INSTITUTE OF DEVELOPING ECONOMIES

Discussion Papers are preliminary materials circulated
to stimulate discussions and critical comments

DISCUSSION PAPER No. 21

Transport Development and the Evolution of Economic Geography*

Masahisa Fujita†,‡ and Tomoya Mori §

January 2005

Abstract

In this paper, based on the recent advances in the new economic geography (e.g., Fujita, Krugman and Venables [12]), we analyze impacts of transport costs on the spatial patterns of economic agglomeration. We first identify prototypes from the existing models, and explain the mechanism of how transport costs influence the balance between economic forces of agglomeration and dispersion. We then investigate the transformation of the agglomeration/dispersion patterns given gradually decreasing transport costs for different goods.

Keywords: new economic geography, transport cost, industrial belt

JEL classification: R12

* The authors are grateful to David Bernstein, Tatsuo Hatta, Komei Sasaki, Tony E. Smith, and two anonymous referees for their valuable comments. This research is partly supported by The Grant in Aid for Research 08403001 of Ministry of Education, Science and Culture in Japan, the Murata Science Foundation, and WESCO Civil Engineering Technology Foundation.

† Institute of Economic Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501 Japan. Phone: +81-75-753-7122. Fax: +81-75-753-7198. Email: fujita@kier.kyoto-u.ac.jp.

‡ Institute of Developing Economies, JETRO, 3-2-2 Wakaba, Mihama-ku, Chiba, 261-8545 Japan. Phone: +81-43-299-9510. Fax: +81-43-299-9723.

§ Institute of Economic Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501 Japan. Phone: +81-75-753-7121. Fax: +81-75-753-7198. Email: mori@kier.kyoto-u.ac.jp.

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, Middle East, Africa, Latin America, Oceania, and East Europe.

INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO
3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI
CHIBA 261-8545, JAPAN

1 Introduction

This paper presents the impact of transport costs on the spatial patterns of economic agglomeration. In particular, we focus on the recent advances in the so-called *new economic geography* (see Fujita et. al [12] a comprehensive survey) which explains the population and industry localization by endogenously generating agglomeration economies through interactions among increasing returns, transport costs, and factor mobility.

Although it has long been pointed out that agglomeration economies play an important role in explaining the localization patterns of population and industries (e.g., Marshall [35], Isard [24], Meyer [36], Cronon [3]), the formalization of the theory has not much proceeded in this direction until the recent outbreak of the new economic geography triggered by Krugman [27].[1] The modeling approach of the new economic geography is very specific in that the agglomeration force is generated by the use of the Dixit-Stiglitz [4] type monopolistic competition model in which the agglomeration force is derived from the monopolistically competitive behavior of many small firms which produce varieties of differentiated goods, and that the transportation of goods is subject to Samuelson's [44] iceberg technology (i.e., transport costs are incurred in the goods shipped).[2] Nonetheless, this approach has unraveled, in a general equilibrium context, the mechanism of several important phenomena of the economy's spatial organization such as the emergence of a core-periphery structure corresponding to the north-south dualism and the formation, spatial distribution and industrial composition of cities.[3]

In such models with endogenous agglomeration economies, the delicate balance between the agglomeration and dispersion forces determines the spatial structure of the economy. In particular, recent studies (e.g., Krugman [27][29], Helpman [19], Mori [38]; Tabuchi [46]) have revealed that the balance between these two opposing forces is critically influenced by transport costs, and that the impacts of transport cost changes on the spatial structure of the economy drastically differ depending on the nature of agglomeration and dispersion forces involved. Especially, it turned out that the interaction of the same agglomeration force with different types of dispersion forces may lead to quite different spatial organizations of the economy. In these studies, however, while the mechanism of the spatial agglomeration has been well investigated, no systematic analysis has been conducted to explain the mechanism of spatial dispersions. Consequently, the role of transport costs in determining the economic geography has been unveiled only fragmentarily.

[1] On one hand, one stream of the traditional urban economic theory (e.g., Mills [37], Papageorgiou and Thisse [42], Fujita [7]) has been successful in explaining the formation of a central business district within a metropolitan area in a partial equilibrium context. However, it has little to say about the relative location, size and specialization pattern of each metropolis within an economy. On the other hand, while the other stream which particularly called the urban systems theory led by Henderson [21][22][23] has been able to generate different specialization patterns and sizes among cities in a general equilibrium context, this approach lacks the spatial dimension. For a detailed review of these theories, see Fujita [8] and Fujita and Thisse [16].

[2] See Ottaviano, Tabuchi and Thisse [41] for an alternative formulation of monopolistic competition using quasi-linear utility function with non-iceberg transport costs.

[3] For the emergence of a core-periphery structure, see Krugman [27][28]. For the formation of a city, see Krugman [29] and Fujita and Krugman [10]. For the spatial distribution and industrial composition of cities, see Fujita and Mori [14] and Fujita et. al. [11], respectively.

In the present paper, we hi-light the nature of dispersion forces incorporated in the new economic geography models. The agglomeration force considered here is the standard one in the new economic geography, i.e., that based on the monopolistic competition model with a variety in consumption goods. The mechanism of agglomeration economies can be described as follows. The production of each variety of differentiated goods (D-goods) is subject to an increasing returns technology using labor as a sole input. Due to scale economies at the plant level, then each variety of D-goods is produced at a single site, while because of transport costs, suppliers prefer to locate closer to markets. On the other hand, the presence of transport costs makes consumers (who are workers at the same time) concentrate at a location where a wider variety of products is available, i.e., the location at which firms producing differentiated goods (D-firms) agglomerate. The concentration of workers thus generates a large demand for D-goods as well as a large labor pool for D-firms, which further attracts more D-firms to the location.

The dispersion force can be generated by incorporating immobile resources in some forms. The models of new economic geography, in particular, introduce homogeneous consumption goods (H-goods) which embody immobile resources. In this context, two types of dispersion forces, i.e., the *local demand pull* and *factor price pull*, may be derived. First, the local demand pull arises when some consumers are inevitably dispersed over space. For instance, H-goods are represented by agricultural goods in Krugman [27] and Fujita and Krugman [10]. In the former, the agricultural goods are produced by a fraction of the total population who is spatially dispersed and immobile, while in the later, they are produced by using labor and land as inputs, where the land is immobile and distributed over space, while labor is supplied by freely mobile workers. In either case, the demand for D-goods generated by workers employed in the agricultural sector is inevitably dispersed over space, and hence a D-firm may find it profitable to move away from competitors and locate in the agricultural area serving primarily this local market where the intensity of competition is lower.

The factor price pull may become a dispersion force when H-goods themselves or whose production inputs are spatially dispersed and immobile, and the transportation of H-goods is costly. An example of the former is the land for housing in Helpman [19], and that for the latter is the agricultural good in the Fujita-Krugman model which is produced using land as an input. In either case, the concentration of D-firms and workers at one location raises the price for these land-intensive goods there, which in turn increases the wage rate there. Thus, a D-firm may be attracted to deviate from the concentration of other D-firms in order to enjoy the labor cost advantage near the production site of land-intensive H-goods.

In this paper, we investigate how the balance between the agglomeration force and these two types of dispersion forces changes and affect the spatial structure of the economy when transport costs gradually decrease. Since the existing studies have already accumulated enough results for the impact of changes in transport cost for D-goods, we first summarize and interpret their findings. We then proceed to investigate the case in which transport costs for both D- and H-goods change by using the Fujita-Krugman framework, in which both types of dispersion forces are incorporated. In particular, we adopt a continuous location space in order to analyze the spatial structure/distribution of agglomeration in a pure form. In the Fujita-Krugman framework, a city is defined as the location of D-firms.

It will be demonstrated that if the transport cost for D-goods decreases relatively faster than that for H-goods, then while D-firms may first agglomerate into a smaller number of (discrete) cities, they eventually disperse over an interval leading to the formation of a *megalopolis* which consists of large discrete cities (or, *core cities*) connected by an *industrial belt* (*a continuum of cities*). Moreover the megalopolis expands given a further relative decrease in the transport cost for D-goods. On the other hand, if the transport cost for H-goods decreases relatively faster, then D-firms agglomerate into a smaller number of cities.

The plan of the paper is as follows. Section 2 summarizes the basic structure of the models of a spatial economy based on the monopolistic competition model with a variety in consumption goods. In section 3, the nature of agglomeration and dispersion forces are presented. In particular, the sources and effects of the two types of dispersion forces are compared. Section 4 introduces the adjustment mechanism which governs the stability and reorganization of the spatial structure of the economy when transport costs change. Section 5, discusses the impact of the decrease in transport costs for D-goods based on the works by Krugman [27][29], Helpman [19], Mori [38] and Tabuchi [46]. In section 6, by using the Fujita-Krugman framework, new findings are presented regarding the impact of simultaneous changes in transport costs for both D- and H-goods on the organization of the economic geography. Concluding remarks for future research directions end the paper.

2 The basic model

In this section, we introduce the common structure of models of a spatial economy which incorporate the Dixit-Stiglitz type monopolistic competition model with varieties of consumption goods.

Consider an economy which produces two types of consumption goods. One is an H-good produced under a constant-returns technology and perfect competition, while the other are varieties of D-goods, each of which is produced under an increasing-returns technology and monopolistic competition.

Every consumer in the economy shares the same tastes, and her welfare is represented by a Cobb-Douglas function of consumption of H-goods, Z_H, and a composite of D-goods, Z_D:

$$U = Z_D^{\mu} Z_H^{1-\mu}, \tag{2-1}$$

where $\mu \in (0,1)$. When a continuum of varieties of size n is supplied, the composite of D-goods takes the following CES function of consumption of each differentiated variety:

$$Z_D = \left\{ \int_0^n z(\omega)^{(\sigma-1)/\sigma} d\omega \right\}^{\sigma/(\sigma-1)}, \tag{2-2}$$

where σ (>1) is the elasticity of substitution, i.e., a smaller value of σ implies that consumers have a stronger preference for variety. Since the number of firms producing D-goods is assumed to be large, the impact of a price change by any individual firm on the total demand for each D-good is negligible, and hence the price elasticity of an individual demand for each variety is also given by a constant, σ.

The production of each D-good uses labor as a sole input in such a way that the total amount of labor, $L(\omega)$, required to produce any individual variety ω of quantity $z(\omega)$ is given by the sum of fixed labor input, α, and variable labor input, $\beta z(\omega)$:

$$L(\omega) = \alpha + \beta z(\omega), \tag{2-3}$$

where α and β are given positive constants.

Next, given a location space, $\mathbb{X}$, it is assumed that there are costs of transporting goods. In particular, transport costs take Samuelson's iceberg form, in which only a fraction of a good shipped arrives. For instance, if a unit of a good is shipped over a distance d, only $e^{-\tau d}$ units reach the destination, where τ is a positive constant. It follows that the transport cost is multiplicative to the mill price of a good: if the mill price of a good is p, then its delivered price at a distance d from the location of the firm is $pe^{\tau d}$.

Due to this multiplicative nature of the transport cost, the price elasticity of an individual demand is independent of location, which in turn implies that the elasticity of aggregate demand is also independent of the spatial distribution of consumers, and equals the price elasticity of an individual demand, σ. It follows that the profit-maximizing mill price, $p_D(x)$, of each D-good produced at location $x \in \mathbb{X}$ is given by

$$p_D(x) = \frac{\beta}{1 - 1/\sigma} W(x), \tag{2-4}$$

where $W(x)$ is the wage rate prevailing at x.

Finally, the free entry of D-firms assures their zero profit in equilibrium, which results in an equilibrium output level, q, of each firm given by a constant:

$$q = (\sigma - 1)\,\alpha/\beta. \tag{2-5}$$

It follows that the equilibrium labor input of each D-firm is also a constant given by

$$L = \alpha\sigma. \tag{2-6}$$

3 Agglomeration force and dispersion force

In the presence of transport costs, consumers and suppliers of D-goods attract each other due to consumers' love of variety in D-goods and increasing returns in the D-good production. Working against this agglomeration force, a dispersion force may arise from the desire of firms to move away from competitors, or that to move to a remote location which offers a lower labor cost. In the two subsections below, we explain in detail the mechanisms that generate these agglomeration and dispersion forces.

3.1 Agglomeration force

In the context of the model introduced in the previous section, there exist two sources of agglomeration force. One is the consumers' love of variety in D-goods. Namely, the utility level of each consumer increases with the size n of varieties. To see this, suppose each variety is sold at the same price p, and let Y denote the fraction of a consumer's income spent on D-goods. Then, the consumption of each variety is given by $z(\omega) = Y/(np)$ for each $\omega \in [0, n]$. It follows that by (2-2) the utility from the consumption of D-goods can be measured by $Yn^{1/(\sigma-1)}/p$, which increases with n. In particular, this increase is larger if goods are more differentiated, i.e., σ is smaller.

The other sources of agglomeration force is increasing returns in the production of each D-good. That is, a larger number of D-firms can cluster at a location, only if the location can offer a sufficiently large demand for their products as well as a sufficiently large labor pool. In other words, producers of D-goods tend to cluster at the location where a large number of consumers (who are workers as well) concentrate.

These two sources of the agglomeration force generate a circular causation of agglomeration economies: a concentration of a larger number of D-firms at a given location implies a larger number of available D-goods there. Due to love of variety, then, more consumers are attracted to the location, which in turn provides a larger demand for D-goods as well as a larger pool of workers. As a result, an even larger number of specialized D-firms can be supported at that location. Agglomeration of more D-firms leads to a further expansion of the product variety available at that location.

3.2 Dispersion force

The dispersion force which works against the agglomeration force can be generated by introducing spatially dispersed immobile resources in some forms. Figure 3.1 identifies prototype models in terms of the way such immobile resources are incorporated. As the figure indicates, one way is to assume that a fixed fraction of the total population is dispersed over space and immobile engaging in the production of H-goods. In the two-region model by Krugman [27], H-goods represent agricultural goods produced by immobile farmers who are evenly distributed between the regions. Another way is to explicitly consider the land use for the purpose of production and/or consumption. In Fujita and Krugman [10], H-goods represent agricultural goods which are produced subject to the Leontief technology using mobile labor and immobile land uniformly distributed over a continuous location space. In the two-region model by Helpman [19], on the other hand H-goods represent the land for housing which is immobile and distributed between the regions.[4] In these ways, the prototype models incorporates either H-goods themselves or inputs for their production as immobile resources.

[4] In Krugman and Livas Elizondo [33] which investigates the impacts of international trade on the industrial localization pattern in the domestic economy, the model which is qualitatively the same as Helpman's is used to describe the spatial structure of the domestic economy. The difference is that in order to generate a dispersion force, the Krugman-Livas Elizondo's model assume a commuting cost and consumption of a fixed amount of land by each worker in each location in the two-point location space rather than assuming zero commuting cost with a fixed amount of land available at each location as in the Helpman's model. See also Krugman [32] for the discussion of the nature of dispersion force which arises in the Krugman-Livas Elizondo model.

<table>
<tr><th>Source of agglomeration forces</th><th colspan="3">Source of dispersion forces</th><th>Prototype models</th></tr>
<tr><td rowspan="3">scale economies in D-production/ product variety of D-goods</td><td rowspan="3">immobile resource</td><td colspan="2">H-workers</td><td>Krugman [27]</td></tr>
<tr><td rowspan="2">land</td><td>H-production</td><td>Fujita and Krugman [10]</td></tr>
<tr><td>H-consumption</td><td>Helpman [19]</td></tr>
</table>

Figure 3.1. Prototypes of spatial monopolistic competition model with consumption good variety.

There are two types of dispersion forces, local demand pull and factor price pull, which may arise from these immobile resources. In the below, we explain how these dispersion forces work.

local demand pull

The local demand pull is generated when there exists a spatially dispersed demand for D-goods. Namely, because of spatially dispersed consumers, a firm producing D-good may find it more profitable to move away from competitors and serve the local demand at a remote location where the intensity of competition is lower. Among the prototype models, the Krugman's and Fujita-Krugman models include this dispersion force. Namely, in the Krugman's model, farmers producing H-goods are assumed to be distributed over the space, i.e., consumers are a priori assumed to be spatially dispersed. In the Fujita-Krugman model, while all workers are freely mobile, those who are employed in the land-intensive H-good sector (H-sector) are inevitably dispersed, and generate a spatially dispersed demand for D-goods.

factor price pull

The factor price pull may work as a dispersion force when the transportation of H-goods is costly. Namely, if the transport cost for H-goods is relatively high, freely mobile workers prefer to reside closer to their production site. Consequently, the wage rate tends to decrease at locations where the production of H-goods takes place. Hence, D-firms (who use labor as a sole input) may find it more profitable to locate closer to the production site of H-goods enjoying the lower labor cost there, than to agglomerate near competitors.

Among the prototype models, the Fujita-Krugman model incorporates this dispersion force by assuming a positive transport cost for agricultural goods, while the Helpman's model generates it by considering (untransportable) land for housing.[5] The Krugman's

[5]It is to be noted that the Helpman's model does not assume a priori spatially distributed demand. In his two region model, all consumers are freely mobile. But, they consume land for housing which is scarce, distributed between regions and immobile.

model, on the other hand, excludes it by assuming a zero transport cost for agricultural goods.

Now, let us illustrate how the local demand pull and factor price pull works as dispersion forces by using the model by Fujita and Krugman which contains both types of dispersion forces.

The location space, $\mathbb{X}$, in their case is represented by a continuous one-dimensional space, $[-\infty, \infty]$, over which land is uniformly distributed. The homogeneous work force of a given size N is endogenously allocated to the production of H-goods and D-goods. Now, let us assume *a priori* that all D-firms are concentrated at location $x = 0$, and call this concentration of D-firms *a city.* By using usual market conditions and by assuring the equal utility level of all workers, we can determine all unknowns associated with this *monopolar* configuration of the economy. In particular, the agricultural fringe distance, $l \in (0, \infty)$, is determined, so that the agricultural hinterland extends over interval $[-l, l]$, over which the price of the agricultural good at each location, y, is given by $p_H(y) = \exp(-\tau_H |y|)$, where $p_H(0)$ is normalized to be 1 and τ_H represents the transport rate for H-goods. Since each unit of H-good is assumed to be produced by 1 unit of land and a units of labor, the land rent at each location is given by $R(y) = \max\{p_H(y) - aW(y), 0\}$, where the equilibrium wage rate at each y can be obtained as

$$W(y) = (1/a)e^{-\mu(\tau_D+\tau_H)l+\{\mu\tau_D-(1-\mu)\tau_H\}|y|}. \tag{3-1}$$

Landlords are assumed to be attached to each location, and spend the entire revenue from the land rent on consumption of D- and H-goods at that location.

Let us now derive the so-called (*market*) *potential function* which measures the profitability of D-firms at each location relative to that of the city. Suppose a D-firm moves away from competitors at $x = 0$, and locates at any $x \in \mathbb{X}$. Let $D(x)$ represent the total potential demand for the product of this firm from the entire economy (under the mill price given by (2-4)), and let $\pi(x)$ represent the potential profit of the firm at location x. Then, it can be shown that $\pi(x) \gtreqless 0 \Leftrightarrow D(x) \gtreqless q$, where q is the (zero-profit) equilibrium output of each firm given by (2-5). The potential function, $\Omega(x)$, is then defined as

$$\Omega(x) \equiv D(x)/q. \tag{3-2}$$

Note that $\pi(x) \gtreqless 0 \Leftrightarrow \Omega(x) \gtreqless 1$. Since D-firms can earn zero profit in the city, they have no incentive to move away from the city if $\Omega(x) < 1$ for all $x\,(\neq 0) \in \mathbb{X}$, while they are better off by deviating to a location $x\,(\neq 0)$ if $\Omega(x) > 1$. A more precise form of the potential function for the monopolar configuration can be given as follows:

$$\Omega(x) = (1/q)\,W(x)^{-\sigma}\left\{\frac{\mu W(0)N_D}{np_D(0)^{1-\sigma}}e^{(1-\sigma)\tau_D|x|} + \int_{-l}^{l}\frac{\mu p_H(y)e^{(1-\sigma)\tau_D|x-y|}}{np_D(0)^{1-\sigma}e^{(1-\sigma)\tau_D|y|}}dy\right\}, \tag{3-3}$$

which is symmetric with respect to the city, i.e., $x = 0$. Each term in (3-3) can be interpreted as follows: $W(x)^{-\sigma}$ represents the *labor cost advantage* of the location x; $\mu W(0)\,N_D$ [resp., $\mu p_H(y)$] the *market size* (for D-goods) at the city [resp., at location y in the agricultural area]; $\exp([1-\sigma]\,\tau_D\,|x|)$ [resp., $\exp([1-\sigma]\,\tau_D\,|x-y|)$] the *accessibility to the market* in the city [resp., at y] from the production location x; $np_D(0)^{1-\sigma}$ [resp.,

$np_D(0)^{1-\sigma}e^{[1-\sigma]\tau_D|y|}$] the *intensity of competition* in the D-goods market at the city [resp., at y].

In the potential function, we can readily see how the local demand pull attracts D-firms to the agricultural area. While the market size in the city may be larger than that at a remote location in the agricultural area, the intensity of competition is higher in the city. Thus, if the intensity of competition in the agricultural area is sufficiently low, it may be profitable for D-firms to locate in the agricultural area, and serve this local market. Note that the existence of local demand (by farmers and landlords) in the agricultural area is essential for the deviating firm to reap the benefit of low intensity of competition there.

The way the factor price pull works is reflected in the term expressing the labor cost advantage, $W(x)^{-\sigma}$. By (3-1), if $\tau_D < \tau_H(1-\mu)/\mu$, then $dW(x)/dx < 0$, i.e., $d\left(W(x)^{-\sigma}\right)/dx > 0$, and hence the labor cost advantage increases as a D-firm moves away from the city. Namely, if the transport cost for D-goods is relatively low, given the same wage rate over space, the utility level tends to be higher in the agricultural area where H-goods are cheaper. It follows that the wage rate tends to be lower in the agricultural area, and hence this labor cost advantage attracts D-firms from the city.

The balance between the agglomeration and dispersion force in the vicinity of the city can be seen by computing the slope of the potential curve at the city:

$$\Omega'(0_+) \equiv \lim_{x\to 0} \partial\Omega(x)/\partial x = -\sigma\left\{(2-1/\sigma)\,\mu\tau_D - (1-\mu)\,\tau_H\right\}. \tag{3-4}$$

Thus, we have

$$\Omega'(0_+) \gtreqless 0 \Leftrightarrow \tau_D \lesseqgtr (1/\mu - 1)\,/\,(2-1/\sigma)\,\tau_H. \tag{3-5}$$

It follows that in order for the concentration of D-firms at $x=0$ to be sustained, the relative transport cost for D-goods needs to be sufficiently high so that $\Omega'(0_+) < 0$. In other words, the labor cost advantage in the agricultural area must be sufficiently small. It is to be noted, however, the dispersion of D-firms do not necessarily take place in the vicinity of the city, but it may take place at a remote location in the agricultural area where the intensity of competition is sufficiently low (recall that the competition intensity becomes lower as a D-firm moves away from the city). In fact, it can be shown that the shape of the potential curve over the interval, $[0,l]$, is first convex and the rest is concave to the below.[6]

4 Stability and adjustment of the spatial system

In this section, we introduce an adjustment mechanism of population distribution which governs the stability and transformation of the economy's spatial structure. In any of the prototype models and their extensions, the following adjustment mechanism or qualitatively the same variants are adopted. Namely, given a set of transport rates, (τ_D, τ_H),

[6] Since the potential curve for the monopolar system is symmetric with respect to the city location, $x=0$, and since the curve beyond the agricultural fringe distance, l, has no importance for our purpose, we can focus on the behavior of the curve over the interval, $[0,l]$.

suppose that each location i in a set $\mathbb{J} \equiv \{1, 2, \cdots, J\}$ of discrete locations are populated with workers employed by D-firms (D-workers) of population size, N_i, in a location space consist of a set $\mathbb{K} \equiv \{1, 2, \cdots, K\}$ of discrete locations where $K \geq J$. Then the population distribution is driven by the following adjustment mechanism:

$$\dot{N}_i = \phi N_i (u_i - \bar{u}) N \quad \text{for } i \in \mathbb{K}, \tag{4-1}$$

where $\bar{u} \equiv \sum_{j \in \mathbb{K}} u_j N_j / N$ is the average utility level of mobile workers given the population distribution, and ϕ is a positive constant.[7]

A population distribution, $\{N_i \mid i \in \mathbb{K}\}$, is in equilibrium if $\dot{N}_i = 0$ for all $i \in \mathbb{K}$. The equilibrium is said to be *stable* if any arbitrarily small perturbation given to the equilibrium population distribution eventually dies out following the adjustment process (4-1), otherwise it is said to be *unstable*, and the adjustment mechanism leads the population distribution to another adjacent stable equilibrium.

In the case of a continuous location space of the Fujita-Krugman model, the set of location indices needs to be modified in such a way that $\mathbb{K} = \mathbb{J} + \{J + 1\}$, where the $(J + 1)th$ location corresponds to any location in the agricultural area, i.e., all H-workers in the agricultural area are assumed to have the same utility level at each moment of time. That is, (4-1) also controls division of labor between differentiated good sector (D-sector) and H-sector.

Also, in order to accommodate the potential formation of new cities, the definition of the stability of an equilibrium in this case should be modified in the following way. Namely, a perturbation to an equilibrium needs to include *the creation of finitely many and arbitrarily small cities at arbitrary points on the location space.* Then, the equilibrium is said to be *stable* if this perturbation dies out under the adjustment process (4-1) defined for this extended set of cities and the agricultural area, otherwise it is said to be *unstable.*

In the context of the Fujita-Krugman model, it has been shown that *no new city is viable at any location x such that $\Omega(x) < 1$.* It follows that *a monopolar equilibrium is stable if the potential value is less than* 1 *everywhere except at the city* (Fujita and Mori [14] Lemma 5.1).

5 Impacts of a decrease in transport costs for differentiated goods

In this section, we present the impacts of changes in transport cost for D-goods on the spatial distribution of agglomerations (of D-firms and workers). Since the existing studies by Krugman [27][29], Helpman [19], Mori [38], and Tabuchi [46] have already provided enough results, our task here is to unify the implications from their findings. Figure 5.1 summarizes the correspondence between the three prototype models, two types of dispersion forces,[8] and impacts of changes in the transport cost for D-goods on the spatial distribution of D-firms.

[7]The adjustment of the population distribution is not affected by the utility level of immobile workers such as farmers in the Krugman's model, or that of immobile landlords of the Fujita-Krugman's model.

[8]More precisely, we mean by dispersion force the force that triggers the dispersion of D-sectors from their full agglomeration (i.e., core-periphery structure).

One extreme result is shown by Krugman [27] which embodies only the local demand pull as a dispersion force by assuming spatially dispersed immobile farmers with zero transport cost for their products, H-goods. In this context, it has been shown that there is a monotonic relation between the transport cost for D-goods and the spatial distribution of D-firms. Namely, in the two-region economy in which farmers are distributed evenly between the regions, D-firms agglomerate in one region given a transport cost for D-goods below some critical value, while they disperse over both regions evenly given a transport cost for D-goods above the critical value.

Krugman [29] has refined this result by extending his original two-region model to the many-(discrete)-region model. Namely, he has shown that D-firms agglomerate [resp., disperse] into a smaller [resp., larger] number of locations under a lower [resp., higher] transport cost for D-goods. In particular, the distance between each pair of agglomerations of D-firms is roughly the same. Here, a change in the magnitude of transport cost induces a change in the intensity of competition over space, which in turn influences the location of D-firms. Namely, given a lower transport cost for D-goods, the intensity of competition becomes more similar over space, and hence there is a smaller incentive for D-firms to relocate. Given a higher transport cost for D-goods, on the contrary, the intensity of competition decreases more rapidly as an D-firm moves away from competitors, and hence they tend to be attracted to a location with less number of competitors, i.e., they tend to disperse over space in order to enjoy the low intensity of competition. Consequently, for a lower transport cost for D-goods, the agglomeration force tends to dominate the dispersion force, and D-firms agglomerate at a larger scale into a smaller number of locations serving an extensive hinterland, while for a higher transport cost for D-goods, they tend to agglomerate at a smaller scale into a larger number of locations dispersed over space serving a smaller hinterland.

prototype models	local demand pull	factor price pull	transport cost for D-goods (lower ⟷ higher)
Krugman [27]	yes	no	agglomeration ¦ dispersion
Fujita and Krugman [10]	yes	yes	dispersion ¦ agglomeration ¦ dispersion
Helpman [19]	no	yes	dispersion ¦ agglomeration

Figure 5.1. Types of dispersion forces and the impact of changes in transport cost for D-goods on the spatial structure of the economy.

The other extreme result (opposite to Krugman's above) is obtained by Helpman [19]. In his model, an H-good represents land for housing. In a two-region economy, the dispersion force is not generated from the local demand pull as in the Krugman's model, but it is generated by the factor price pull due to the difference in rent for untransportable land allocated (and immobile) in each region. In this context, Helpman has found the monotonic relation between the magnitude of transport costs for D-goods and the spatial distribution of D-firms, which is opposite to the result obtained by Krugman. Namely, there is a critical value of transport cost for D-goods below which D-firms disperse over the two regions, and above which they agglomerate into one region. The underlying mechanism for this result can be understood as follows. When the transport cost for D-goods is lower, given freely mobile workers, the wage rate at each location tends to reflect the local land rent. Since the land rent is lower with a lower density of D-firms per unit of land, then, D-firms disperse over the two regions exploiting the labor cost advantage. On the other hand, given a higher transport cost, the wage rate in each region tends to reflect the local price of D-goods, and hence the region with more D-firms gets the labor cost advantage, and the agglomeration force is reinforced there.

Tabuchi [46] has combined the above two prototype models of Krugman's and Helpman's, and hence incorporated the two types of dispersion force in one model.[9] He has shown that in this case the relation between the magnitude of the transport cost for D-goods and the spatial distribution of D-firms becomes non-monotonic unlike the above two models. Namely, D-firms disperse over the two regions for a transport cost for D-goods which is either sufficiently high or sufficiently low, while they agglomerate into one region when the transport cost for D-goods takes an intermediate value. The mechanism causing such phenomena can be understood by combining the mechanisms generating the results in the Krugman's and Helpman's models.

Namely, when the transport cost for D-goods is low, since the local demand pull is not an effective dispersion force, since the intensity of competition is similar over space, the factor price pull incurred by the lower land rent in a less inhabited location tends to dominate. Thus, as in the Helpman's case, D-firms tend to disperse between the two regions seeking for a cheaper land rent. When the transport cost for D-goods is high, on the contrary, the high intensity of competition near competitors tends to disperse D-firms. Notice that since the wage rate tends to be lower in the region where D-firms concentrate, the factor price advantage encourages a further agglomeration. That is, the local demand pull is the only effective dispersion force as in the Krugman's model, and the demand from the immobile farmers eventually attract D-firms to disperse over space. For an intermediate transport cost, all D-firms agglomerate in one region, since neither dispersion force is strong enough to dominate the agglomeration force.

Not surprisingly, similar results can be obtained by using another prototype model by Fujita and Krugman which embodies the local demand pull by assuming a labor input in the land-intensive production of H-goods, and the factor price pull by introducing a positive transport cost for H-goods. However, the setting of this model is more general (in terms of interregional spatial structure) than other models in that a continuous location

[9] Tabuchi introduced the factor price pull in a way different from that of Helpman's. Namely, he introduced commuting costs while assuming the existence of infinite amount of land in each location. This formulation generates essentially the same dispersion force which is incorporated in the Helpman's model.

space is assumed, and that all workers are homogeneous and the division of labor between D- and H-sectors is endogenously determined. Consequently, more general results regarding the spatial distribution of agglomerations can be derived from this model.

First, it is to be noted that in the context of the Fujita-Krugman model, while the dispersion of D-firms always takes place given a sufficiently low transport cost for D-goods as in the case of Helpman's and Tabuchi's models, a higher transport cost for D-goods does not always disperse D-firms unlike the case of the models by Krugman and Tabuchi. The former result is obvious from (3-5), i.e., $\Omega'(0_+) > 0$ for $\tau_D < (1/\mu - 1)/(2 - 1/\sigma)\tau_H$. For the latter result, it can be shown that if the degree of product differentiation is sufficiently high, i.e., if σ is sufficiently small so that $\sigma < 1/(1-\mu)$, then the monopolar equilibrium is stable for all τ_D larger than $(1/\mu - 1)/(2 - 1/\sigma)\tau_H$ (at which $\Omega'(0_+) = 0$) (Fujita and Mori [14, Lemma 4.1(2)]). Thus, in this case, starting from a monopolar equilibrium, any increase in the transport cost for D-goods never disperses D-firms. In the Fujita-Krugman model, unlike Krugman's and Tabuchi's model, all workers are homogeneous and freely mobile, and the division of labor between D- and H-sectors is endogenous. That is, consumers can indefinitely substitute H-goods for highly differentiated D-goods by agglomerating to the city and producing more D-goods.

When the degree of product differentiation is not too high, i.e., σ is sufficiently large (i.e., $\sigma > 1/(1-\mu)$), an increase in transport cost for D-goods will eventually disperse D-firms as in the case of the Krugman's and Tabuchi's models. To see how the balance between agglomeration force and two types of dispersion forces changes in this case given different transport costs for D-goods, let us show in Figure 5.2 illustrative examples of potential curves (for $x \geq 0$) for a monopolar system under various values of transport rate, τ_D, for D-goods, while fixing the transport rate for H-goods at $\tau_H = 0.8$, where other parameter values are set as follows:

$$N = 3; \mu = 0.5; \sigma = 5; \alpha = \beta = 1; a = 0.5 \tag{5-1}$$

Thus, we have $\sigma = 5 > 1/(1-\mu) = 2$. In the figure, we can see that at $\tau_D = 1$, $\Omega(x) < 1$ for all $x \neq 0$, i.e., no D-firm has an incentive to move out the city at $x = 0$.[10]

As τ_D increases, however, while $\Omega'(0_+)$ decreases (i.e., the kink of the potential curve at the city becomes sharper), the potential value at a remote location in the agricultural area becomes larger. Notice that since the intensity of competition is very high in the small vicinity of the city, it is not profitable for a D-firm to move away from the city (giving up the proximity to the large market there). However, since the transport cost for D-goods is high, the intensity of competition becomes rapidly lower as a D-firm moves away from the city, which means that the market potential rapidly increases once the D-firm departs from the small vicinity of the city. In other words, *ceteris paribus the urban shadow of the city is smaller when the transport cost for D-goods is higher.* As indicated in the figure, the potential value reaches 1 at $x = 0.485$ at $\tau_D = 4.145$ and it exceeds 1 over an interval given a larger value of τ_D as in the case of $\tau_D = 5.0$. Thus for

[10] It can be shown that if $\tau_D > (1/\mu - 1)(2 - 1/\sigma)\tau_H$, the monopolar system is a stable equilibrium for a sufficiently small population size, N, if $\sigma > 1/(1-\mu)$, and it is a stable equilibnum for any $N > 0$ if $\sigma < 1/(1-\mu)$ (Fujita and Mori [14, Lemma 4.1]).

$\tau_D > 4.145$, D-firms will no longer continue agglomerating at $x = 0$, but move towards locations in the agricultural area offering a positive profit.

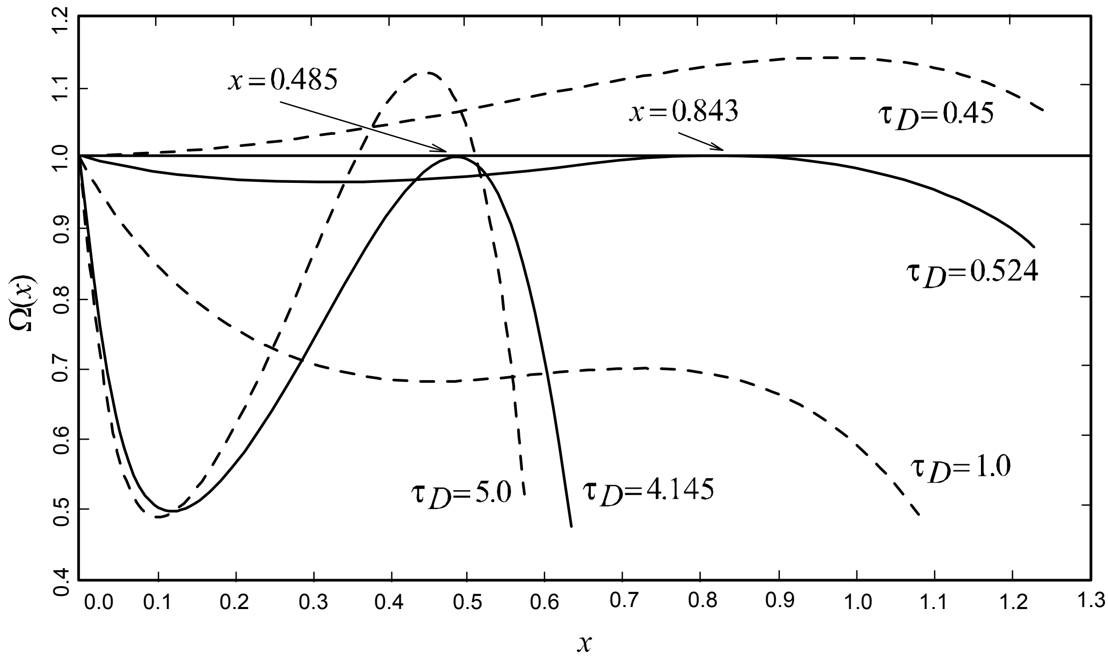

Figure 5.2. Potential curves for various transport rates for D-goods.

Next, as τ_D decreases, the slope of the potential curve at the city, $\Omega'(0_+)$, increases, and eventually reaches 1 at $\tau_D = (1/\mu - 1)(2 - 1/\sigma)\ \tau_H$. Thus, the monopolar system cannot be held given a further decrease in τ_D. Moreover, it can be shown that $\lim_{x\downarrow 0} d^2\Omega(x)/dx^2 > 0$ at $\tau_D = (1/\mu - 1)(2 - 1/\sigma)\ \tau_H$ if the degree of product differentiation is sufficiently low, i.e., σ is sufficiently large. Our example shown in the figure corresponds to this case. Namely, when transport rate τ_D for D-goods decreases to 0.524, the potential value reaches 1 at $x = 0.843$, and it exceeds 1 over an interval given a smaller value of τ_D as in the case of $\tau_D = 0.45$. Thus, D-firms will no longer continue agglomerating in the city at $x = 0$, but move towards a remote location in the agricultural area avoiding the high intensity of competition near competitors.

Mori [38] has demonstrated, in the context of a continuous location space assumed in the Fujita-Krugman model, that the dispersion of D-firms caused by the local demand pull and that caused by the factor price pull are qualitatively different. In the former case, as in the case of Krugman [29], the agglomeration of D-firms takes place at a smaller scale but at a larger number of discrete locations, and each small mass of D-firms serves primarily its nearby area. In our example, at $\tau_D = 4.145$, an arbitrarily small perturbation to the equilibrium population distribution triggers a bifurcation of the spatial system via the adjustment mechanism (4-1), and results in the formation of one or two new discrete masses of D-firms, i.e., new cities, at either or both of locations $x = \pm 0.485$. Figure 5.3(a) and (b) depict respectively the potential curve and wage curve when a new city emerges only at $x = 0.485$ after the bifurcation.[11] In diagram (a), while sharp kinks of the potential curve at the two cities at $x = 0$ and 0.485 indicate that the strong agglomeration force is at work near the cities, the potential value starts increasing rapidly in the agricultural area a little away from the cities (around $x = \pm 0.6$) due to the rapidly decreasing intensity of competition in the agricultural area. Thus, the shape of the potential curve implies that when the transport cost for D-goods is high, the hinterland of the city tends to be smaller. As shown in diagram (b), due to the high transport cost for D-goods, the cities have a labor cost advantage, and hence the effective dispersion force is only the local demand pull.

In the latter case, i.e., when the dominating dispersion force is the factor price pull, the dispersion takes place over an interval generating a so-called megalopolis which consists of two (discrete) core cities connected by an industrial belt, i.e., a continuum of cities. The megalopolis formation is triggered when the transport rate for D-goods decreased to $\tau_D = 0.524$. The potential curve which corresponds to $\tau_D = 0.524$ indicates, however, that location $x = 0.843$ (and -0.843) is the only possible location for D-firms except the city. Namely, the interval $(0, 0.843)$ is under the urban shadow of the existing city due to the high intensity of competition. Thus, an arbitrarily small perturbation to the population distribution at this moment can generate a new city or cities only at either or both of locations, $x = \pm 0.843$, via the adjustment mechanism (4-1). However, the transformation of the spatial system does not stop here. The reason for this is illustrated by Figure 5.4(a-1) and (a-2) which depict respectively the potential curve and wage curve of a bipolar economy with two discrete cities at $x = 0$ and 0.843. Diagram (a-1) indicates that the potential value exceeds 1 along interval $(0, 0.843)$, and hence D-firms are attracted to the interval. We can interpret this situation as follows. Once the new city emerged at $x = 0.843$, given a low transport cost for D-goods, locations along the interval $(0, 0.843)$ between the two cities become indifferent for D-firms and workers in terms of the competition intensity and market proximity to both cities. However, given the relatively high transport cost for H-goods, the interval can offer a relatively low price for H-goods. It follows that since workers attain the same utility level, the wage rate along the interval tends to be lower than that in the cities, which is in fact confirmed by diagram (a-2). This labor cost advantage, i.e., the factor price pull then attracts D-firms to migrate into the interval between the existing two cities.

Notice, however, that *the labor cost advantage continues to exist as long as exclusively agricultural area remains along the interval (0,0.843).* Hence, the interval between the oldest two cities at $x = 0$ and 0.843 eventually filled with D-firms and their workers generating an industrial belt. In equilibrium, then, while H-goods continue to be produced at each location along the industrial belt, there will be no transportation of H-goods along it. The excess supply of H-goods produced outside the megalopolis is then transported to and consumed in the cities at $x = 0$ and 0.843 (see Appendix for the conditions for a megalopolis equilibrium). The D-worker distribution and wage curve of the resulting megalopolis equilibrium are depicted in diagram (b-1) and (b-2) respectively. There is a discrete mass of D-workers of size 0.532 at each of the two core cities at $x = 0$ and

[11] Three equilibria, i.e., two bipolar system with cities at $x = 0$ and 0.485 or -0.485, and one tripolar system with cities at $x = 0$ and ± 0.485, are all realizable as a result of the bifurcation at $\tau_D = 4.145$.

0.843 at the edge of the megalopolis, while D-workers are distributed along the industrial belt connecting the two core cities with density ranging from 0.81 to 0.83. The D-worker density is higher towards the center of the industrial belt, reflecting the slightly better market proximity there. As we can see in diagram (b-2), the wage differential within the megalopolis (i.e., in interval $(0, 0.843)$) is mitigated, since all excess factor price advantage is exploited there.[12]

A further decrease in τ_D results in an expansion of the megalopolis. Diagram (c-1) indicates that the potential value at $x = 1.150$ (and $x = -0.307$) in the agricultural area reaches 1 at $\tau_D = 0.451$. Diagram (c-2) implies that the lower labor cost in the agricultural area is attracting D-firms. As a result, the megalopolis expands to either one of intervals $(-0.307, 0.843)$, $(0, 1.150)$, or $(-0.307, 1.150)$. It should be noted that since core cities always locate at the edge of a megalopolis, *the location of core cities changes as the megalopolis expands*. Diagram (d-1) and (d-2) present the resulting D-worker distribution and wage curve when the megalopolis expanded to the interval $(0, 1.150)$. In this case, the core city at $x = 0.843$ disappeared and a new core city emerged at $x = 1.150$. Diagram (d-1) indicates that the size of core cities decreased from 0.532 to 0.332, reflecting that *the agglomeration economies become less effective under the decreasing transport cost for D-goods.*[13] Diagram (d-2) shows that the wage differential is mitigated within the expanded megalopolis. In this way, a decrease in τ_D continues to expand the megalopolis, and eventually covers entire agricultural area.

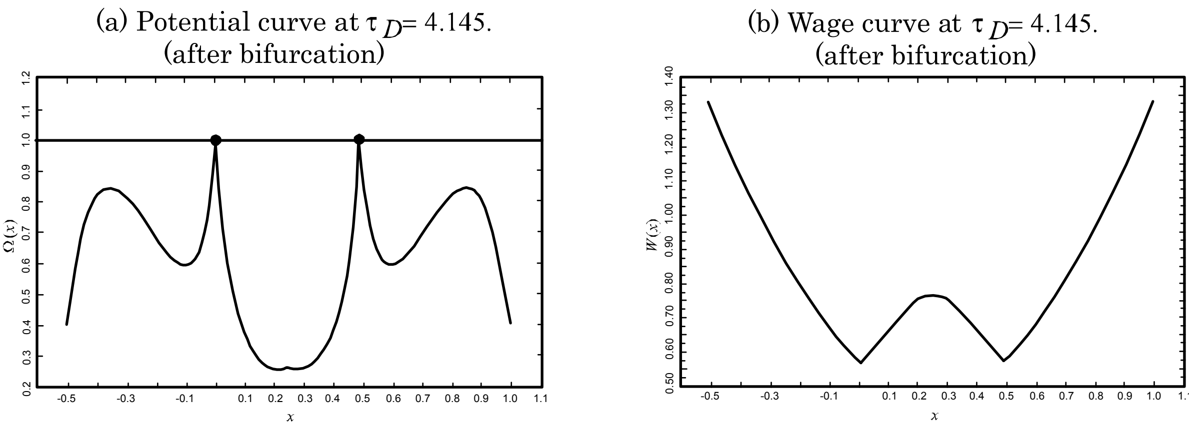

Figure 5.3. Formation of multiple discrete cities under a high transport cost for D-goods.

[12] The filling-in process of new cities between old cities which takes place in the megalopolis formation is in fact observed in the formation process of the megalopolis along the North-East cost of the US from Boston to Washington D.C. via New York and Philadelphia during the 19th century (Gottmann [18, Fig.51]). Also, the formation process of so-called *edge cities* emerging around the old metropolitan area of the US cities resembles the megalopolis formation, while the mechanism generating edge cities may differ from that makes a megalopolis in the context of our model. [See Garreau [17] and Urban Land Institute [48] for detailed descriptions of edge cities.]

[13] Note that a positive transport cost for D-goods is essential for generating agglomeration economies in the models we deal with.

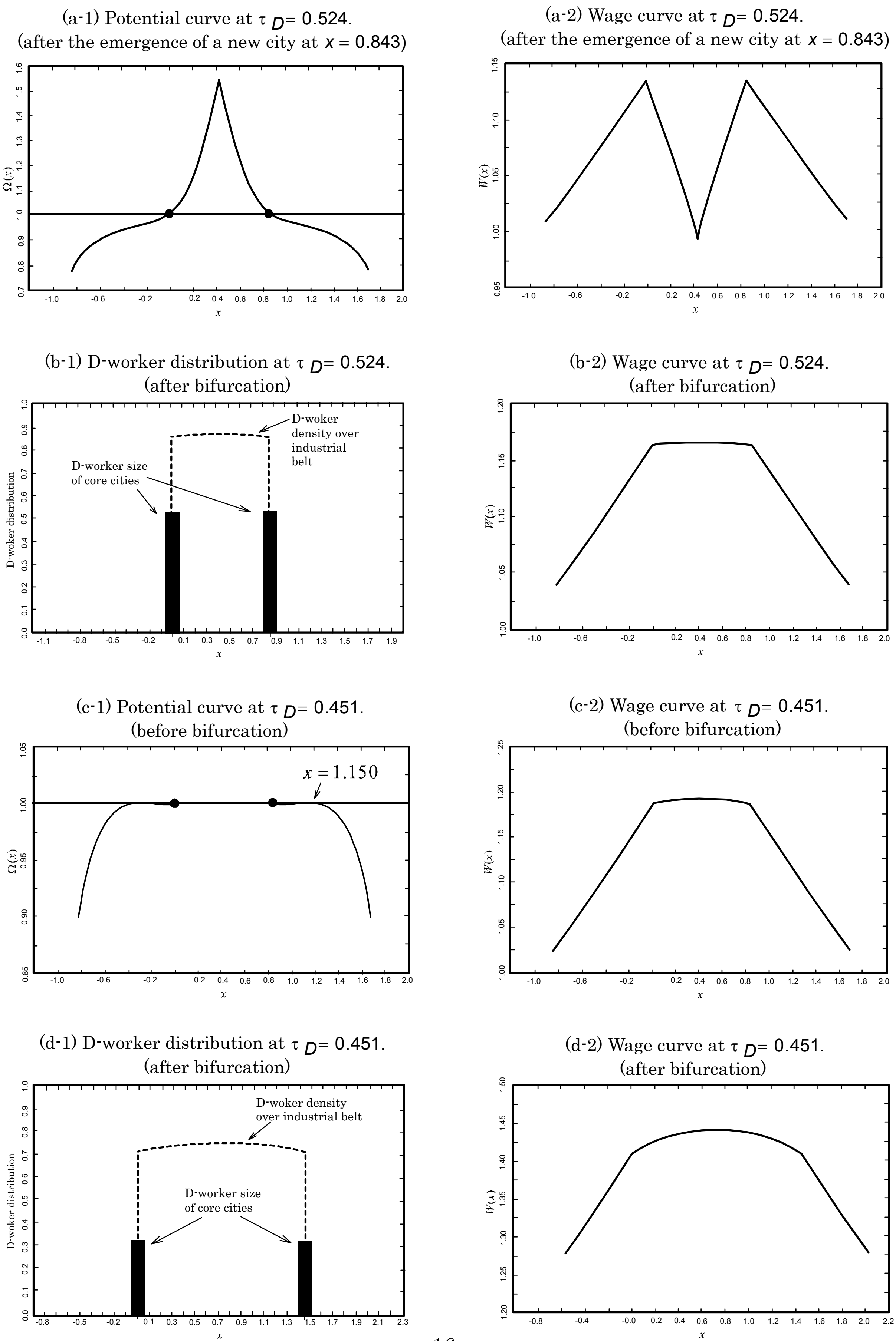

Figure 5.4. Megalopolitanization under the decreasing transport cost for D-goods.

So far, by reviewing and extending the results obtained by existing studies, we summarized the impact of changes in the transport cost for D-goods on the location pattern of D-industry. It has been shown that depending on the type of dispersion force, the impact of transport cost changes turned out to be quite different. Namely, if the local demand pull is the only dispersion force, then given that the degree of product differentiation is not too high, a higher [resp., lower] transport cost for D-goods tends to disperse D-firms in such a way that they concentrate at a smaller [resp., larger] scale at a larger [resp., smaller] number of discrete locations, and each mass of D-firms serves a smaller [resp., larger] hinterland around it. On the other hand, if the factor price pull is the only dispersion force present in the economy, D-firms tend to agglomerate [resp., disperse] into a smaller [resp., larger] number of locations given a higher [resp., lower] transport cost for D-goods. When both of the dispersion forces are involved, the effect of transport cost change is not monotonic as in the cases with either one of dispersion force. Namely, D-firms tend to disperse when the transport cost is either sufficiently high or low, while they tend to concentrate into a smaller number of locations for an intermediate transport cost. Moreover, in the context of a continuous location space, it has been shown that depending the dominating dispersion force, the patterns of spatial dispersion of D-firms are qualitatively different. Namely, if the transport cost for D-goods is high, i.e., the local demand pull is the dominating dispersion force, then D-firms tend to disperse into a larger number of discrete locations. On the other hand, if the transport cost for D-goods is low, the factor price pull becomes the dominating dispersion force, then D-firms tend to disperse over an interval leading to the formation of a megalopolis.

6 Impacts of a transport cost decrease: megalopolitanization or polarization

Unlike the previous section, here, we consider decreases in transport costs for both D- and H-goods, and demonstrate that depending on the relative rate of their decrease, the resulting spatial structure of the economy differs drastically. For this purpose, we use the framework of the Fujita-Krugman model, and examine the impact of transport cost changes focusing on two illustrative situations. Namely, in subsection 6.1, we present how an initially monopolar economy (i.e., D-firms are concentrated at one discrete point on the continuous location space) transforms when transport costs decrease, and in subsection 6.2, we extend our analysis for the case in which the economy is initially bipolar (i.e., D-firms are concentrated at two discrete points).

6.1 Monopolar economy and transport cost decrease: possible emergence of a megalopolis

Suppose a monopolar system is in equilibrium, where all D-firms are concentrated at $x = 0$ (see footnote 10 for the condition of a monopolar equilibrium). Recall that this situation can be realized when $(\tau_D, \tau_H) = (1.0, 0.8)$ under the set of parameter values given by (5-1). The potential curve for the monopolar system under this set of parameter values is shown in Figure 5.2. Since $\Omega(x) < 1$ for all $x \neq 0$, this equilibrium is indeed stable.

Figure 6.1 summarizes the result of our numerical simulation which experiments how initially a monopolar system at $(\tau_D, \tau_H) = (1.0, 0.8)$ transforms given gradually decreasing transport rates, τ_D and τ_H. The figure indicates the transformation of the spatial structure of the economy as (τ_D, τ_H) moves away from the right-top corner, i.e., $(1.0, 0.8)$. Here, we only focus on the case in which both transport rates monotonically decrease.

Curve OA, OB and OC represent the bifurcation points of the spatial system. Namely, if (τ_D, τ_H) is in the right of OA, then the monopolar system is in equilibrium, while if (τ_D, τ_H) crosses the curve from the right, then the monopolar system is no longer an equilibrium, and a megalopolis emerges. In particular, the bifurcation point A corresponds to the megalopolis formation at $(\tau_D, \tau_H) = (0.524, 0.8)$ which was explained in the previous section (Figure 5.4(a,b)). The process of the megalopolis formation is qualitatively the same along curve OA. Curve OB represents the bifurcation curve at which the expansion of the megalopolis takes place. The expansion which takes place along this curve is qualitatively the same as that happens at point B, i.e., $(\tau_D, \tau_H) = (0.451, 0.8)$, which has been explained in the previous section (Figure 5.4(c,d)). Similarly curve OC represents the bifurcation curve at which the next expansion of the megalopolis takes place. In this way, if the transport rate, τ_D, for D-goods decreases relatively faster, then the dispersion of D-industry proceeds in the form of megalopolitanization. Since the size of core cities of the megalopolis becomes smaller as τ_D becomes smaller (because the agglomeration economies become less relevant), the expansion of the megalopolis becomes smoother and finally continuous.[14] It is to be noted, however, for $\tau_H \leq 0.28$, the formation and expansion of megalopolis is always continuous, since τ_D reaches $(1/\mu - 1)(2 - 1/\sigma)\tau_H$, i.e., $\Omega'(0_+)$ becomes 0, before the potential value at any location $x \neq 0$ in the agricultural area reaches 1.

[14] Recall that in the previous section the size of core cities decreased from 0.532 to 0.332 as the megalopolis expands when τ_D decreased from 0.524 to 0.451, which implies the reduction in the agglomeration force at core cities.

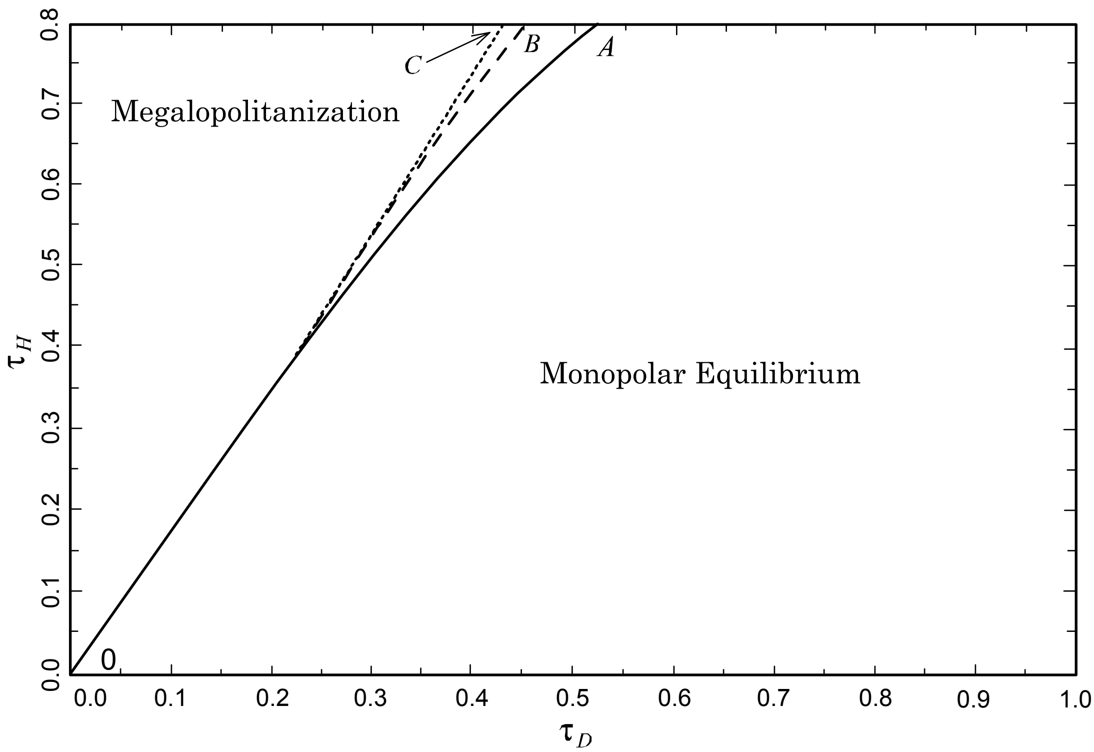

Figure 6.1. Impact of transport cost decrease on a monopolar economy: possible emergence of a megalopolis.

Next, let us consider the case in which the transport cost for H-goods decreases faster. Since τ_H decreases faster than τ_D, the wage rate in the agricultural area increases relatively to that in the city (refer to (3-1)), which in turn implies that the city tends to have a labor cost advantage over the agricultural area. In fact, we can see in (3-4) that $\Omega'(0_+)$ decreases as τ_H becomes relatively smaller, i.e., at least the market potential in the city relative to its nearby agricultural area increases. However, given a decrease in the relative transport cost for H-goods, the relative H-good price increases in the agricultural area. Then due to the income effect, the demand for D-goods at each location in the agricultural area tends to increase, and hence the local demand pull becomes stronger in the agricultural area. Thus, these two opposing effects leaves the resulting change in the market potential in the agricultural area ambiguous.

Our simulation indicates that if (τ_D, τ_H) does not reach OA, then no D-firms disperse from the city at $x = 0$, and hence the monopolar system is sustained. Namely, when the transport cost for H-goods decreases relatively faster, even though the transport cost for D-goods becomes relatively high, i.e., the local demand pull becomes relatively stronger, no dispersion of D-firms takes place unlike the case in which the transport cost for D-goods increases (Figure 5.2). A part of the reason for this result can be seen by examining the potential curve, (3-3), for a monopolar system. Namely, for $\tau_D > (1/\mu - 1)\,(2 - 1/\sigma)\,\tau_H$ (so that $\Omega'(0_+) < 0$), the monopolar system is a stable equilibrium, i.e., $\Omega'(x) < 1$ for all $x \neq 0$, if the population size, N, is sufficiently small (footnote 10). Since a small population size implies a small demand for H-goods, which in turn implies that a small agricultural area is required. However, the local demand pull can be effective only if the agricultural area is sufficiently large so that the intensity of competition sufficiently decreases within the agricultural area. It follows that the dispersion of D-firms cannot happen under a sufficiently small population size.

In summary, if the transport cost for D-goods decreases relatively faster, then eventually a megalopolis forms, and the megalopolis continues to expand if the transport cost for D-goods continues to decrease faster. On the other hand, if the transport cost for the H-good decreases relatively faster, then the agglomeration force remains to dominate the dispersion force, and the monopolar system continues to be sustained.[15]

6.2 Multipolar economy and transport decrease: megalopolitanization or polarization

If the transport cost for D-goods is relatively high so that $\tau_D > (1/\mu - 1)(2 - 1/\sigma)\tau_H$, and the degree of product differentiation is sufficiently small, i.e., $\sigma > 1/(1-\mu)$, then many-city system inevitably emerge as the population size of the economy becomes larger (Fujita and Mori [14, Sec.5.3,5.6]).[16] In the below, rather than attempting to explain all the possible transformation processes of the spatial system involving multiple cities, we focus on a simple example with two cities.

Suppose we have a monopolar equilibrium which is exactly the same as the initial spatial system assumed in the previous section. It has been shown by Fujita and Mori [14, Sec.5.3] that if the population size increases from $N = 3$ to 4.36, in order to meet the increased local demand for D-goods in the expanded agricultural hinterland, a new city or cities should emerge at either or both of locations, $x = \pm 1.10$, where the agricultural area extends to $l = 1.40$. For simplicity, let us assume that a new city emerged only at $x = 1.10$ when the population size reached $N = 4.36$. Thus, the monopolar system transforms to a bipolar system with cities at $x = 0$ and 1.10. Let us call the city at $x = 0$ [resp., $x = 1.10$] city 1 [resp., city 2]. Not surprisingly, the two cities in a new equilibrium have the same size, and the agricultural hinterland for these cities is divided at $x = 0.55$ ($= 0.5 \times 1.10$), in the left [resp., right] of which all excess H-goods are transported to city 1 [resp., city 2]. We take this bipolar system at $N = 4.36$ as the initial state of the economy's spatial structure, and investigate its transformation given a gradual decrease in transport costs.

[15] In interpreting the figure, it should be noted that due to the hysteresis or irreversibility of agglomeration processes, bifurcation curves, OA, OB, and OC, are valid only when (τ_D, τ_H) crosses them from the right. If the transport cost for the H-good starts decreasing faster than that for D-goods after (τ_D, τ_H) reached the left of OA, OB, and OC, then the megalopolis possibly contracts, or reduces to a system of discrete cities.

[16] Recall that by (3-5), $\tau_D > (1/\mu - 1)(2 - 1/\sigma)\tau_H \Leftrightarrow \Omega'(0_+) < 0$, and this condition together with $\sigma > 1/(1-\mu)$ implies that $\Omega(x) < 1$ for $x \neq 0$ for sufficiently small population size, N.

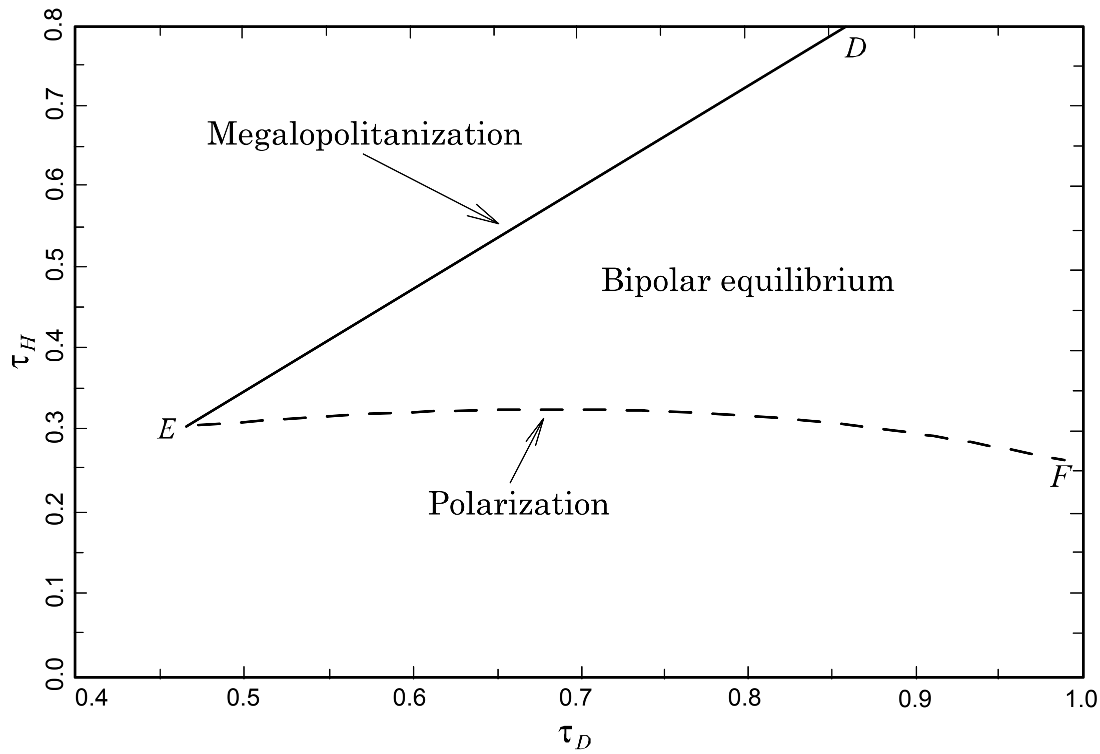

Figure 6.2. Impact of transport cost decrease on a bipolar economy: megalopolitanization or polarization.

Figure 6.2 summarizes the result. The initial situation is represented by the right-top corner, i.e., $(\tau_D, \tau_H) = (1.0, 0.8)$. First, suppose τ_D decreases relatively faster. Then, on one hand, D-firms and consumers become rather indifferent among locations along the interval between the two cities for their proximity to the market in these cities. On the other hand, given a relatively high transport cost for the H-good, the wage rate at each location tends to reflect the H-good price, i.e., the wage rate relatively decreases in the agricultural area. Hence, the agricultural area along the interval between the two cities attracts D-firms for its market proximity and labor cost advantage. When (τ_D, τ_H) reaches bifurcation curve DE, D-firms start moving toward the interval, and the process of the megalopolis formation sets in, following the same mechanism which generated a megalopolis from a monopolar system in the previous subsection (Figure 5.4(a,b)).

Next, suppose τ_H decreases relatively faster. As mentioned in the previous subsection, the impact on the balance between agglomeration force and dispersion force is ambiguous in this case. The simulation result indicates, however, that when (τ_D, τ_H) reaches bifurcation curve EF, one of the two cities collapse, and the bipolar system reduces to a monopolar system.[17] Once the spatial structure of the economy becomes monopolar, the impact of a further decrease in transport costs is qualitatively the same as the one shown by Figure 6.1. Hence, if τ_H continues to decrease faster, then the monopolar system continues to be sustained.

[17] Which of the two cities at $x = 0$ and 1.10 survives is arbitrary, since this bipolar system is symmetric.

In order to understand this polarization process, let us plot in Figure 6.3 the relative utility level, U_1/U_2, of city 1 given each D-worker share, $N_1/(N_1 + N_2)$, of city 1 for various value of τ_H, while fixing τ_D at 1.0, i.e., the change in the $N_1/(N_1 + N_2)$-U_1/U_2-curve along the (TD = 1.0)-locus in Figure 6.2. By investigating this curve, we can find the equilibrium population distributions as well as their stability in terms of the adjustment mechanism (4-1) when the location of cities are restricted at $x = 0$ and 1.10. Namely, $N_1/(N_1 + N_2)$ with $U_1/U_2 = 1$ corresponds to an equilibrium D-worker share of city 1, and the equilibrium is stable if the slope of the U_1/U_2-curve is negative.

At $\tau_H = 0.8$ (i.e., at the right-top corner of Figure 6.2), as shown in diagram (a), the symmetric bipolar system is the only stable equilibrium. When τ_H decreases to 0.52, however, as shown in diagram (b), four more asymmetric bipolar equilibria emerge: the two which are closer to the symmetric bipolar one are unstable, while the other two are stable. The emergence of these asymmetric equilibria indicates that a relative decrease in the transport cost for H-goods made the large concentration of D-firms at one of the cities possible due to the increase in labor cost advantage near a large variety of D-goods. At $\tau_H = 0.3$, as shown in diagram (c), the two stable asymmetric equilibria are absorbed into monopolar equilibria changing the stability of the monopolar equilibria. Thus, the monopolar system is now a stable equilibrium, and a further decrease in relative transport cost for H-goods magnifies the labor cost advantage generated by the concentration of a larger variety of D-goods at one location (recall the wage curve for the monopolar system (3-1) which increases towards the agricultural area when $\mu\tau_D > (1 - \mu)\tau_H$). At $\tau_H = 0.26$, eventually, the symmetric bipolar equilibrium loses its stability, and only monopolar equilibria remain to be stable.

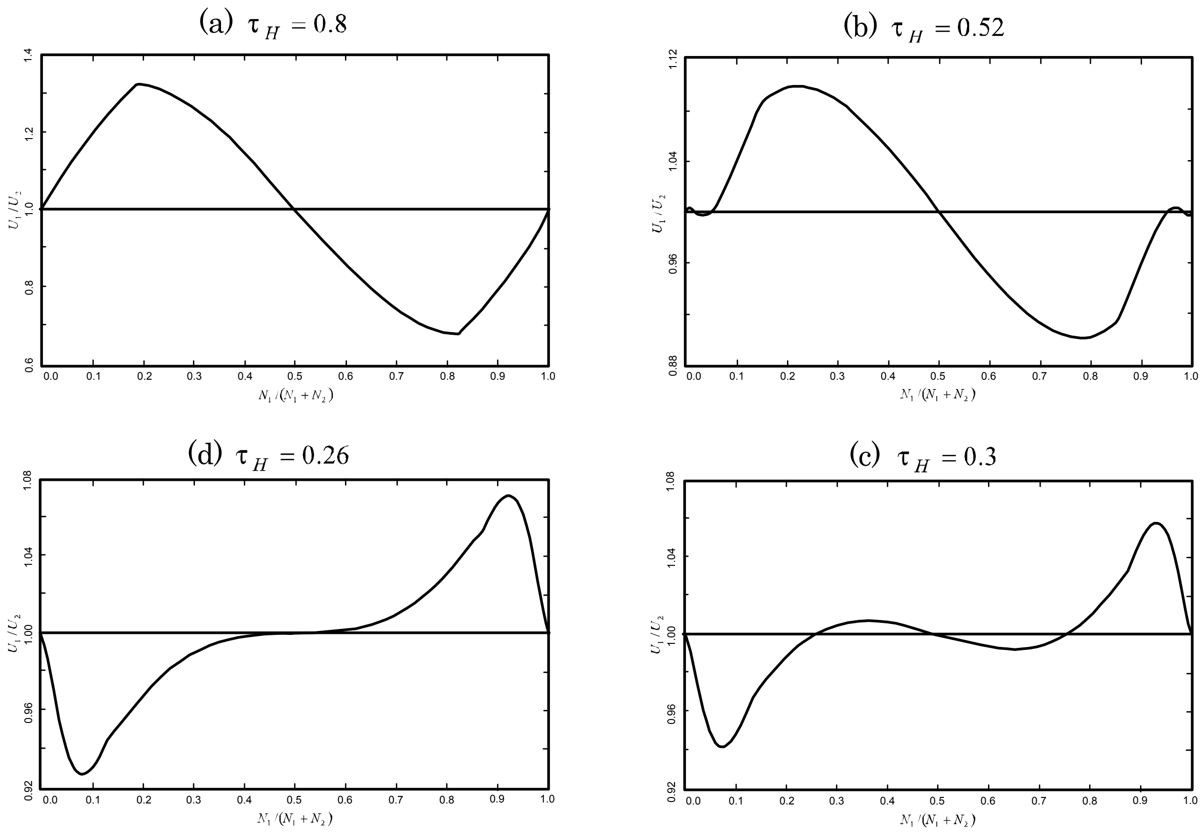

Figure 6.3. Transportation from a bipolar system to a monopolar system given a decrease in transport cost for H-goods.

Thus, when the transport cost for H-goods decreases relatively faster, although the increase in the relative H-good price in the agricultural area induces a larger demand for D-goods there, generating a local demand pull towards the agricultural area, this dispersion force tends to be dominated by the factor price pull towards the city generated by the labor cost advantage which stems from the proximity to a large variety of D-goods in the city. As a result, no dispersion of D-firms takes place when the transport cost for H-goods decreases faster.

7 Concluding remarks

In this paper we presented the impact of a transport cost decrease on the spatial pattern of economic agglomeration. In particular, by focusing on an economy in which the agglomerative industry consists of firms producing differentiated consumption goods, we have shown that locational differences in the market proximity, intensity of competition, and factor price play an important role in determining the agglomeration patterns. Our analysis, however, dealt with a rather limited situation. Thus, in applying our results to understand the organization of the economic geography in the real world the following remarks have to be kept in mind.

Remark 1. While our model involves only one type of differentiated consumption goods in terms of the degree of product differentiation (represented by σ) and transport cost (represented by τ_D), we observe several different varieties coexist in reality. For instance, varieties of restaurants, operas and musicals, boutiques, and movie theaters are all different from one another in both of the degree of product differentiation and transport cost. As a result, strengths of agglomeration economies which arise in these different industries are not the same. In particular, the industry which produces more highly differentiated products is subject to stronger agglomeration economies, and hence it is more likely to be polarized at a smaller number of locations. Thus, when multiple types of differentiated goods are introduced, the megalopolitanization and polarization induced by transport cost changes may be industry specific. [See Fujita, Krugman and Mori [11] for a preliminary attempt to investigate the spatial organization of the economy when multiple types of differentiated consumption goods are involved.]

Remark 2. While we generated the agglomeration force from product varieties of consumption goods, there are a number of other important sources of agglomeration economies. Examples are *product varieties of intermediate goods* (e.g., Fujita and Hamaguchi [9], Krugman and Venables [34], Venables [49]); *search and matching in local labor market* (Helsley and Strange [20]); *information spillover through face-to-face communications within an industry* (see Fujita and Smith [15] for a survey); *improvement in localized labor productivity through learning-by-doing* (e.g., Soubeyran and Thisse [45]).[18] The agglomeration economies which emerge from these different sources are affected differently by the improvement of transportation and communication technologies, which in turn implies that the impacts of transport development on the spatial organization of an economy are also different depending on which type of agglomeration economies is relevant for the economy in question.[19]

Remark 3. The two dispersion forces we considered in this paper are both market mediated forces such as the demand pull from spatially dispersed consumers, transport costs for land-intensive goods, land rent and commuting costs. However, non-market forces such as *congestion* and *pollution* are also important. There are some attempts to incorporate congestion in the context of the monopolistic competition model with varieties of consumption goods. In a two-point location space, Alonso-Villar [1] incorporates the traffic congestion by assuming that the cost of shipping in and out of a location increases as the population at the location increases. Brakman et. al [2] assume, in a discrete location space, that the productivity of labor decreases as a location becomes more crowded. Kofuji [25] assumes, in a two-point location space, that each variety of differentiated goods is produced by using labor and social overhead capital, where the latter input is distributed between the two locations and subject to congestion. Each study indicates that when the congestion is the effective dispersion force, the impact of a decrease in transport costs for differentiated goods is similar to the case in which the factor price pull is the effective dispersion force in our model. Namely, firms tend to disperse given a lower transport

[18] See Duranton and Puga [5] for a recent survey of micro foundations of agglomeration economies.

[19] See Tofflemire [47] for the treatment of telecommunication technologies and their impact on the urban structure.

cost for differentiated goods. However, we conjecture that the dispersion due to factor price pull and that due to congestion are qualitatively different. Namely, while the former has been shown to take the form of megalopolis formation, where two distinctively large masses of industrial agglomerations are connected by an industrial belt, the latter is likely to take the form of *urban sprawl* without generating any distinctive industrial agglomeration point.

Remark 4. In reality, the transportation of goods is subject to *transport density economies*. That is, the cost for shipping on a given link (or via a given hub) becomes smaller as the transport density via the link/hub increases. A study by Mori and Nishikimi [39] suggest that such *agglomeration economies in transportation* generates transport hubs in a location space, which in turn attract firms and consumers through their *hub effect*, and hence concentrations of industries and population tend to happen at transport hubs. When the transport development is considered in this context, its impact on the spatial organization of the economy may be quite different from the one we have obtained in this paper.

Remark 5. While the transport cost change assumed in our model takes place uniformly between any pair of locations, it is not likely to be so in reality. For instance improvements in railway, highway, and airway networks proceeds first by connecting well established economic centers. Such uneven development in transport networks may affect the transformation of the economic geography in different ways from the ones we obtained in this paper. [For discussions of related topics, see Fujita and Mori [13], Konishi [26], and Krugman [31]].

Remark 6. While we ignored the impact of international trade on the spatial structure of a domestic economy, Krugman and Livas Elizondo [33] has shown that the international trade may have a large impact on the industrial localization pattern and spatial population distribution within the domestic economy.

Remark 7. So far, no model is offered for the case in which industrial agglomeration endogenously takes place within an otherwise residential area, and the consumers' substitution among proximity to their job location, that to the market for consumption goods, and the size of land for housing is explicitly considered. However, especially in the developed countries, such a framework is more appropriate than that of the existing models including ours.

Remark 8. While each firm in our model is a single entity, in reality, it may consist of several units, each of which is subject to a different location behavior, and hence the impact of transport cost changes on their location patterns may differ.[20] [For a model with multi-unit firms, see Ota and Fujita [40] and Duranton and Puga [6].]

[20]Ross [43], for instance, has shown that the ranking of cities in terms of their sizes and that in terms of corporate control linkages are positively correlated in the case of the U.S. city system.

References

[1] Alonso-Villar, O., 1996, "Configuration of cities: the effects of congestion cost and government," Discussion paper 96-17(09), Departmento de Economia, Universidad Carlos III de Madrid.

[2] Brakman, S., Garretsen, R., van Marrewijk, C. and R. Vagenvoort, "Negative feedbacks in the economy and industrial location," *Journal of Regional Science* 36, 631-651.

[3] Cronon, W., 1991, *Nature's metropolis: Chicago and the Great West*, New York: Norton.

[4] Dixit, A. K. and J. E. Stiglitz, 1977, "Monopolistic competition and optimum product diversity," *American Economic Review* 67, 297-308.

[5] Duranton, G. and D. Puga, "Microfoundations of urban agglomeration economies," in J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics* vol.4, forthcoming.

[6] Duranton, G. and D. Puga, 2003, "From sectoral to functional specialization," Discussion paper No.2971, Center for Economic and Policy Research.

[7] Fujita, M., 1988, "A monopolistic competition model of spatial agglomeration: differentiated product approach," *Regional Science and Urban Economics* 18, 87-124.

[8] Fujita, M., 1990, "Spatial interactions and agglomeration in urban economics," in Chatterjii and R. E. Kunne (eds.), *New frontiers in regional science*, London: Macmillan, 185-222.

[9] Fujita, M. and N. Hamaguchi, 2001, "Intermediate goods and the spatial structure of an economy," *Regional Science and Urban Economics* 31, 79-109.

[10] Fujita, M. and P. Krugman, 1995, "When is the economy monocentric?: von Thünen and Chamberlin unified," *Regional Science and Urban Economics 25*, 505-528.

[11] Fujita, M., and Krugman, P. and T. Mori, 1999, "On the evolution of hierarchical urban systems," *European Economic Review* 43, 209-251.

[12] Fujita, M., Krugman, P. and A.J. Venables, 1999 *The Spatial Economy: Cities, Regions and International Trade*, Cambridge, Massachusetts: MIT Press.

[13] Fujita, M. and T. Mori, 1996, "The role of ports in the making of major cities: self-agglomeration and hub-effect," *Journal of Development Economics* 49, 93-120.

[14] Fujita, M. and T. Mori, 1997, "Structural stability and evolution of urban systems," *Regional Science and Urban Economics* 42, 399-442.

[15] Fujita, M. and T.E. Smith, 1991, "Additive-interaction models of spatial agglomeration," *Journal of Regional Science* 30, 51-74.

[16] Fujita, M. and J.-F. Thisse, 1996, "Economics of agglomeration," discussion paper No.1344, Center for Economic Policy Research, London.

[17] Garreau, J., 1991, *Edge city*, New York: Doubleday.

[18] Gottmann, J., 1961, *Megalopolis: the urbanized northern seaboard of the United States,* New York: The Twentieth Century.

[19] Helpman, E., 1998, "The size of regions," in D.Pines, E.Sadka and I.Zilcha (eds.), *Topics in Public Economics: Theoretical and Applied Analysis*, Cambridge: Cambridge University Press, 33-54.

[20] Helsley, R. and W. Strange, 1990, "Matching and agglomeration in a system of cities," *Regional Science and Urban Economics* 20, 189-212.

[21] Henderson, J. V., 1974, "The size and types of cities," *American Economic Review* 64, 640-656.

[22] Henderson, J. V., 1977, *Economic theory and the cities*, Oxford: Oxford University Press.

[23] Henderson, J. V., 1988, *Urban development*, Oxford: Oxford University Press.

[24] Isard, W., 1956, *Location and space economy*, Cambridge, MA: MIT Press.

[25] Kofuji, H., 1997, "Industrial Location and Traffic Infrastructures in Spatial Structures," *Keizaigaku Ronso* 49(2), 254-271 (in Japanese).

[26] Konishi, H., 2000, "Hub cities: city formation without economies of scale," *Journal of Urban Economics 48, 1-28.*

[27] Krugman, P., 1991a, "Increasing returns and economic geography," *Journal of Political Economy* 99, 483-499.

[28] Krugman, P., 1991b, *Geography and Trade*, Cambridge, MA: MIT Press.

[29] Krugman, P., 1993a, "On the number and location of cities," *European Economic Review* 37, 293-298.

[30] Krugman, P., 1993b, "First nature, second nature, and metropolitan location," *Journal of Regional Science* 33, 129-144.

[31] Krugman, P., 1993c, "The hub effects: or, threeness in interregional trade," in W. J. Ethier, E. Helpman and P. Neary (eds.), *Theory, Policy and Dynamics in International Trade. Essays in the Honor of R. W. Jones*, Cambridge University Press, 29-37.

[32] Krugman, P., 1996, "Urban concentration: the role of increasing returns and transport costs," *International Regional Science Review* 19, 5-30.

[33] Krugman, P. and R. Livas Elizondo, 1996, "Trade policy and the Third World metropolis," *Journal of Development Economics* 49, 137-150.

[34] Krugman, P. and A. Venables, 1996, "Integration, specialization and adjustment," *European Economic Review* 40, 959-968.

[35] Marshall, A., 1920, *Principles of economics*, London: Macmillan.

[36] Meyer, D. R., 1983, "Emergence of the American manufacturing belt: an interpretations," *Journal of Historical Geography* 9, 145-174.

[37] Mills, E. S., 1967, "An aggregate model of resource allocation in a metropolitan area," *American Economic Review* 57, 197-210.

[38] Mori, T., 1997, "A modeling of megalopolis formation: the maturing of city systems," *Journal of Urban Economics* 42, 133-157.

[39] Mori, T. and K. Nishikimi, 2002, "Economies of transport density and industrial agglomeration," *Regional Science and Urban Economics* 32, 167-200.

[40] Ota, M. and M. Fujita, 1993, "Communication technologies and spatial organization of multi-unit firms in metropolitan areas," *Regional Science and Urban Economics* 23, 695-729.

[41] Ottaviano, G., Tabuchi, T. and J.-F. Thisse, 2002, "Agglomeration and trade revisited," *International Economic Review* 43, 409-436.

[42] Papageorgiou, Y. Y. and J.-F. Thisse, 1985, "Agglomeration as spatial interdependence between firms and households," *Journal of Economic Theory* 37, 19-31.

[43] Ross, C., 1991, *The urban system and networks of corporate control*, Greenwich, CT: JAI press.

[44] Samuelson, P. A., 1952, "The transfer problem and transport costs: the terms of trade when impediments are absent," *Economic Journal* LXII, 278-304.

[45] Soubeyran, A. and J.-F. Thisse, 1996, "Learning-by-doing and the formation of Marshallian industrial districts," mimeo., CORE, University of Catholique de Louvain.

[46] Tabuchi, T., 1998, "Urban agglomeration and dispersion," *Journal of Urban Economics* 44, 333-351.

[47] Tofflemire, J. M., 1992, "Telecommunication external economies, city size and optimal pricing for telecommunications," *Journal of Regional Science* 32, 77-90.

[48] Urban Land Institute, 1994, "Suburban activity centers/edge cities," InfoPacket No.332, ULI Information Services, Urban Land Institute, Washington, D.C..

[49] Venables, A. J., 1996, "Equilibrium locations of vertically linked industries," *International Economic Review* 37, 217-235.

Appendix: Condition for a megalopolis equilibrium

Here we consider an economy with a megalopolis. Let the locations of two symmetric core cities be $x = \pm b$ (where $b > 0$), and denote the population size of these cities by N_b. At each location between these cities, both D-goods and H-goods are produced where the D-worker density is denoted by $N_D(y)$ for $y \in (-b, b)$. Along this interval, H-goods produced at each location are consumed at the same location, i.e., there is no transportation of H-good. Thus, all the excess H-goods on $(b, l]$ and that on $[-l, -b)$ are consumed in core cities at b and $-b$ respectively, where $\pm l$ represents the agricultural fringe distance. The H-good price for $x \in (b, l] \cup [-l, -b)$ is then given by $p_H(x) = \exp(-\tau_H |b - |x||)$, where $p_H(b)$ and $p_H(-b)$ are normalized to be 1. By using (2-4) and (2-6), the price index, $T_D(x)$, for the D-goods at location x can be obtained as follows:

$$\begin{aligned} T_D(x) &= (\alpha\sigma)^{1/(\sigma-1)} \frac{\beta}{1-1/\sigma} \\ &\times \left\{ \int_{-b}^{b} N_D(s) W(s)^{1-\sigma} e^{(1-\sigma)\tau_D|x-s|} ds + N_b W(b)^{1-\sigma} \left(e^{(1-\sigma)\tau_D|x-b|} + e^{(1-\sigma)\tau_D|x+b|} \right) \right\}^{1/(1-\sigma)} \end{aligned} \quad \text{(A-1)}$$

The conditions that the land rent at the agricultural fringe is zero, i.e., $R(\pm l) = 0$, and that all workers attain the same utility level assure that the wage rate at each location x must equal to

$$W(x) = (1/a) p_H(l)^{\mu} p_H(x)^{1-\mu} \left[T_D(x)/T_D(l) \right]^{\mu}. \quad \text{(A-2)}$$

In particular, $W(b) = (1/a) \exp(-\mu [\tau_H + \tau_D][l - b])$. Next, since all the H-goods consumed in the core city at b, $(1-\mu) N_b W(b)$, is produced in the agricultural hinterland on $(b, l]$, and all the excess H-good on this interval, $\mu \int_b^l \exp(-\tau_H [x - b]) dx$, is consumed in the core city at b, then it must hold that $(1-\mu) N_b W(b) = \mu \int_b^l \exp(-\tau_H [x - b]) dy$. Solving this equation for N_b, we obtain

$$N_b = \frac{\mu}{1-\mu} \frac{a}{\tau_H} \left\{ e^{\mu(\tau_H + \tau_D)(l-b)} - e^{(\mu\tau_D - [1-\mu]\tau_H)(l-b)} \right\}, \quad \text{(A-3)}$$

which is a function of only l. On the interval $(-b, b)$, the demand for H-goods at each location x, $(1-\mu) N_D(x) W(x)/p_H(x)$, should be met by the supply of H-goods at the same location x which in fact is a constant μ, and hence:

$$N_D(x) = \frac{\mu}{1-\mu} \frac{p_H(x)}{W(x)}. \quad \text{(A-4)}$$

Now, by using (2-1), (2-2), (2-4), (2-6) and (A-4), the total demand for the D-good produced by the firm at location x, $D(x)$, can be calculated as follows:

$$D(x) = \int_{-l}^{l} Z(x, y) dy, \quad \text{(A-5)}$$

where $Z(x, y)$ is the demand at location y for the D-good produced at location x, and is expressed by

$$\begin{aligned} Z(x,y) &= \frac{\alpha}{\beta} (\sigma - 1) \mu Y(y) W(x)^{-\sigma} e^{(1-\sigma)\tau_D|x-y|} \\ &\Big/ \left\{ \frac{\mu}{1-\mu} \int_{-b}^{b} W(s)^{1-\sigma} p_H(s) e^{(1-\sigma)\tau_D|s-y|} ds + N_b W(b)^{1-\sigma} \left(e^{(1-\sigma)\tau_D|y+b|} + e^{(1-\sigma)\tau_D|y+b|} \right) \right\}. \end{aligned} \quad \text{(A-6)}$$

Then, by substituting (A-5) into (3-2), the clearing condition for the D-good market can be obtained as

$$\Omega(x) = 1 \quad \text{for } x \in [-b, b] \tag{A-7}$$

Finally, since the total number of D-workers is the total population less the number of agricultural workers, i.e., $N - 2al$, then the demand and supply of labor meet if the following equation is satisfied.

$$2\left(N_b + \int_0^b N_D(x)dx\right) = N - 2al. \tag{A-8}$$

Now, the real unknowns are the agricultural fringe distance, l, the H-good price, $p_H(x)$, and the wage rate, $W(x)$, for each $x \in [0, b)$, which are determined by using (A-2), (A-7) and (A-8). Finally, in order for the equilibrium to be stable, it must hold that $\Omega(x) < 1$ for all $x \in (b, l] \cup [-l, -b)$.