expenses (national medical expenses)”	"National Medical Expenses" (Ministry of Health, Labor and Welfare)
Social Capital (”Net Social Capital Stock by 16 sectors”)			
1. Road	5. Sewerage	9. Schooling Facilities	13. Agriculture/Forestry/Fisheries
1. Port	6. Waste	10. Social Education	14. National Forest
2. Aviation	7. Water Supply	11. Hydraulic Engineering/Restoration	15. Industrial Water
3. Pulbic Housing	8. City Park	12. Coast	16. Government Buildings

NOTE) This table was made using ” Social Demographic System; Statistics of Prefectures and Municipalities.” (Ministry of Internal Affairs and Communications) and “Net Capital Stock by 16 sectors”(Cabinet Office).

Table. A4: Classification of regional characteristics

Name of Variables	Divisions		Significance(10%)
Amenities (Basic Resident Register Population)			
Average Temperature			○
Highest Temperature			○
Lowest Temperature			
Percentage of Daylight Hours			
Block Park Area Ratio	**(1)**	**Environment**	
Neighborhood Park Area Ratio			
Sports Park Area Ratio			
Green Space Area Ratio			
Green Road Area Ratio			
Highschool Graduation Rate			
Vocational School Graduation Rate	**(2)**	**Education**	○
University/Graduate School Graduation Rate			○
Labor Force Population Ratio			○
Unemployment Rate	**(3)**	**Urban**	○
Number of crimes per 1000 people			○
Number of pregnant women instructors per 1000			
Number of infant health instructors per 1000			○
Number of hospitals per 1000 people	**(4)**	**Welfare**	○
Number of clinics per 1000			○
Number of hospital beds per 1000 people			
Number of clinic beds per 1000 people			
Amenities (Census)			
Lake Area Ratio			○
Forest Area Ratio			
Average Temperature			○
Highest Temperature	**(1)**	**Environment**	○
Lowest Temperature			
Average Temperature			
Number of Retail Stores per 1000 People			○
Sales Floor Area per Retail Store	**(3)**	**Urban**	
Road Density			
Block Park Area Ratio			
Neighborhood Park Area Ratio			
Sports Park Area Ratio	**(1)**	**Environment**	
Green Space Area Ratio			○
Green Road Area Ratio			
Highschool Graduation Rate			
Vocational School Graduation Rate			
University/Graduate School Graduation Rate	**(2)**	**Education**	
Junior High School Enrollment Rate			
High School Enrollment Rate			○

Learning Time Ratio			
Commuting / School Time Ratio			
Labor Force Population Ratio	**(3)**	**Urban**	○
Unemployment Rate			○
Number of crimes per 1000 people			
Birth to Population Ratio			○
Number of pregnant women instructors per 1000			○
Number of infant health instructors per 1000			○
Number of hospitals per 1000 people	**(4)**	**Welfare**	
Number of clinics per 1000			
Number of hospital beds per 1000 people			
Number of clinic beds per 1000 people			○
Number of Doctors per 1000			
Medical Expenses (Logarithmic)			○
Social Capital			
Road			
Port	**(1)**	**Transportation**	
Aviation			
Public Housing			○
Sewerage			
Waste	**(2)**	**Living**	
Water Supply			○
City Park			○
Schooling Facilities	**(3)**	**Education**	○
Social Education			○
Hydraulic Engineering/Restoration			
Coast	**(4)**	**Environment**	○
Agriculture/Foresty/Fishcers			○
National Forest			
Industrial Water	**(5)**	**Industry**	○
Government Buildings	**(6)**	**Living**	

NOTE) Created by the author based on Diamond (2016).

Table. A5: Local Characteristics

Amenities		**Social Capitals**	
Top 5 Prefectures	**Bottom 5 Prefectures**	**Top 5 Prefectures**	**Bottom 5 Prefectures**
1. Nagano	43. Fukuoka	1. Tottori	43. Kanagawa
1. Iwate	44. Nara	2. Yamanashi	44. Hokkaido
2. Yamagata	45. Osaka	3. Shimane	45. Aichi
3. Shizuoka	46. Kyoto	4. Nara	46. Osaka
4. Aomori	47. Tokyo	5. Miyazaki	47. Tokyo

NOTE) The author created this table.