Historical Trends of Agglomeration to the Capital Region and New Economic Geography*

Takatoshi Tabuchi†

30 November 2013

Abstract

This paper shows that a family of the Dixit-Stiglitz type of new economic geography models is capable of simulating the real-world tendency for agglomeration to the primate city. It is often observed that while regional populations were dispersed in early times, they have been increasingly concentrated into one capital region over recent years. The present paper thus characterizes the stable equilibrium distribution for any number of regions, any set of interregional distances, and any distribution of immobile demand for sufficiently small or large transport costs. It also demonstrates that multi-region new economic geography models are able to simulate the real-world population distribution trends witnessed over the past few centuries.

Keywords: agglomeration; new economic geography; historical population distribution

JEL Classification: R1, R3

---

*I wish to thank Steven Brakman, Masahisa Fujita, Tomoya Mori, Michael Pflüger, Daniel Sturm, Jens Suedekum, and two anonymous referees. I also thank H. Kobayashi and W. Wanichthaworn for excellent research assistance. We acknowledge the financial support by the JSPS Grants-in-Aid for Scientific Research B and RIETI.

†Faculty of Economics, University of Tokyo and Research Institute of Economy, Trade and Industry. Address: Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: ttabuchi@e.u-tokyo.ac.jp
1 Introduction

Since the pioneering work of Krugman (1991), new economic geography (NEG) has been developed and sophisticated in several directions in order to show how the spatial distribution of economic activities is evolving in the real world. Specifically, the alternative modeling strategies proposed by Ottaviano, Tabuchi and Thisse (2002), Forslid and Ottaviano (2003), and Pützer (2004), among others, have improved analytical tractability, which has enabled researchers to gain further insights into the space economy and its transition. Furthermore, NEG has been enriched by introducing important ingredients such as intermediate goods, commuting costs, land for housing, agricultural transport costs, firm heterogeneity, and economic growth.

The scopes of most of the theoretical studies published thus far have been limited to two regions in order for researchers to reach meaningful analytical results. The two-region NEG models tend to demonstrate that spatial distribution is dispersed in the early period (high trade costs or low manufacturing share) and agglomerated in one of the two regions in the late period (low trade costs or high manufacturing share).

However, it is no doubt that the two-region NEG models are too simple to describe the spatial distribution of economic activities in real-world economies. Since there are only two regions, their geographical locations are necessarily symmetric, and thus diverse spatial distributions cannot occur. Moreover, it is unlikely that, say, Eastern regions have been growing at the expense of Western regions in a country. Many regions interact both in trade and in migration in the real world, where geographical locations of regions are asymmetric suggesting that their respective transport costs are different. In order to describe such geography, it is indispensable to assume many regions and unequal transport costs between them.

Some scholars have already attempted to extend two-region to multi-region NEG models. Multi-region NEG models are somewhat analytically tractable in a racetrack economy, where regions are symmetrically located on the circumference of a circle (Krugman, 1993; Picard and Tabuchi, 2010), as well as in a linear economy, where regions are located on a line (Venables and Limão, 2002; Ago, Isono and Tabuchi, 2006). In order to depict
the long-run evolutionary process of multi-regional development, we start from a position of autarky with the dispersed distribution of economic activities. The decrease in transport costs would enable firms to trade between regions, which alters equilibrium prices, wages and profits and fosters the migration of firms and thereby workers. This would enable some regions to gain manufacturing share and may allow one region to attract all the manufacturing activities. This scenario is analytically confirmed in section 3 of this paper.

Although there have been attempts to simulate the spatial distribution of economic activities in the real world (notably Bosker et al., 2010), previous studies have not yet succeeded in obtaining practical analytical results for the long-term transition of the spatial distribution of economic activities in a multi-region economy. The present paper thus bridges this gap in the current body of knowledge on this topic by describing the transition of the spatial distribution of economic activities over the past few centuries. To do so, consider NEG models using an arbitrary number of regions in order to maintain analytical tractability.

The spatial distribution of economic activities has certain characteristics. The first is Zipf’s law or the rank-size rule of city size distribution. The microfoundations of this law have been explained by Rossi-Hansberg and Wright (2007), among others. However, this is beyond the scope of the present paper, which focuses on explaining long-term changes in the size distribution of regions rather than of cities. In this paper, each region consists of (manufacturing) cities and (agricultural) rural areas, which is to be analyzed by NEG models.

The second characteristic is the robustness of long-run regional population distribution trends to large temporary shocks, such as the bombing of Japan during World War II (Davis and Weinstein, 2002). However, although regional population distribution is robust and stable during an intermediate period of time, it has been shown to be gradually changing over very long periods of time, as described in the next section. In particular, it is often observed that the capital region has experienced distinct growth patterns over centuries especially after the Industrial Revolution.

Therefore, this paper simply plots the long-term regional population distributions for
several countries in order to assess how the aftermath of the Industrial Revolution and the recent IT revolution have influenced decreases in shipping and communications costs and thereby affected urbanization and agglomeration to core regions (Bairoch, 1988). It then presents NEG models with asymmetric transport costs and immobile demand that are capable of explaining the described changes in the spatial distributions of economic activities.

The remainder of this paper is organized as follows. The regional population distribution trends for the postwar period and for a few centuries in several countries are investigated in the next section. In order to explain these trends, a multi-region extension of Krugman’s (1991) model is presented and the existence and stability of the spatial equilibrium are examined and interpreted in relation to an actual multi-region economy for symmetric and asymmetric regions in section 3. Section 4 concludes the paper.

2 Long-term trends in population distribution

2.1 Postwar trends

I first considered the rates of population growth and decline of the largest city in each sample country. Since the largest cities in these countries often spread beyond municipal boundaries, I chose metropolitan areas rather than municipal city areas as the unit of analysis. Although the definitions of metropolitan areas differ by country, the United Nations database of urban agglomerations includes both central cities and suburbs, and provides a universal definition of metropolitan areas.\footnote{According to World Urbanization Prospects (http://esa.un.org/unup/index.asp?panel=6), the term “urban agglomeration” refers to the de facto population contained within the contours of a contiguous territory inhabited at urban density levels without regard to administrative boundaries. It usually incorporates the population in a city or town plus that in the suburban areas lying outside of but adjacent to the city boundaries.} The data sources are listed in Appendix 1.

From the UN database, I chose the top 30 countries according to GDP in 2010 and then selected the largest metropolitan area (= agglomeration) in each country. I collected these
data on every fifth year between 1950 and 2010, resulting in a dataset of 13 observations per country. Even though the national populations in each sample country increased during the study period, the population shares of the largest metropolitan areas also increased owing to interregional migration. In order to confirm this trend, I calculated the correlation coefficients between the population share of the largest metropolitan area and sample years of 1950, 1955, . . . , 2010. It was found that these correlation coefficients were significantly positive in 24 countries, significantly negative in 4 countries, and insignificant in 2 countries out of the 30 countries at the 5 percent level. This finding implied that the population shares in most of the largest cities in the top 30 countries by GDP have grown since World War II.

Figure 1a displays the 25 countries in which the population share of the largest metropolitan area is increasing over time, while Figure 1b shows the five countries in which these population shares are decreasing over time (one of which is statistically insignificant). The marked growth in Tokyo, Ar-Riyadh, and Seoul, which have increased by approximately 15 percentage points over the 60-year study period, is notable.

Most of the largest agglomerations experienced gradual postwar growth with a few exceptions such as New York. The regional evolution in the United States may be explained by immigration from Europe and settlement to the West for more than four centuries. Such an exogenous increase in population was modeled under the NEG framework by Fujita, Krugman and Mori (1999). Rather than focusing on these few exceptions, however, the present paper pays attention to the gradual increases in the largest agglomerations that started agrarian societies and then became industrialized societies.

### 2.2 Historical trends

For assessing the longer-term changes in the spatial distribution of economic activities, it is necessary to consider a spatial unit that is fixed over time. For this reason, metropolitan areas such as Metropolitan Statistical Areas are unsuitable.² It is also necessary for a spatial unit to be large enough to include metropolitan areas that contain a central city and suburbs. In fact, according to the population density plots of the Tokyo and New

---

²For information on how varying definitions yield different population totals, see Forstall et al. (2009).
York metropolitan areas plotted by Nakamura (http://blogs.yahoo.co.jp/shohei_tokyo_1980/32233480.html), these areas are continuously and smoothly decreasing from each city center for a 100-km radius, implying that they may spread across an area that has a radius of more than 100 km. This finding implies that, for example, the London region or Greater London is too small to cover the London metropolitan area and Tokyo prefecture is too small to cover Tokyo metropolitan area.

Based on these observations, I collected historical data on the regional population and aggregated some of them for the following eight sample countries: Brazil (sample period 1872–2010), France (1851–2009), Great Britain (1701–2010), Italy (1881–2001), Japan (1721–2010), Spain (1900–2010), Canada (1834–2012), and the United States of America (1900–2008). As before, the data sources are listed in Appendix 1. The historical population shares in each region are plotted in Figures 2a–2h, respectively.

A visual inspection of Figures 2a–2g reveals the striking feature that the region containing the largest metropolitan area (i.e., Sao Paulo in Brazil, Île-de France in France, London+East+South East in Great Britain, Lazio in Italy, Minami Kanto in Japan, and Ontario in Canada) has experienced significant population growth in recent years in comparison with that in the rest of the country. In particular, the region that contains the largest metropolitan area in Brazil (Figure 2a), France (Figure 2b), and Japan (Figure 2e) showed remarkable growth relative to the rest of the country. Indeed, the share of the studied region tripled in France and was 2.5 times larger in Brazil and Japan during their respective study periods. This growth pattern was similar in Spain (Figure 2f). However, Spain also presents rapid population growth in its two central regions, Madrid and Barcelona, possibly because of their different cultures (although Barcelona has lost share since the 1980s). Therefore, the same basic forces seem to be at work in these countries. I may thus conclude that the region containing the largest metropolitan area has significantly increased in population share at the expense of the rest of the country in the long-term in these seven countries.

---

3 These historical regional data are those I was able to source. For the other countries, the sample period of the data either is too short or the number of regions is too few.

4 This is true for Ontario in Canada after World War II (Figure 2g), but not necessarily so before World War II. The latter trend is similar to that in the United States for the reason explained below.
The observations presented in Figures 2a–2g are consistent with those in Figures 1a–1b with the exception of Great Britain. The share of the capital region in Great Britain increased after 1800 (Figure 2b), whereas the largest metropolitan area of London has lost share since World War II (Figure 1b). Such a discrepancy occurs because of the definition of the regional boundary: the latter region is not large enough to cover the whole area in which daily commuting to the center of the capital is possible.

Finally, it is fair to say that not all countries have experienced the same regional growth patterns. The notable exception is the United States. Its largest metropolitan area (New York) and capital (Washington, D.C.) belong to the Mideast. According to Figure 2h, the share of the Mideast has been declining since early in the 19th century. This decrease is in sharp contrast to other countries and partly similar to Ontario in Canada in the first half of the 20th century. The declining share in the Mideast and Ontario may be attributed to the settlement to the West mentioned previously, while there are millennia of settlement history in the other countries.

The foregoing raises the question of why the population shares in the regions that contain the largest metropolitan areas have significantly increased. The level of urban concentration to capital regions may be partly determined by political factors (Ades and Glaeser, 1995), while the growth in urban concentration to these regions may be determined by the extent-of-the-market and aggregate demand (Ades and Glaeser, 1999). I attempt to explain and simulate these historical trends using NEG models in the next section.

\[ \chi^2 = 188.6, \text{ degrees of freedom} = 1, \text{ and a probability} = 6.54 \times 10^{-43}. \]
Obviously, there are too few NEG papers in the United States relative to economics papers, which may be affected by differences in historical agglomeration trends.
3 NEG models to explain historical trends

I consider the multi-region NEG model of Fujita, Krugman and Venables (1999) with an arbitrary number of regions $n(\geq 2)$ and extend it to an arbitrary distribution of immobile farmers. Following the established tradition of NEG, each region is divided into traditional and modern sectors. Agricultural production is carried out by farmers (or unskilled workers) in the traditional sector, whereas manufacturing production is carried out by workers (or skilled workers) in the modern sector. The size of the workforce is fixed and split according to given shares between farmers and workers. Each person supplies one unit of labor inelastically. The given masses of farmers and workers are equal to $1 - \mu > 0$ and $\mu > 0$, respectively. Farmers are immobile, while workers are assumed to be spatially mobile. The number of farmers in region $r(= 1, 2, \ldots, n)$ is exogenously given by $(1 - \mu) a_r$, where $\sum_{r=1}^{n} a_r = 1$.

The agricultural good is homogeneous and is supplied under constant returns and perfect competition. In order to produce one unit of the homogenous good, one unit of farming labor is required. This good can be traded freely between regions and thus its price is identical across regions. Hence, this good may be chosen as the numéraire. As a result, farmers’ wages are equal to one in each region.

The manufacturing good is horizontally differentiated and is supplied under increasing returns and monopolistic competition. Manufacturing technology is identical for all varieties and in all regions and involves a fixed input requirement $F$ and marginal input requirement $c$. Thus, the production of a quantity $q$ requires labor input $l$, given by

$$l = cq + F.$$ 

There is a continuum of potential firms, implying that the impact of each firm on market outcomes is negligible. Because of increasing returns to scale in production, each variety is produced by a single firm. Since firms are symmetric within a region, each firm’s output is equalized within a region in equilibrium. Therefore, the total number of firms and varieties in region $r$ is given by $m_r = \mu \lambda_r / l$, where the percentage distribution of workers is denoted by

$$\Lambda \equiv (\lambda_1, \lambda_2, \ldots, \lambda_n), \quad \lambda_r \in [0, 1], \quad \sum_{r=1}^{n} \lambda_r = 1.$$
Interregional shipments of any variety are subject to iceberg trade costs, where a constant fraction $T$ of each variety melts away per unit of distance. In other words, $T_{rs} \equiv \exp(\tau d_{rs})$ units have to be shipped for one unit to reach its destination, where $d_{rs}$ is the distance between regions $r$ and $s$.

The preferences of a typical resident of region $s$ are represented by

$$U_s = \mu^{-\mu} (1 - \mu)^{-\mu - 1} \left[ \sum_{r=1}^{n} \int_{0}^{m_r} q_{rs}(v)^{\frac{\mu-1}{\mu}} dv \right]^{\frac{\mu}{\mu - 1}} A^{1 - \mu}, \quad (1)$$

where $q_{rs}(v)$ is the consumption of variety $v$ in region $s$ that is produced in region $r$, $\sigma > 1$ measures the elasticity of substitution between any two varieties, and $A$ is the consumption of the homogeneous good. The budget constraint of a consumer earning a wage $w_s$ in region $s$ is as follows:

$$\sum_r \int_{0}^{m_r} p_{rs}(v) q_{rs}(v) dv + A = w_s, \quad (2)$$

where $p_{rs}(v)$ is the delivered price of variety $v$ in region $s$ that is produced in region $r$. Given these assumptions, firms differ only by their locations in equilibrium. Accordingly, I drop the variety label $v$ hereafter.

The maximization of (1) subject to the budget constraint (2) yields the following demand for a variety produced in region $r$:

$$q_{rs} = \frac{P_{rs}^{-\sigma} P_s^{1 - \sigma}}{P_{s}^{1 - \sigma}} \mu Y_s, \quad (3)$$

where $Y_s = \mu \lambda_s w_s + (1 - \mu) a_s$ is the total income in region $s$ and $P_s = \left( \sum_t m_t P_{ts}^{1 - \sigma} \right)^{1/(1 - \sigma)}$ is the price index in region $s$. Accordingly, the indirect utility of a consumer residing in region $s$ is

$$V_s = \frac{w_s}{P_s^\mu}. \quad$$

Because of the iceberg assumption, a typical firm established in country $r$ has to produce $x_{rs} = T_{rs} q_{rs}$ units to satisfy final demand $q_{rs}$ in country $s$. The firm in region $r$ takes (3) into account when maximizing its profit given by

$$\pi_r = \sum_s (p_{rs} q_{rs} - w_r c x_{rs}) - w_r F = \sum_s (p_{rs} - w_r c T_{rs}) \frac{P_{rs}^{-\sigma}}{P_{s}^{1 - \sigma}} \mu w_s - w_r F. \quad (4)$$
The maximization of (4) yields the equilibrium price:

\[ p_{rs} = \frac{\sigma}{\sigma - 1} cT_{rs}w_r = T_{rs}w_r. \tag{5} \]

where I normalize the marginal labor requirement \( c = 1 - 1/\sigma \). Assuming the free entry of firms, \( \pi^* = 0 \) holds, which leads to the equilibrium output:

\[ \sum_s x_{rs} \leq F\sigma = l. \tag{6} \]

Let \( \phi_{rs} \equiv T_{rs}^{1-\sigma} = \exp[(1-\sigma)\tau d_{rs}] \) be the freeness of trade. By substituting the demand (3) and price index into (6), multiplying both sides by \( p_{rr} > 0 \), and using (5), I derive

\[ \sum_s w_r^{-\sigma} \phi_{rs} [\mu \lambda_s w_s + (1 - \mu) \lambda_s] \leq 1 \tag{7} \]

with equality if \( \lambda_r > 0 \). Hence, \( n \) wage equations (7) determine the equilibrium wages \( w_r \).

According to the Walras law, one of these conditions is redundant.

Workers must reach the same utility level in the long-run equilibrium. A spatial equilibrium is such that there exists a constant \( \bar{V} \) for which

\[ V_r \leq \bar{V} \quad \text{and} \quad (V_r - \bar{V}) \lambda_r = 0 \quad \forall r = 1, \ldots, n. \tag{8} \]

Because \( V_r \) is continuous with respect to \( \lambda_s \) (\( s = 1, \ldots, n \)), it follows from Ginsburgh et al. (1985) that a spatial equilibrium exists for all parameter values.

For the stability of equilibrium, I choose the replicator dynamics of \( n - 1 \) equations and \( n - 1 \) variables:

\[ \dot{\lambda}_r = \lambda_r \left( V_r - \sum_s \lambda_s V_s \right) \equiv J_r \quad \text{for } r = 1, \ldots, n - 1, \]

where \( \dot{\bullet} \) denotes the time derivative of \( \lambda_r \) and the equation for the \( n \)th region is out of consideration because of the identity \( \lambda_n = 1 - \sum_{s=1}^{n-1} \lambda_s \). Since unstable equilibria are rarely observed in the real world, stability is used to refine the equilibria.

Because the model setting is quite general in that both interregional transport costs and the immobile farmer’s distribution are asymmetric, it is difficult to characterize stable equilibria analytically. In order to obtain meaningful results, I thus consider two extreme cases of near autarky \( \tau \to \infty \) and near free trade \( \tau \to 0 \) in the following section.
3.1 Asymmetric regions

First, consider the case of near autarky. When transport costs are sufficiently large \( \tau \to \infty \), it can be easily shown from (7) that \( w_r = a_r / \lambda_r \) and

\[
V_r = a_r^{1-\mu} \lambda_r^{\frac{\mu}{\sigma-1}-1}
\]  
(9)

hold for all \( r \). Assume the no-black-hole condition

\[
\mu < \mu_{\text{max}} = \frac{\sigma-1}{\sigma}
\]  
(10)

so that increasing returns at the level of the economy as a whole are not so strong that manufacturing agglomerates regardless of transport costs. Then, equating \( V_r \) for all \( r \) yields a spatial equilibrium given by

\[
\lambda_r^* = \frac{(1-\mu)(\sigma-1)}{\sum_s a_s \sigma^{s-\mu(\sigma-1)}} > 0.
\]  
(11)

Because the exponent of the stable equilibrium (11) is positive and greater than one under the no-black-hole condition (10), mobile manufacturing workers are more strongly concentrated (distributed more unevenly) than immobile farmers when trade costs are very high. This is a manifestation of the home market effect observed in the trade literature (Krugman 1980). Hence, I ascertain the following proposition for near autarky.

**Proposition 1** When transport costs are sufficiently large, there exists a unique stable equilibrium (11), where manufacturing in each region is nonempty. Due to the home market effect, manufacturing workers are more strongly concentrated than farmers.

This exponent in (11) is decreasing in \( \mu \) and increasing in \( \sigma \), implying that the degree of concentration is the stronger, the larger is the aggregate manufacturing sector and the stronger is product differentiation. This result is consistent with the standard results under NEG.

The spatial equilibrium (11) is unique and stable because \( \partial V_r / \partial \lambda_r < 0 \) and \( \partial V_r / \partial \lambda_s = 0 \) hold for all \( s \neq r \) from (9).

Second, consider the other extreme case of near free trade. When transport costs are sufficiently small \( \tau \to 0 \), all regions are fully integrated, and thus, indirect utility \( V_r \),
becomes the same for all $r$. This means that any distribution of manufacturing workers is a spatial equilibrium when $\tau = 0$. Further, the NEG literature suggests that a fully agglomerated equilibrium occurs for sufficiently small transport costs.

Let $A_r^{agg} \equiv (0, \ldots, 0, 1, 0, \ldots, 0)$ be the agglomerated configuration of manufacturing workers in region $r$, where the $r$th argument is 1. I show the existence and stability of this full agglomeration equilibrium below. The agglomeration sustain conditions are given by

$$\Delta V_{rs} (\tau) \equiv V_r - V_s|_{A = A_r^{agg}} > 0, \quad \forall s.$$  

(12)

Since the utility differentials $\Delta V_{rs} (\tau)$ are zero for all $r$ and $s$ when $\tau = 0$, I consider the utility differentials for sufficiently small transport costs. Using the Taylor series expansion about $\tau = 0$ and ignoring the second and higher orders, we have

$$\Delta V_{rs} (\tau) = \tau [\mu (2\sigma - 1) d_{rs} + (1 - \mu) (\sigma - 1) (D_s - D_r)].$$  

(13)

where $D_r \equiv \sum_{t=1}^{n} a_t d_{rt}$ is the sum of the distances to all immobile farmers from region $r$. See Appendix 2 for derivation of (13).

We have

$$\Delta V_{rs} (\tau) \gtrless \Delta V_{sr} (\tau) \iff D_r \gtrless D_s.$$  

If $D_r$ is the smallest, then the last term in (13) is nonnegative, and hence (13) is positive for all $s \neq r$. That is, the agglomeration sustain conditions (12) always hold, verifying that the fully agglomerated configuration in region $r$ is a stable equilibrium. This implies that agglomeration is likely to emerge in the best location from which to transport products to all of them. In this sense, this region $r$ is regarded as the geographical center of this economy.

On the other hand, if $D_r$ is the largest and is sufficiently large, then (13) becomes negative, and hence, the agglomeration sustain conditions (12) are violated. That means that agglomeration is unlikely to emerge in a peripheral region. In sum, agglomeration of economic activities is more likely to emerge in the geographical center than peripheral regions. I have thus shown the following in the neighborhood of the free trade condition.

**Proposition 2** When transport costs are sufficiently small, there exists a fully agglomerated stable equilibrium. The agglomeration is likely to emerge in the geographical center of the economy.
Proposition 2 says that manufacturing agglomeration is likely to occur in a region with large immobile demand and/or good access to all other regions. However, full agglomeration in a region other than the geographical center can also be a stable equilibrium. A remote region with few farmers could accommodate agglomeration. For example, if the manufacturing share $\mu$ is very large, then the first term on the RHS of (13) becomes large relative to the geographical advantages of the second term. In this case, no individual worker has an incentive to move once full agglomeration has occurred by historical accidents. This is so-called the lock-in effect in NEG.

The two propositions approximate the early and late periods of the historical trends experienced in Brazil, France, Great Britain, Italy, Japan, Spain, and postwar Canada (see Figures 2a–2g). Thus, I conclude that these NEG models are capable of replicating the real-world tendency for urban agglomeration to the primate city.

The results obtained thus far hold for a family of the Dixit-Stiglitz type of NEG models such as Forslid and Ottaviano (2003) and Pflüger (2004). In other words, the stable equilibrium distribution of economic activities is characterized for any number of regions, any set of interregional distances, and any distribution of immobile farmers when transport costs are close to zero or infinity. Through the continuity of the overall model, similar outcomes may be expected when transport costs are small or large. It is therefore of interest to investigate and characterize the equilibrium distribution in the case of intermediate transport costs. However, it may no longer be possible to obtain meaningful analytical results in such a general setting.

Finally, employing the fact that the nominal wages of workers and farmers are shown to be equal to one for the symmetric and agglomerated configurations, it can be shown that (i) workers prefer agglomeration to symmetric dispersion, (ii) farmers in the core prefer agglomeration to dispersion whereas farmers in the periphery prefer dispersion to agglomeration, and (iii) neither the agglomerated configuration nor the symmetric configuration is shown to be Pareto dominant. These results were shown by Charlot et al. (2006) for two-region models, and are subsequently extended to an arbitrary number of regions here. Moreover, these results also hold in the models of Forslid and Ottaviano (2003) and Pflüger (2004).
3.2 Symmetric regions

In order to extract further results and generate more insights into falling transport costs, in this section I assume that regions are symmetric such that \( a_r = 1/n \) for all \( r \) and

\[
T_{rs} = \begin{cases} 
T & \text{if } r \neq s, \\
1 & \text{if } r = s.
\end{cases}
\]

where \( T > 1 \) and \( \phi \equiv T^{1-\sigma} \in [0,1) \). Farmers are evenly distributed across regions and the transport cost of one unit of good is the same regardless of the origin and destination. The former assumption may be acceptable if regions are divided in a manner that ensures that the numbers of farmers are equal in all regions. The second one may be justified by the fact that distance-related shipping costs are low, whereas distance-unrelated costs of insurance, loading and unloading are relatively high (Boyer, 1997). Therefore, I focus on an arbitrary number of fully symmetric regions and assess what happens when the fully symmetric equilibrium breaks.

As a thought experiment, I consider trade costs to steadily fall, namely the freeness of trade gradually increases from 0 (autarky) to 1 (free trade).\textsuperscript{6} First, the symmetric configuration defined by \( \Lambda_{\text{sym}} \equiv (1/n, \ldots, 1/n) \) is clearly a spatial equilibrium for any values of the freeness of trade. The first task is thus to study the conditions under which this symmetric configuration is a stable equilibrium as follows.

From Appendix 3(i), the following lemma is obtained:

**Lemma 1** Assume that the no-black-hole condition (10) holds. The symmetric equilibrium is stable only if

\[
0 \leq \phi < \phi_B \equiv \frac{(1-\mu)(\sigma-1-\mu\sigma)}{(1-\mu)(\sigma-1-\mu\sigma) + n\mu(2\sigma-1)}.
\]

Here, \( \phi_B \) is called the symmetry break point and is in the interval of \((0,1)\) under the no-black-hole condition (10).

Next, we pay attention to the full agglomeration equilibrium \( \Lambda_{\text{agg}}^n \equiv (0, \ldots, 0, 1) \). From Appendix 3(ii), the following lemma is shown:

\textsuperscript{6}The thought experiment may be conducted with respect to manufacturing share \( \mu \) rather than to trade costs \( \tau \) in order to describe the structural change in industrial composition or rural-to-urban migration. The main conclusions of this paper remain the same, however.
Lemma 2 The agglomerated equilibrium is stable if $\phi > \phi_S$ or if the no-black-hole condition (10) is violated.

Here, $\phi_S$ is called the agglomeration sustain point and is implicitly defined in Appendix 3. The research interest of this paper is assessing what happens when the symmetry breaks in the case of an arbitrary number of regions. Because it is not possible to find and characterize all equilibria when the symmetry breaks, this paper proceeds by investigating the full agglomeration $\Lambda_n^{agg}$ only. In other words, the question is whether the full agglomeration is a stable equilibrium when the symmetry breaks at $\phi = \phi_B$. This can be answered by finding whether the following inequality holds:

$$g|_{\phi=\phi_B} < 1,$$

where $g$ is defined in Appendix 3. Note that the LHS is a function of the three parameters: $n$, $\mu$ and $\sigma$. The full agglomeration is a stable equilibrium if inequality (14) holds. Similar to Robert-Nicoud (2005), this statement can be proven in the following manner. First, $h \equiv \log g|_{\phi=\phi_B}$ is shown to be decreasing in $n$ by differentiation. Then, by letting $\tilde{h} \equiv h|_{n=2}$, it can be readily verified that $\partial^2 \tilde{h}/\partial \mu^2 < 0$ for all $0 < \mu < \mu_{\text{max}}$ and $\partial \tilde{h}/\partial \mu|_{\mu=0} = \tilde{h}|_{\mu=0} = 0$. Hence, I get $\tilde{h} < 0$ and $h < 0$ for all $0 < \mu < \mu_{\text{max}}$.

Thus, this procedure has extended the well-known NEG result on dispersion and agglomeration by Krugman (1991) to an arbitrary number of fully symmetric regions as follows:

**Proposition 3** When trade costs steadily fall, the symmetric equilibrium breaks and full agglomeration to one city occurs for any parameter value of $n$, $\mu$ and $\sigma$ in Krugman’s (1991) model that has an arbitrary number of regions.$^7$

An identical result can be shown to hold in the case of Forslid and Ottaviano (2003) that has an arbitrary number of regions. A similar result also holds in the case of Pflüger (2004) that has an arbitrary number of regions (see Appendix 4). In the latter case, when the symmetry breaks, either partial agglomeration or full agglomeration occurs

---

$^7$For three regions, this proposition is the same as Proposition 3 in Castro, Correia-da-Silva and Mossay (2012).
depending on the parameter values of \( n, \mu \) and \( \sigma \). Partial agglomeration is expressed as
\[
\Lambda_{\text{par}}^n = \left( \frac{1-\lambda}{n-1}, \ldots, \frac{1-\lambda}{n-1}, \lambda \right)
\]
with \( 0 < \frac{1-\lambda}{n-1} < \lambda \), implying that the number of workers in one region is larger than that in the remainder of each region. The following theorem is thus implied from these three major NEG models:

**Theorem 1** When trade costs steadily fall, the symmetric equilibrium breaks and partial or full agglomeration to one city occurs for any parameter value of \( n, \mu \) and \( \sigma \) in the models of Krugman (1991), Forslid and Ottaviano (2003), and Pflüger (2004) that have an arbitrary number of regions.

Because regional populations consist of farmers and workers, the population distributions of the symmetric and agglomerated configurations are given by \( \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \) and \( \left( \frac{1-\mu}{n}, \ldots, \frac{1-\mu}{n}, \frac{1+(n-1)\mu}{n} \right) \), respectively. In the latter, while the manufacturing workers are fully agglomerated, the population is partially agglomerated due to the existence of immobile farmers.

### 4 Conclusion

There is consensus that while regional populations were dispersed in early times, they have been increasingly concentrated into one capital region over recent years. Although two-region NEG models can depict such an agglomeration tendency, the two-region economy itself is unrealistic. This paper thus extended the two-region NEG models to multi-region ones, which can reasonably simulate these historical trends of urban agglomeration to the primate city. It also shows that the multi-region models of NEG are similar to the two-region ones in terms of market outcomes.

I have been abiding by the traditional NEG models of Krugman (1991), Forslid and Ottaviano (2003), and Pflüger (2004). One may think of several extensions by introducing other dispersion forces such as urban costs, taste heterogeneity and migration costs. However, due to analytical difficulties of asymmetric regions, they are left for future research.
Appendix 1: Data sources and the largest cities

Figure 1a–1b: United Nations

World Urbanization Prospects, the 2011 Revision, Data on Cities and Urban Agglomerations (http://esa.un.org/unup/unup/index_panel2.html).

Figure 2a: Brazil

Instituto Brasileiro de Geografia e Estatística (http://seriesestatisticas.ibge.gov.br/).

Figure 2b: France

Population of metropolitan France at the census, by region, Institut National de la Statistique et des Études Économiques

Figure 2c: Great Britain

For 1701–1951. Table 7.1 in Lee (1986).

Figure 2d: Italy


Figure 2e: Japan

For 1721–1846. Table 1 in Kito (1996).

Figure 2f: Spain


Figure 2g: Canada

Statistics Canada. Table 051-0001 - Estimates of population, by age group and sex for July 1, Canada, provinces and territories, annual.

Figure 2h: United States of America

The largest cities in the top 30 countries by national GDP in 2010 are New York in the USA, Shanghai in China, Tokyo in Japan, Mumbai in India, Berlin in Germany, Moscow in Russia, London in the UK, São Paulo in Brazil, Paris in France, Rome in Italy, Mexico City in Mexico, Seoul in South Korea, Madrid in Spain, Toronto in Canada, Jakarta in Indonesia, Istanbul in Turkey, Teheran in Iran, Sydney in Australia, Warsaw in Poland, Amsterdam in the Netherlands, Buenos Aires in Argentina, Ar-Riyadh in Saudi Arabia, Bangkok in Thailand, Johannesburg in South Africa, Cairo in Egypt, Karachi in Pakistan, Bogota in Colombia, Kuala Lumpur in Malaysia, Brussels in Belgium, and Lagos in Nigeria.

Appendix 2: Derivation of equation (13)

When \( \Lambda = \Lambda^\text{agr} \), it can be easily shown that

\[
P_s = w_r T_{rs}, \quad \forall s
\]
\[
Y_r = \mu + (1 - \mu) a_r \quad \text{and} \quad Y_s = (1 - \mu) a_s, \quad \forall s \neq r.
\]

Substituting them into (7) yields

\[
w_r = 1
\]
\[
w_s = \left[ \frac{\mu}{\phi_{rs}} + \sum_t (1 - \mu) a_t \frac{\phi_{rt}}{\phi_{st}} \right]^{1/\sigma}, \quad \forall s \neq r.
\]

When \( \tau = 0 \), \( w_r = w_s \) obviously holds. Plugging these wages into the utility differentials, we have

\[
\Delta V_{rs} (\tau) \equiv V_r - V_s|_{\Lambda=\Lambda^\text{agr}} = 1 - \phi_{rs}^\mu \left[ \frac{\mu}{\phi_{rs}} + \sum_t (1 - \mu) a_t \frac{\phi_{rt}}{\phi_{st}} \right]^{1/\sigma}.
\]

Using \( \phi_{rs} = \exp [(1 - \sigma) \tau d_{rs}] \), applying the Taylor series expansion about \( \tau = 0 \), and ignoring the second and higher orders,

\[
\Delta V_{rs} (\tau) = \Delta V_{rs} (0) + \tau \Delta V'_{rs} (0) + o(\tau^2)
\]
\[
\approx \tau \left[ \mu (2\sigma - 1) d_{rs} + (1 - \mu) (\sigma - 1) (D_s - D_r) \right].
\]
Appendix 3: Proof of Lemmas

(i) Proof of Lemma 1

I substitute $\lambda_n = 1 - \sum_s \lambda_s$ into $J_r$ so that there are $n - 1$ equations $J_r = 0$ and $n - 1$ variables $\lambda_1, \ldots, \lambda_{n-1}$. Because $J_r$ is also a function of $w_1, \ldots, w_n$, the Jacobian should be computed as

$$\frac{dJ_r}{d\lambda_s} = \frac{\partial J_r}{\partial \lambda_s} + \sum_t \frac{\partial J_r}{\partial w_t} \frac{\partial w_t}{\partial \lambda_s}. \tag{15}$$

Since $w$’s are determined by the wage equations (7), their marginal changes can be computed by the implicit function theorem as follows:

$$\begin{pmatrix} \frac{\partial w_1}{\partial \lambda_s} \\ \vdots \\ \frac{\partial w_n}{\partial \lambda_s} \end{pmatrix} = \left( \begin{pmatrix} \frac{\partial K_1}{\partial w_1} & \cdots & \frac{K_1}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial K_n}{\partial w_1} & \cdots & \frac{\partial K_n}{\partial w_n} \end{pmatrix} \right)^{-1} \begin{pmatrix} \frac{\partial K_1}{\partial \lambda_s} \\ \vdots \\ \frac{\partial K_n}{\partial \lambda_s} \end{pmatrix},$$

where $K_r$ defined by the LHS of (7). Substituting the above into the Jacobian (15) and evaluating it at $\Lambda = \Lambda_{\text{sym}}$ yields

$$\frac{dJ_r}{d\lambda_s} \bigg|_{\Lambda = \Lambda_{\text{sym}}} = \frac{\partial J_r}{\partial \lambda_s} + \sum_t \frac{\partial J_r}{\partial w_t} \frac{\partial w_t}{\partial \lambda_s} \bigg|_{\Lambda = \Lambda_{\text{sym}}} \forall r, s = 1, \ldots, n - 1$$

$$= \begin{cases} B (1 - \phi) (\phi - \phi_B) & \text{if } s = r \\ 0 & \text{if } s \neq r \end{cases} \forall r, s = 1, \ldots, n - 1,$$

where $B$ is a positive constant and $\phi_B$ is defined in Lemma 1.

(ii) Proof of Lemma 2

Plugging $\Lambda = \Lambda_n^{\text{agg}}$ into the $n$th wage equation (7), one gets $w_n = 1$. Substituting $\Lambda = \Lambda_n^{\text{agg}}$ and $w_n = 1$ with $w_r = w$ for all $r = 1, \ldots, n - 1$ into the other wage equations yields

$$w = \left[ \frac{1 + (n - 2) \phi + \phi^2 - \mu (1 - \phi) (1 + n \phi)}{n \phi} \right]^{1/\sigma}.$$

Therefore, the agglomeration at the symmetry break point $\phi = \phi_B$ is sustained if

$$V_n - V_1 \bigg|_{\Lambda = \Lambda_n^{\text{agg}}, \phi = \phi_B, w_1 = \cdots = w_{n-1} = w, w_n = 1} = 1 - g^{1/\sigma} > 0, \tag{16}$$

which is (13).
where
\[
g \equiv \frac{\phi^2}{n - 1} \left[ 1 + (n - 2) \phi + \phi^2 - \mu (1 - \phi) (1 + (n - 1) \phi) \right].
\]

The inequality (16) can be rewritten as \( \phi > \phi_S \), where \( \phi_S \) is implicitly defined by the unique solution of \( g = 1 \).

**Appendix 4: Pflüger’s (2004) model with multiple regions**

Extending Pflüger’s (2004) model to three or more regions, I get the indirect utility of a consumer living in region \( r \):

\[
V_r = \frac{1}{\sigma - 1} \log \left( \sum_s \lambda_s \phi_{sr} \right) + \frac{1}{\sigma} \sum_s \lambda_s + \frac{1 - \mu}{\mu \sigma} \sum_t \phi_{ts}^r \phi_{sr} + \text{constant}.
\]

The symmetry break point is computed as
\[
\hat{\phi}_B \equiv \frac{\sigma - 1 - \mu (2\sigma - 1)}{\sigma - 1 - \mu (2\sigma - 1) (n - 1)} \in (0, 1)
\]

and the no-black-hole condition is given by
\[
\mu < \hat{\mu}_{\text{max}} \equiv \frac{\sigma - 1}{2\sigma - 1}.
\]

Let \( \hat{\mu}_{\text{min}} \) be a unique solution of \( V_n - V_1|_{\lambda = \lambda^\text{agg}, \phi = \hat{\phi}_B} = 0 \). Because \( \mu_{\text{min}} < \mu_{\text{max}} \) holds, the interior equilibrium configuration with \( n - 1 \) small cities and 1 large city is shown to be stable for all \( \mu_{\text{min}} < \mu < \mu_{\text{max}} \) and that the fully agglomerated configuration is a stable equilibrium for all \( 0 < \mu \leq \mu_{\text{min}} \). Hence, I establish the following proposition.

**Proposition 4** When the trade costs steadily fall, the symmetric equilibrium breaks and partial or full agglomeration in one city appears for any parameter values of \( \mu \) and \( \sigma \) in Pflüger’s (2004) model with three or more regions. More precisely, at \( \phi = \hat{\phi}_B \),

(i) if \( \mu \geq \hat{\mu}_{\text{max}} \), the no-black-hole condition (17) is violated and there exists a stable equilibrium with full agglomeration in one city;

(ii) if \( \hat{\mu}_{\text{max}} > \mu > \hat{\mu}_{\text{min}} \), there exists a stable equilibrium with one big city and \( n - 1 \) small cities of equal size;
(iii) if $0 < \mu \leq \hat{\mu}_{\text{min}}$, there exists a stable equilibrium with full agglomeration in one city at $\phi = \hat{\phi}_B$.

A proof of the proposition is available upon request to the author.

References


Figure 1a:  Increasing population share of the largest metropolitan area

Figure 1b:  Decreasing population share of the largest metropolitan area
Figure 2e: Regional population share in Japan

Figure 2f: Regional population share in Spain

Figure 2g: Regional population share in Canada

Figure 2h: Regional population share in USA