DISCUSSION PAPER SERIES

No. 3386

ON THE NUMBER AND SIZE OF CITIES

Takatoshi Tabuchi, Jacques-François Thisse and Dao-Zhi Zeng

INTERNATIONAL TRADE

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP3386.asp

ISSN 0265-8003

ON THE NUMBER AND SIZE OF CITIES

Takatoshi Tabuchi, University of Tokyo
Jacques-François Thisse, CERAS, Paris and CORE, Université Catholique de Louvain and CEPR
Dao-Zhi Zeng, Kagawa University

Discussion Paper No. 3386
May 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3386

May 2002

ABSTRACT

On the Number and Size of Cities*

We study the effects of a decrease in trade costs on the spatial distribution of industry in a multi-regional economy, when a rise in the regional population of workers generates higher urban costs. When the number of cities is unaffected by falling trade costs, small cities become smaller for large trade costs, medium-sized cities become smaller for medium values of trade costs, and large cities become smaller for small trade costs. Furthermore, when urban costs are 'identical,' we show that there exists a path of stable equilibria such that the industry, first, experiences progressive agglomeration into a decreasing number of cities and, then, dispersion into a growing number of cities. The second phase arises because of the increasing urban costs associated with the process of agglomeration.

JEL Classification: F12, L13 and R13
Keywords: agglomeration, city, multiplicity of equilibria, multi-regional system, transport costs and urban costs

Takatoshi Tabuchi
Faculty of Economics
University of Tokyo
Hongo 7-3-1
Bunkyo-ku
Tokyo 113-0033
JAPAN
Tel: (81 3) 5841 5603
Fax: (81 3) 5841 5521
Email: ttabuchi@e.u-tokyo.ac.jp

Jacques-François Thisse
CORE
Université Catholique de Louvain
34 Voie du Roman Pays
B-1348 Louvain-la-Neuve
BELGIUM
Tel: (32 10) 474312
Fax: (32 10) 474301
Email: thisse@core.ucl.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=156606

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=104425

Dao-Zhi Zeng
Faculty of Economics
Kagawa University
1-1, Saiwai
Takamatsu
760-8521
JAPAN
Tel: (81 8) 7832 1905
Email: zeng@ec.kagawa-u.ac.jp

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=157596

* The first author is grateful to the Japan Society for the Promotion of Science through the Future Program, whereas the second author acknowledges support from the Ministère de l éducation, de la recherche et de la formation (Communauté française de Belgique), Convention 00/05-262. All authors thank Masa Fujita, Olivier Gossner and Tomoya Mori for helpful comments and discussions. This Paper is produced as part of a CEPR research network on 'The Economic Geography of Europe: Measurement, Testing and Policy Simulations', funded by the European Commission under the Research Training Network Programme (Contract No: HPRN-CT-2000-00069).

Submitted 15 April 2002

1 Introduction

There is a vast literature in economic history showing how the secular decline in communication and transport costs has fostered the making of cities (Bairoch, 1988). As all the impediments to the exchange of goods have decreased over a long enough period of time, one has witnessed a massive concentration of economic activities within a fairly small number of urban regions (Pollard, 1981). Yet, it has been argued recently that the continuous rise of congestion within such regions has triggered a process of counterurbanization (Berry, 1976). This paper addresses these issues by using a simple spatial general equilibrium model.

The primary purpose of this paper is to study the impact of the fall of trade costs on the spatial distribution of economic activities when the number of regions is arbitrary. Indeed, models of economic geography have so far focussed on a two-region setting (Krugman, 1991; Fujita, Krugman and Venables, 1999). This makes the dynamic analysis very simple since moving away from one region automatically implies that workers and firms go to the other region. Furthermore, it is not clear what the main result obtained in economic geography, namely the existence of a core-periphery structure, becomes when there are more than two regions. Indeed, a multi-regional economy is able to sustain a much richer hierarchy. To our knowledge, this is the first time that an analytical treatment of a multi-regional economy with mobile factors is addressed.

Our secondary purpose is to allow for urban costs to be paid by workers when residing in a particular region. Indeed, the core-periphery model has been criticized because it does not account for the growing urban costs associated with the concentration of firms and workers within the same region (Helpman, 1998; Tabuchi, 1998; Papageorgiou and Pines, 1999). By ignoring the costs imposed by urban life, this model would remain in the tradition of international trade theory, and would thus fail to provide a fair description of the working of a spatial economy. Introducing urban costs is both reasonable and meaningful. It is reasonable because an increasing concentration of workers and firms within a region generates rising congestion costs. It is meaningful because, in the absence of such costs, when trade costs decrease the economy might move from full dispersion to full agglomeration without passing through intermediate stages, a result that strikes us as being very implausible. Urban costs typically arise when each region has a central business district (CBD) whose existence is due to the agglomeration economies that induce the regional firms to cluster (Fujita and Thisse, 2002), as well as a land market that governs the residential allocation of workers around the CBD (Fujita, 1989). In other words, when firms and workers agglomerate within a region, they form a *city*.[1]

Specifically, we extend here the two-region model proposed by Ottaviano, Tabuchi and Thisse (2002) to study the impact of falling trade costs on the equilibrium distributions of firms and workers in the case of n regions, while permitting each region to have specific urban costs (e.g., commuting and housing), which vary with the number of workers. When the number of regions exceeds two, determining the equilibrium prices, wages, and (indirect) utilities in each region becomes a hard task. Indeed, these expressions typically depend on the *whole* distribution of the manufacturing sector across regions. In order to be able to work with a tractable model, we use a simple transportation geometry in which regions are pairwise equidistant. This means that the trade cost of a unit of good is the same regardless of the origin and destination regions. Such an assumption may be justified by the fact that distance-related transportation costs have become low enough while distance-unrelated costs such as tariffs, insurance, loading and unloading are still relatively high. Likewise, communication costs are not very sensitive to distance, but often involve high fixed costs (think of portable telephones). Of course, the value of trade costs varies with the volume of trade and, therefore, with the size of the population in the region of destination.

Regarding urban costs, our modeling strategy is as follows. Although we acknowledge the fact that both trade and commuting costs have been decreasing since the beginning of the Industrial Revolution, we assume that interregional transport costs decrease while urban commuting costs are constant for a given population size. This assumption is made to capture the idea that, in modern economies, trade costs of manufactured goods keep decreasing at a fast pace, while the decrease in commuting costs tends to slow down (and maybe to rise) due to growing congestion and higher opportunity time cost for urban residents.

Our concept of equilibrium is standard, while we borrow a dynamics that has been used in migration analysis (Ginsburgh, Papageorgiou and Thisse, 1985; Tabuchi, 1986; Zeng, 2002). This means two things. First, like in all economic geography models, we assume that markets for goods and labor adjust instantaneously (for a given distribution of firms and workers, it is readily verified that the market outcome is globally stable under a standard Marshallian dynamics). Besides simplifying the analysis, this assumption allows

[1]In this perspective, our model is somewhat comparable to Henderson (1974, 1988) in that we deal with the existence of an urban system within the interregional economy. However, here, cities are specialized in the production of differentiated varieties, which are sold under monopolistic competition and shipped between cities at a positive cost. In Henderson, cities supply different goods, which are sold under perfect competition and shipped at zero cost.

us to capture the idea that adjustments on markets are (much) faster than adjustments in migrations. Second, *the incentives to migrate away or toward a particular region are given by the sum of utility differentials between this region and the others*, the utility levels being evaluated at the equilibrium prices and wages corresponding to the prevailing distribution.

Even in such a context, proving the existence of a stable equilibrium when there are more than two regions may be a problematic issue. More generally, characterizing the eigenvalues of a nonnumerical system is often a formidable task. However, our model displays some nice features that allow us to apply recent stability theorems without having to compute eigenvalues. In particular, a stable equilibrium can be shown to exist for almost all values of the trade costs. To our knowledge, such a result has not been proven for the original core-periphery model developed by Krugman (1993).

Previewing our main results, we study in Section 4 how the size of cities changes as trade costs fall in the case where the number of cities is unaffected. More precisely, we show that, starting from large values, the decrease of trade costs triggers an agglomeration process in which large cities attract workers and firms from the small cities which shrink. By contrast, when trade costs are small, the large cities lose workers and firms while the small cities grow. Hence, agglomeration takes place in the early stages of the economic integration process, while re-dispersion should occur in the late stages. When trade costs take intermediate values, medium size cities lose workers and firms while large and small cities grow. In Section 5, we consider the case in which the number of cities changes as trade costs fall. This makes the analysis much more difficult because stable equilibria may vanish while new stable equilibria emerge for a marginal decrease in trade costs. In this context, we show that both an upper bound and a lower bound on the number of cities decrease when trade costs fall from high values to intermediate values; when trade costs take low values, this process is reversed. Furthermore, in the special, but meaningful, case of identical urban costs, the actual number of cities initially decreases and increases later. Hence, in the early stages of the integration process, the core of the economy is made of a shrinking number of cities. However, once trade costs are sufficiently low, the market solves the congestion problem induced by the agglomeration of industry by redistributing firms and workers among a larger number of cities. It should be clear that the implications of such results are important for the formation of integrating economies, such as the European Union or NAFTA.

Although we do not deal with differential regional growth, it seems fair to say that our paper contributes to the debate regarding the spatial implications of economic development. In the development literature, a high degree of urban concentration is expected to arise during the early phases of growth. As development proceeds, deconcentration would occur because the economy can afford to spread infrastructure, while the initial urban giants become high cost and congested places that are less attractive locations for producers and workers (Vining and Kontuly, 1978; Alonso, 1980). Since it is reasonable to interpret the value of internal trade costs as an index of economic development, we may conclude that our results suggest the existence of such a $\cap$-shaped relationship between economic development and the spatial distribution of activities. Interestingly, this relationship accords with the observations made in some developed economies, according to which industry would relocate outside the main cities (Champion, 1994; Geyer and Kontuly, 1996).

The remainder of the paper is organized as follows. The model is presented in Section 2, while existence and stability of an equilibrium are dealt with in Section 3. Sections 4 and 5 contain our main results. Section 6 concludes.[2]

2 The model

The space-economy is made of $n \geq 2$ regions ($i = 1, \ldots, n$). Each region has one city that has a given center but a variable size. As in urban economics, the city center stands for the CBD in which all firms locate once they have chosen to set up in the corresponding region. The CBDs are given by n points of the location space.

There are two factors, denoted A and L. Factor A is evenly distributed across regions (A/n) and is spatially immobile. The assumption of a uniform distribution of A is made in order to focus on the impact of differential urban costs on the distribution of activities. Factor L is mobile between any two regions. Let $\lambda_i \in [0, 1]$ denotes its share in region i and let

$$\Lambda \equiv \left\{ \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n); \quad \sum_{i=1}^{n} \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0 \right\}$$

For expositional purposes, we refer to the first sector as "agriculture" and to the second sector as "manufacturing". Accordingly, we call "farmers" the immobile factor A and

[2]Before proceeding, we would like to clarify how this paper differs from the models developed by Fujita and Krugman (1995) as well as by Fujita, Krugman and Mori (1999). Although these authors also use monopolistic competition (but of the Dixit-Stiglitz genre), they use a very different setting in which the industrial good is produced within cities with no spatial extension while the agricultural good is produced in the country by means of land and labor. They also have a different purpose because they aim at explaining the emergence and location of cities as the total population of mobile workers rises. Therefore, it is fair to say that the two approaches are quite different.

"workers" the mobile factor L.[3] For this reason, we will refer to regions accommodating workers ($\lambda_i > 0$) as *cities*, while regions with no workers ($\lambda_i = 0$) are called *rural regions*.

There are three goods in the economy. The first good is homogeneous and is produced in the agricultural sector using factor A as the only input under constant returns to scale and perfect competition. Technology in agriculture requires one unit of A in order to produce one unit of output. Consumers also have a positive initial endowment of this good. We assume that this good can be traded freely between regions so that its price is the same across regions. Hence this good is chosen as the numéraire. As a result, farmers' income is equal to one in all regions.[4]

The second good is a horizontally differentiated product; it is supplied by using L as the only input under increasing returns to scale and monopolistic competition. Technology in manufacturing is such that producing $q(i)$ units of variety i requires ϕ units of L and $vq(i)$ units of A. Given the demand structure described below, we may assume without loss of generality that the marginal cost of production of a variety is equal to zero ($v = 0$). Each firm in the manufacturing sector has a negligible impact on the market outcome in the sense that it can ignore its influence on, and hence reactions from, other firms. To this end, we assume that there is a continuum of potential firms. There are no scope economies so that, due to increasing returns to scale, there is a one-to-one relationship between firms and varieties. Clearly, regardless of the regional distribution of firms and the value of trade costs, the total number of firms in the whole economy is given by $N = L/\phi$. Although this might seem restrictive at first sight, this allows us to focus on the spatial redistribution of industry *per se* without accounting for the possible variations in industry size.

Because each firm sells a differentiated variety, it faces a downward sloping demand. Since there is a continuum of potential firms, each one is negligible and the interaction between any two firms is zero. However, as will be seen below, aggregate market conditions of some kind affects any single firm. This provides a setting in which individual firms are not competitive in the classic economic sense of having infinite demand elasticity but, at the same time, have no strategic interactions with one another. Finally, interregional trade flows go from one CBD to another. As discussed in the introduction, the corresponding

[3]We want to stress the fact, however, that the role of factor A is to capture the idea that some inputs (such as land or some services) are nontradable while some others have a very low spatial mobility (such as low-skilled workers). For example, the first sector could be reinterpreted as the traditional one and the second sector as the modern one.

[4]Recall that the choice of the numéraire is a difficult issue in general equilibrium model with imperfect competition.

trade costs per unit shipped are assumed to be identical between any two regions:

$$\tau_{ij} = \begin{cases} \tau > 0 & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}$$

Thus, each variety can be traded at a positive cost of τ units of the numéraire for each unit carried from one region to another, regardless of the variety, τ accounting for all the impediments to trade. The underlying geography is simple: the n regions are located along a circumference, while shipping a good from one region to another involves going through the center of the circumference.

Housing should be the third good in our economy. When they live in a city, workers use housing and commute to the regional CBD where they work. To keep things simple, *all the urban costs borne by a worker who chooses to reside in region* i (land rents, commuting and congestion costs, pollution) *are subsumed in a cost function* $\theta_i(\lambda_i)$, which varies with the size of the corresponding population of workers and enters the budget constraint (3) given below. This function is assumed to be twice continuously differentiable and to satisfy the following properties:

$$\theta_i(0) = 0 \qquad \theta_i(1) < \infty \qquad \theta_i'(\lambda_i) \geq 0 \qquad i = 1, \ldots, n \quad \text{and} \quad \lambda_i \in [0, 1] \tag{1}$$

Unlike trade costs that are the same between any pair of regions, urban costs are region-specific, reflecting the fact that living conditions may vastly differ across cities for the same population size due to differences in the amount of natural amenities, the quality of transport facilities, the supply of local public services, and/or the quantity of land available for housing (think of the shape of Chicago, Paris, or Perth which all have very different patterns). In doing so, we break the symmetry of the model in order to account for the heterogeneity of the set of natural sites that can be used for urban and industrial development.

Preferences over the first two goods are identical across individuals and described by a quasi-linear utility with a quadratic subutility, which is supposed to be symmetric in all varieties:

$$U(q_0; q(x), x \in [0, N]) = \alpha \int_0^N q(x)dx - \frac{\beta - \gamma}{2} \int_0^N [q(x)]^2 dx - \frac{\gamma}{2} \left[\int_0^N q(x)dx \right]^2 + q_0 \tag{2}$$

where $q(x)$ is the quantity of variety $x \in [0, N]$ and q_0 the quantity of the numéraire. The parameters in (2) are such that $\alpha > 0$ and $\beta > \gamma > 0$. In this expression, α expresses the intensity of preferences for the differentiated product, whereas $\beta > \gamma$ means that consumers are biased toward a dispersed consumption of varieties (*varietas delectat*).

If the consumption of the homogeneous good is positive, maximizing (2) under the budget constraint

$$\int_0^N p(x)q(x)dx + q_0 = w_i + \overline{q}_0 - \theta_i(\lambda_i) \tag{3}$$

(where w_i denotes the wage prevailing in region i and $\overline{q}_0$ is the initial endowment of the numéraire) yields the following first-order conditions:

$$\alpha - (\beta - \gamma)q(x) - \gamma \int_0^N q(y)dy = p(x) \qquad x \in [0, N]$$

or

$$q(x) = a - bp(x) + c\int_0^N [p(y) - p(x)]dy \qquad x \in [0, N] \tag{4}$$

where

$$a \equiv \frac{\alpha}{\beta + (N-1)\gamma} \qquad b \equiv \frac{1}{\beta + (N-1)\gamma} \qquad c \equiv \frac{\gamma}{(\beta - \gamma)[\beta + (N-1)\gamma]}$$

Substituting (3) and (4) into (2), we obtain the indirect utility of a worker residing in this region:

$$V_i = \frac{a^2 N}{2b} - a\int_0^N p(x)dx + \frac{b + cN}{2}\int_0^N [p(x)]^2 dx - \frac{c}{2}\left[\int_0^N p(x)dx\right]^2 + \overline{q}_0 + w_i - \theta_i(\lambda_i) \tag{5}$$

In accord with empirical evidence (Head and Mayer, 2000; McCallum, 1995), we assume that markets are regionally segmented so that each firm chooses a delivered price which is specific to the region in which its variety is sold. Let $p_{ij}(x)$ be the price of variety x produced in region i and sold in region j, and $q_{ij}(x)$ the demand in region j for variety x produced in region i. To ease the burden of notation, we drop x hereafter. Consequently, operating profits of a firm established in region i can be written as

$$\Pi_i(\boldsymbol{\lambda}) = \sum_{j=1}^{n}(p_{ij} - \delta_{ij}\tau)q_{ij}\left(\frac{A}{n} + \lambda_j L\right)$$

where $\delta_{ij} = 1$ when $i \neq j$ and 0 otherwise.

As to equilibrium wages, they are determined as follows. First, by maximizing firms' profits with respect to prices, we obtain[5]

$$
\begin{aligned}
p_{ii} &= \frac{2a + c\tau(1-\lambda_i)N}{2(2b+cN)} \\
p_{ji} &= p_{ii} + \frac{\tau}{2} \qquad \text{for } i \neq j \\
q_{ii} &= a - (b+cN)p_{ii} + cN\sum_{\ell=1}^{n} \lambda_\ell p_{\ell i} = (b+cN)p_{ii} \\
q_{ji} &= a - (b+cN)p_{ji} + cN\sum_{\ell=1}^{n} \lambda_\ell p_{\ell i} = (b+cN)(p_{ji}-\tau) \qquad \text{for } i \neq j
\end{aligned}
$$

Since the firms' prices net of trade costs are to be positive for any distribution of workers, we assume throughout this paper that

$$
\tau < \tau_{trade} \equiv \frac{2a\phi}{2b\phi + cL}
$$

This condition also guarantees that it is always profitable for a firm to export to any other region. Second, due to free entry and exit, profits net of fixed costs are zero in equilibrium. As in Krugman (1991), the equilibrium wages are determined by a bidding process between firms for workers, which ends when no firm can earn a strictly positive profit at the equilibrium market prices. In other words, all operating profits are absorbed by the wage bills. Hence, the wage prevailing in region i is determined as follows:

$$
\begin{aligned}
w_i(\boldsymbol{\lambda}) &= \frac{\Pi_i}{\phi} = \frac{1}{\phi}\sum_{j=1}^{n}(p_{ij} - \delta_{ij}\tau)q_{ij}\left(\frac{A}{n} + \lambda_j L\right) \\
&= \frac{(b+cN)N}{L}\sum_{j=1}^{n}\left(p_{jj} - \frac{\delta_{ij}\tau}{2}\right)^2\left(\frac{A}{n} + \lambda_j L\right) \\
&= \frac{(b+cN)N}{L}\left[\sum_{j=1}^{n}\left(p_{jj} - \frac{\tau}{2}\right)^2\left(\frac{A}{n} + \lambda_j L\right) + \left(p_{ii}\tau - \frac{\tau^2}{4}\right)\left(\frac{A}{n} + \lambda_i L\right)\right]
\end{aligned}
$$

[5] It is reasonable to assume that each firm's demand is decreasing in the total number of varieties because consumers spread their purchases over more varieties. Furthermore, it is also reasonable to assume that a consumer's demand for the differentiated product increases with N because more varieties makes this good more attractive compared to the numéraire. Computing the partial derivatives of the above demand function, we immediately see that $\partial q_{ii}/\partial N < 0$ and $\partial (q_{ii}N)/\partial N > 0$.

Accordingly, the indirect utility of a worker living in region i can be computed as follows:

$$
\begin{aligned}
V_i(\boldsymbol{\lambda}) =& \frac{a^2 N}{2b} - a\sum_{j=1}^{n}\lambda_j N p_{ji} + \frac{b+cN}{2}\sum_{\ell=1}^{n}\lambda_j N p_{ji}^2 - \frac{c}{2}\left(\sum_{j=1}^{n}\lambda_j N p_{ji}\right)^2 \\
&+ \overline{q}_0 + w_i - \theta_i(\lambda_i) \\
&= \frac{a^2 N}{2b} - aN\left[p_{ii} + \frac{\tau(1-\lambda_i)}{2}\right] + \frac{(b+cN)N}{2}\left[p_{ii}^2 + (1-\lambda_i)\tau\left(p_{ii} + \frac{\tau}{4}\right)\right] \\
&- \frac{cN^2}{2}\left[p_{ii} + \frac{\tau(1-\lambda_i)}{2}\right]^2 + \frac{(b+cN)N}{L}\left[\sum_{j=1}^{n}\left(p_{jj} - \frac{\tau}{2}\right)^2\left(\frac{A}{n} + \lambda_j L\right)\right. \\
&\left. + \tau\left(p_{ii} - \frac{\tau}{4}\right)\left(\frac{A}{n} + \lambda_i L\right)\right] + \overline{q}_0 - \theta_i(\lambda_i) \qquad (6)
\end{aligned}
$$

As expected, the indirect utility depends on the whole distribution $\boldsymbol{\lambda}$.

3 Existence and stability of a spatial equilibrium

We now move to the definition and stability of a spatial equilibrium. The distribution $\boldsymbol{\lambda}^* \in \Lambda$ is a *spatial equilibrium* when no worker may get a higher utility level by moving to another region. Formally, a distribution $\boldsymbol{\lambda}^*$ is an equilibrium if V^* exists such that

$$
\begin{aligned}
V_i(\boldsymbol{\lambda}^*) &= V^* \quad \text{if} \quad \lambda_i^* > 0 \\
V_i(\boldsymbol{\lambda}^*) &\leq V^* \quad \text{if} \quad \lambda_i^* = 0
\end{aligned}
\qquad (7)
$$

In words, this means that, in equilibrium, workers' utility in cities is (weakly) higher than in rural regions, while the utility level is constant across cities. Since $V_i(\boldsymbol{\lambda})$ is continuous in $\boldsymbol{\lambda} \in \Lambda$ as shown by (6), Proposition 1 of Ginsburgh *et al.* (1985) implies that *a spatial equilibrium always exists.*

In order to study the stability of a spatial equilibrium, we assume that local labor markets adjust instantaneously when some workers move from one region to the other. More precisely, wages are adjusted in each region for each firm located therein to earn zero profits. Hence, during the adjustment process, the utility level of a worker residing in region i is given by $V_i(\boldsymbol{\lambda})$.

The above spatial equilibrium conditions turn out to be equivalent to the following migration conditions:

$$
d\lambda_{ji}(t) = 0 \qquad \text{for all } j, i = 1, \ldots, n \qquad (8)
$$

where $d\lambda_{ji}(t)$ is the (net) migration from region j to region i during the infinitesimal time interval dt at time t. Following a now well-established tradition in migration modeling, we focus on an adjustment process in which *workers spread themselves among several cities, being attracted (resp. repulsed) by cities providing high (resp. low) utility levels.* In particular, we assume that migration $d\lambda_{ji}$ is proportional to the utility difference if the population of workers in region j is positive. Then, the dynamic system is such as

$$\frac{d\lambda_i}{dt} \equiv \sum_{j=1}^{n} \frac{d\lambda_{ji}}{dt} \qquad \text{for } i = 1, \ldots, n \tag{9}$$

where the speed of adjustment has been normalized to one, and where

$$\frac{d\lambda_{ji}}{dt} \equiv \begin{cases} V_i(\boldsymbol{\lambda}) - V_j(\boldsymbol{\lambda}) & \text{if } \lambda_i > 0, \lambda_j > 0 \\ \min\{0, V_i(\boldsymbol{\lambda}) - V_j(\boldsymbol{\lambda})\} & \text{if } \lambda_i > 0, \lambda_j = 0 \\ \max\{0, V_i(\boldsymbol{\lambda}) - V_j(\boldsymbol{\lambda})\} & \text{if } \lambda_i = 0, \lambda_j > 0 \\ 0 & \text{if } \lambda_i = 0, \lambda_j = 0 \end{cases}$$

It is readily verified that $\sum_{i=1}^{n} d\lambda_i/dt = 0$ since the total population of workers remains constant during the adjustment process. This dynamics can be justified by the assumption that migration decisions are made on the basis of pairwise comparisons between regions in that the net migration from region j to region i is proportional to their utility differential $V_i - V_j$. As a consequence, the sum of the net migration flows of region i is such that

$$\frac{d\lambda_i}{dt} = \sum_{j=1}^{n} [V_i(\boldsymbol{\lambda}) - V_j(\boldsymbol{\lambda})] = n \left[V_i(\boldsymbol{\lambda}) - \frac{1}{n} \sum_{j=1}^{n} V_j(\boldsymbol{\lambda}) \right] \tag{10}$$

if population in region $j \neq i$ is positive. Expression (10) means that regions with a utility level higher (resp. lower) than the average level across regions have a growing (resp. declining) population of workers and firms.[6]

In order to study the stability of a spatial equilibrium, we must evaluate the sum of the pairwise utility differentials used in (10). To this end, we set

$$S_i(\lambda_i) \equiv (C_1\tau - C_2\tau^2)\lambda_i - C_3\tau^2\lambda_i^2 - \theta_i(\lambda_i) \tag{11}$$

[6]Observe that (10) bears some resemblance the replicator dynamics used recently by Fujita *et al.* (1999). The two dynamics yield identical stationary states. However, the replicator is not suitable in depicting the appearance and disappearance of cities because the speed of adjustment gets very slow for sufficiently small λ_i.

where

$$C_1 \equiv \frac{aN(b+cN)(3b+2cN)}{(2b+cN)^2}$$
$$C_2 \equiv \frac{N(b+cN)}{8(2b+cN)^2}\left[4(2b+cN)\frac{cNA}{nL} + 12b^2 + 4bcN - 3c^2N^2\right]$$
$$C_3 \equiv \frac{cN^2(b+cN)(8b+5cN)}{8(2b+cN)^2}$$

Clearly, $S_i(0) = 0$. It is readily verified that $C_1 > 0$, $C_3 > 0$, $C_2 + C_3 > 0$.

Unlike $V_i(\boldsymbol{\lambda})$ that depends on the whole distribution $\boldsymbol{\lambda}$, the function $S_i(\lambda_i)$ depends only upon the size of city i. In addition, the following lemma will allow us to use $S_i(\lambda_i)$ instead of $V_i(\boldsymbol{\lambda})$ in the stability analysis of equilibria. The proof is given in Appendix A.

Lemma 1 *For $i = 1, \dots, n$, we have:*

$$\sum_{j=1}^{n}[V_i(\boldsymbol{\lambda}) - V_j(\boldsymbol{\lambda})] = \sum_{j=1}^{n}[S_i(\lambda_i) - S_j(\lambda_j)] \tag{12}$$

Hence, the RHS of (12) is additively separable with respect to the λ_i's, i.e., there are no crossed terms $\lambda_i\lambda_j$ with $i \neq j$. This lemma implies that

$$V_i(\boldsymbol{\lambda}) - \frac{1}{n}\sum_{j=1}^{n} V_j(\boldsymbol{\lambda}) = S_i(\lambda_i) - \frac{1}{n}\sum_{j=1}^{n} S_j(\lambda_j) \qquad i = 1, ..., n$$

For a given distribution $\boldsymbol{\lambda}$, this means that city i yields a welfare level higher (lower) than the average welfare if and only if $S_i(\lambda_i)$ is larger (smaller) than the average value of the $S_j(\lambda_j)$'s. Hence, the migration equation (10) becomes

$$\frac{d\lambda_i}{dt} = n\left[S_i(\lambda_i) - \frac{1}{n}\sum_{j=1}^{n} S_j(\lambda_j)\right]$$

thus making the stability analysis much simpler. From now on, we follow the tradition of urban economics and refer to $S_i(\lambda_i)$ as the "surplus" of region i.

If $\boldsymbol{\lambda}^*$ is a spatial equilibrium with $m \leq n$ cities i_j ($j = 1, \dots, m$), then $S^* \geq 0$ exists such that

$$S_{i_1}(\lambda^*_{i_1}) = \cdots = S_{i_m}(\lambda^*_{i_m}) = S^* \tag{13}$$
$$S_j(\lambda^*_j) = 0 \leq S^* \quad \text{for all } j \neq i_1, \dots, i_m \tag{14}$$

So, (7) is equivalent to (13) and (14). In spite of (14), there may exist an equilibrium at which all the surpluses are negative and equal. However, if the equilibrium involves at least one rural region, the surpluses of all cities are nonnegative and equal. In other words, there exists no spatial equilibrium such that $m < n$ and $S^* < 0$.

Consider an equilibrium with m cities such that

$$S'_{i_1}(\lambda^*_{i_1}) \leq \cdots \leq S'_{i_{m-1}}(\lambda^*_{i_{m-1}}) \leq S'_{i_m}(\lambda^*_{i_m})$$

Tabuchi and Zeng (2001) have shown that $\boldsymbol{\lambda}^*$ is (locally) *stable* if

$$S'_{i_{m-1}}(\lambda^*_{i_{m-1}}) < 0 \qquad \text{and} \qquad \sum_{j=1}^{m-1} \frac{S'_{i_m}(\lambda^*_{i_m})}{S'_{i_j}(\lambda^*_{i_j})} > -1 \tag{15}$$

$\boldsymbol{\lambda}^*$ is unstable if one of the inequalities in (15) is reversed.[7] We also know from Tabuchi and Zeng (2001) that a stable equilibrium generically exists for the dynamic system (10). When the manufacturing sector is concentrated into a single region ($m = 1$), the stability condition (15) ceases to hold and the equilibrium conditions (13) and (14) boil down to $S_1(1) \geq 0$.

4 The size of cities

In this section, we consider the evolution of a stable equilibrium such that the set of cities remains the same (so the number $m \leq n$ of cities does not change) whatever the values of τ and study how their *size* is affected by a marginal decrease in trade costs. Tabuchi and Zeng (2001) show that a stable equilibrium exists for almost all values of τ Without loss of generality, we may assume that cities are indexed such that:

$$\lambda^*_i \geq \lambda^*_{i+1} > 0 \qquad i = 1, \ldots, m-1 \tag{16}$$

As the urban cost curves $\theta_i(\lambda_i)$ may intersect, the order of cities by size may change as τ falls. When this is so, we re-index them for (16) to hold.

Let s be the city such that $S'_s(\lambda^*_s) = \max_i S'_i(\lambda^*_i)$. The stability condition (15) implies that $S'_i(\lambda^*_i) < 0$ for all $i \neq s$.

Definition 1 *The stable spatial equilibrium* $\boldsymbol{\lambda}^*$ *is said to be* regular *when* $S'_s(\lambda^*_i) \leq 0$*; otherwise, it is called* irregular.

Regular equilibria in which $S'_s < 0$ and $S'_s = 0$ are represented in Figures 1a and 1b, respectively. The case of an irregular equilibrium ($S'_s > 0$) is depicted in Figure 1c. In what follows, both cases are discussed in order.

[7] When either one of the two inequalities in (15) becomes an equality, the equilibrium may be stable or unstable. We will reurn to this case in Section 5.

Figure 1: Regular and irregular equilibria

Trade costs being given by τ, we denote the corresponding interior equilibrium by

$$\boldsymbol{\lambda}^*(\tau) = (\lambda_1^*(\tau), \dots, \lambda_m^*(\tau))$$

with $\lambda_i^*(\tau) > 0$. Since $S_i(\lambda_i)$ is also a function of τ, we denote it as $S_i(\lambda_i, \tau)$. For convenience, we also set

$$S_i' \equiv \left.\frac{\partial S_i(\lambda_i, \tau)}{\partial \lambda_i}\right|_{\lambda_i=\lambda_i^*}$$

$$z_i \equiv \sum_{j=1}^m \frac{\partial(S_i - S_j)}{\partial \tau} = (C_1 - 2C_2\tau)(m\lambda_i^* - 1) - 2C_3\tau\left[m(\lambda_i^*)^2 - \sum_{j=1}^m (\lambda_j^*)^2\right] \tag{17}$$

The evolutionary process is indeterminate when the equilibrium is fully symmetric ($\lambda_i^* = \lambda_j^*$ holds for all $i, j = 1, \dots, m$). Therefore, we restrict our attention to asymmetric equilibrium in which $\lambda_i^* \neq \lambda_j^*$ holds for some i, j. Such an asymmetry ensures that "large" and "small" cities exist. We also assume that

$$\lambda_i^* \neq \lambda_s^* \qquad \text{for all } i \neq s \quad \text{when } S_s' = 0 \tag{18}$$

As $S_i' < 0$ holds for all $i \neq s$ when $S_s' = 0$, (18) does not entail much loss of generality.

In equilibrium, it must be that $\sum_{j=1}^m [S_i(\lambda_i^*(\tau), \tau) - S_j(\lambda_j^*(\tau), \tau)] = 0$. Differentiating this equation yields the following system of linear equations whose unknowns are $d\lambda_i^*(\tau)/d\tau$:

$$-(m-1)S_i' \frac{d\lambda_i^*(\tau)}{d\tau} + \sum_{\substack{j=1 \\ j \neq i}}^m S_j' \frac{d\lambda_j^*(\tau)}{d\tau} = z_i \qquad i = 1, \dots, m \tag{19}$$

Let

$$D \equiv \begin{pmatrix} -(m-1)S_1' - S_m' & S_2' - S_m' & \cdots & S_{m-1}' - S_m' \\ S_1' - S_m' & -(m-1)S_2' - S_m' & \cdots & S_{m-1}' - S_m' \\ \vdots & \vdots & \ddots & \vdots \\ S_1' - S_m' & S_2' - S_m' & \cdots & -(m-1)S_{m-1}' - S_m' \end{pmatrix}$$

and

$$E \equiv \begin{cases} \dfrac{1}{m^3|D|} \displaystyle\prod_{\ell=1}^{m}(-mS'_\ell) \sum_{j=1}^{m} 1/\left(-S'_j\right) > 0 & \text{for } S'_s \neq 0 \\ \dfrac{1}{m^2|D|} \displaystyle\prod_{\substack{\ell = 1 \\ \ell \neq s}}^{m} (-mS'_\ell) & \text{for } S'_s = 0 \end{cases}$$

Since $|D| > 0$ from Lemma B in Appendix B, $E > 0$ must hold at any stable equilibrium. Also define the weighted average of z_j as follows:

$$\bar{z} \equiv \begin{cases} \dfrac{\sum_{j=1}^{m} z_j/\left(-S'_j\right)}{\sum_{j=1}^{m} 1/\left(-S'_j\right)} & \text{for } S'_s \neq 0 \\ z_s & \text{for } S'_s = 0 \end{cases}$$

We now establish the following result which will be useful when studying the evolution of the industry distribution.

Lemma 2 *Assume that the number of cities does not change as τ falls. If $S'_i \neq 0$, then*

$$\frac{d\lambda_i^*(\tau)}{d\tau} = E\,\frac{z_i - \bar{z}}{-S'_i} \qquad \textit{for } i = 1, \ldots, m \tag{20}$$

If $S'_s = 0$, then

$$\frac{d\lambda_i^*(\tau)}{d\tau} = \begin{cases} E\,\dfrac{z_i - \bar{z}}{-S'_i} & \textit{for } i \neq s \\ E \displaystyle\sum_{\substack{j = 1 \\ j \neq s}}^{m} \dfrac{\bar{z} - z_j}{-S'_i} & \textit{for } i = s \end{cases} \tag{21}$$

The proof is given in Appendix C.

It is readily verified that the difference between z_i and z_{i+1} is given by

$$z_i - z_{i+1} = m(\lambda_i^* - \lambda_{i+1}^*)\{C_1 - 2[C_2 + C_3(\lambda_i^* + \lambda_{i+1}^*)]\tau\} \qquad \text{for } i = 1, \ldots, m-1 \tag{22}$$

We are interested in the sign of $z_i - z_{i+1}$. To this end, set

$$\tau_1^* \equiv C_1/2\,(C_2 + C_3) \qquad \text{and} \qquad \tau_2^* \equiv C_1/2C_2$$

where $\tau_2^* > \tau_1^* > 0$ (resp. $\tau_1^* > 0 > \tau_2^*$) if $C_2 > 0$ (resp. $C_2 < 0$); $\tau_2^* > \tau_1^*$ is likely to hold when varieties are sufficiently differentiated or when the number of farmers is not too low.

In this case, if trade costs are so high that $\tau > \tau_2^*$, then the RHS of (22) is nonpositive since $\lambda_i^* \geq \lambda_{i+1}^*$ from (16). Thus, a decrease in τ makes larger cities larger, while smaller cities become smaller. By contrast, if trade costs are low ($\tau < \tau_1^*$), the opposite holds. This argument is developed in a more systematic way in the following result.

Lemma 3 *For any τ, there exists a city k such that*

$$z_i \geq z_{i-1} \quad \text{for } i < k$$
$$z_i \geq z_{i+1} \quad \text{for } i \geq k$$

The proof is given in Appendix D.

Lemma 3 shows that $z_k = \max_i z_i$ and that z_i is "single-peaked" in i. If $k = 1$ (m), then z_i is decreasing (increasing) in i. If $k \neq 1$ and m, then z_i is increasing in i for $i < k$ and decreasing for $i \geq k$. From now on, we say that city size "increases" when it strictly increases for decreasing τ, and "decrease" when it decreases or does not change for decreasing τ.

4.1 Regular equilibrium ($S_s' \leq 0$)

Since the population of workers and firms is fixed, when some cities become larger due to a decrease in τ, some others must become smaller. Among the cities that experience a decreasing population, let k_- be the largest city (the city with the smallest number if there are several of them) and k_+ be the smallest one (the one with the largest number if there are several of them). Both k_- and k_+ exist, otherwise there would be no city having a decrease in population; if there is only one city whose population decreases, then $k_- = k_+$.

At a regular equilibrium, some city k must become smaller and, hence, $k_- \leq k \leq k_+$ must hold. Indeed, if $S_k' < 0$, then $d\lambda_i^*(\tau)/d\tau > 0$ from (20) because of $z_k = \max_i z_i \geq \bar{z}$, and hence city k becomes smaller. If $S_k' = 0$, then it follows from the first equation of (21) that all the other cities ($i \neq 1, \ldots, k-1, k+1, \ldots, m$) do not become smaller. Therefore, city k must become smaller since $d\lambda_i^*(\tau)/d\tau = 0$. More precisely, the following result is shown in Appendix E.

Lemma 4 *Assume that the number of cities does not change as τ falls and $\tau_2^* \in (0, \tau_{trade})$. Then we have:*

(i) when $\tau \geq \tau_2^*$, the large cities $1, \ldots, k_- - 1$ become larger, whereas the small cities $k_-, \ldots, m$ become smaller;

(ii) when $\tau_1^* < \tau < \tau_2^*$, the large cities $1, \ldots, k_- - 1$ and the small cities $k_+ + 1, \ldots, m$ become larger, whereas the medium size cities $k_-, \ldots, k_+$ become smaller.

(iii) when $\tau \leq \tau_1^*$, the large cities $1, \ldots, k_+$ become smaller whereas the small cities $k_+ + 1, \ldots, m$ become larger.

In this lemma, the large, medium size and small cities have not been identified because the cities k_- and k_+ may change when τ falls. In any case, city 1 is the largest region and city m is the smallest one. Accordingly, the large region experiences a population increase in the first stage of the integration process whereas the small region experiences a population increase in the second stage, a result which agrees with the 2-region case studied by Ottaviano *et al.* (2002). In the case of three regions, suppose that the large and medium size cities first grow while the small city decreases. Then, (i) the medium size city is the one first that switches from growth to decline; then (ii) the small region switches from decline to growth; finally (iii) the large region switches from growth to decline. When the large city grows while the small and medium size cities first decline, case (ii) arises first and then (iii) occurs.

4.2 Irregular equilibrium ($S_s' > 0$)

By comparison with the regular case, the sign of (20) in Lemma 2 is reversed only in region s since the corresponding denominator ($-S_s'$) is negative. However, the sign of the other denominators ($-S_i'$ for $i \neq s$) is the same as before. Therefore, we obtain the same result as in Lemma 4, except for region s (see Appendix F for a proof).

Lemma 5 *Assume that the number of cities does not change as τ falls and $\tau_2^* \in (0, \tau_{trade})$. At any irregular and interior equilibrium with $S_s' > 0$, Lemma 4 holds for all cities $i \neq s$. When s is the smallest region ($s = m$), the lemma may be not hold as $d\lambda_s^*(\tau)/d\tau > 0$ only in city s for $\tau_1^* < \tau < \tau_2^*$.*

In the very special case covered by Lemma 5, the large cities $1, \ldots, k_- - 1$ and the small cities $k_+ + 1, \ldots, m - 1$ become larger, while the medium size cities $k_-, \ldots, k_+$ together with the smallest city m become smaller.

4.3 Dispersion/agglomeration/re-dispersion

Putting together Lemmas 4 and 5, we obtain our main result.

Theorem 1 *Assume that the set of cities does not change as τ falls and $\tau_2^* \in (0, \tau_{trade})$. If $S_s' \leq 0$, then large cities become larger and small cities become smaller as long as $\tau \geq \tau_2^*$; large and small cities become larger and medium size cities become smaller when*

$\tau_1^* < \tau < \tau_2^*$; and large cities become smaller and small cities become larger once $\tau \leq \tau_1^*$. Furthermore, when $S_s' > 0$, the direction of migration is reverse for at most one city.

This result is illustrated in Figure 2. Starting from large values of τ, we see that large (resp. small) cities become larger (resp. smaller) as τ decreases. In the interval $[\tau_1^*, \tau_2^*]$ of intermediate values of τ, Figure 2 shows that net migration changes sign, thus sizes change accordingly, as trade costs fall from small to large regions. For low trade costs, large (resp. small) cities lose (resp. gain) workers and firms.

Figure 2: Evolution of city size distributions

Our theorem has several important implications. First, when trade costs are high ($\tau \geq \tau_2^*$), their decrease triggers an agglomeration process in which each large region attracts workers and firms from the small regions which shrink. By contrast, when trade costs are small ($\tau \leq \tau_1^*$), the large cities lose workers and firms while the small cities grow. Hence, agglomeration takes place in the early stages of economic integration, while re-dispersion should occur in the late stages of the economic integration process. Second, when trade costs take intermediate values ($\tau_1^* < \tau < \tau_2^*$), medium size cities lose workers and firms while large and small cities grow. To sum up: *small cities shrink in the early stages, medium size cities shrink in the next stages, and large cities shrink in the late stages of the process of economic integration.*

Finally, when there are few farmers, or when the varieties are close substitutes, or both, we have $\tau_2^* \notin [0, \tau_{trade})$. Then, the early stages of agglomeration do not arise. Furthermore, when the number of farmers is very low, or when the degree of differentiation between varieties is very low, we have $\tau_1^* \geq \tau_{trade}$. Then, the early and the intermediate stages do not arise. In this case, the integration process triggers a dispersion of the industrial sector from the large cities to the small ones, through the medium size cities, thus generalizing Helpman's (1998) numerical results.

5 The number of cities

So far, we have assumed that the set and number of cities was not affected by a decrease in trade costs. This may be justified when τ varies within a small interval. However, it is important to figure out how the number of cities is affected when the decrease in trade costs is large. In order to achieve this goal, we impose a minor restriction on the urban cost functions: we assume that $S_i(\lambda_i)$ is *strictly quasi-concave* in the population size λ_i. This is satisfied when the urban cost functions $\theta_i(\lambda_i)$ are convex ($S_i(\lambda_i)$ is then strictly concave), a condition which holds under fairly general assumptions in urban economics (Fujita, 1989, p.145).

Despite its simplicity, our model exhibits a high degree of complexity in that a stable equilibrium may become unstable or may even disappear, whereas a new stable equilibrium may emerge, for a marginal decrease in τ. In order to illustrate the nature of the difficulty, consider the following example. There are three regions ($n = 3$) and the surplus function is identical across regions. Assume that the initial equilibrium involves the 3 cities and is such that $(1/3, 1/3, 1/3)$, as shown in Figure 3(a). This equilibrium becomes unstable for a marginal decrease in τ. When the surplus function at $\lambda = 1/6$ is flatter than at $\lambda = 5/6$, then the equilibrium $(5/6, 1/6, 0)$ with 2 cities is a stable and is shown in Figure 3(b); note that other stable equilibria might exist. This equilibrium, in turn, becomes unstable for a further decrease in τ. Indeed, there are now three stable equilibria given by $(2/5, 2/5, 1/5)$, $(1/2, 1/2, 0)$ and $(1, 0, 0)$, which are all displayed in Figure 3(c). Hence, there exist several stable equilibria, some of them having a larger number of cities, while others involve a smaller number. In particular, falling trade costs may lead to dispersion ($m = 3$) or to full agglomeration ($m = 1$). Thus, monotonicity in the number of cities cannot be shown.

In the multiregional case, unlike the 2-region case, the multiplicity of stable equilibria does no longer allow one to tell nice stories about the evolution of the spatial economy. Rather, it becomes a nightmare that prevents us from deriving any general result. In this respect, the numerical results presented by Krugman (1993) as well as by Brakman *et al.* (2001, ch. 4) are somewhat misleading in that they do not make it clear how the transition from full dispersion to partial agglomeration occurs.

Figure 3: Multiple equilibria in the number of cities

In order to gain more insights about the possible evolution of the urban system, we consider the case of identical urban costs: $\theta_i(\lambda) = \theta(\lambda)$ for all i with $\theta''(\lambda) \geq 0$ and $\theta'''(\lambda) \leq 0$ - the latter condition means that urban costs are not "too convex". The convexity of $\theta(\lambda)$ implies that $S(\lambda)$ is strictly concave (and not strictly quasi-concave as in Figure 3). In this case, our next lemma shows that irregular equilibria never arise (the proof is contained in Appendix G).

Lemma 6 *If the urban costs are identical, convex and $\theta'''(\lambda) \leq 0$, then there is no irregular equilibrium.*

Given Lemma 6, any stable equilibrium is regular. Consequently, we can fully characterize the number and size of cities in terms of τ. Indeed, it can readily be shown that stable equilibria are always symmetric $\lambda_i^* = 1/m$ $(m \leq n)$, and hence the spatial equilibrium condition becomes $S_i(1/m) \geq 0$ whereas the stability condition is $S_i'(1/m) < 0$. Therefore, assuming that a stable equilibrium describes the actual evolution of the economy until it becomes unstable, we have the following result (the proof is contained in Appendix H).

Proposition 1 *Assume that the urban costs are identical, convex and $\theta'''(\lambda) \leq 0$. Then, there exists a threshold value τ^* such that*
(i) the number of cities decreases as trade costs decrease for $\tau \geq \tau^*$;
(ii) the number of cities increases as trade costs decrease for $\tau < \tau^*$.[8]

This shows that *urban concentration first arises while re-dispersion comes afterwards.* In other words, as trade costs decrease, rural-urban flows first arise and are then followed by urban-rural flows. As discussed in the introduction, the first phase is historically well documented while there are on-going debates about whether or not the second phase has already started.

In addition, for sufficiently large or sufficiently small trade costs, dispersion is likely to arise, whereas the highest degree of agglomeration within the economy may involve several cities. In other words, *the presence of urban costs may prevent the full agglomeration into a single core region.* On the other hand, if the urban costs in region 1 are sufficiently low, then the industry is fully agglomerated into a single region for intermediate values of the trade costs. Consequently, when trade costs decrease while urban costs do not, *the economy would move from dispersion to the emergence of an urban giant and, then, would display gradual deconcentration.*

Note that τ^* may exceed τ_{trade}, and hence case (i) in the proposition may not arise. For example, when there are very few farmers, the mere existence of urban costs is sufficient

[8] To illustrate, the above assumptions are satisfied by $\theta(\lambda) = \sum_j a_j \lambda^{b_j}$ $(a_j \geq 0,\ 1 \leq b_j \leq 2)$, which includes the cases of linear and quadratic urban costs.

for the decrease in trade costs to induce right away a gradual dispersion of the industrial sector over a growing number of regions, a result which extends Helpman (1998) to an arbitrary number of regions.

Finally, even in the special case of identical urban costs, the number of cities does not necessarily decrease (resp. increase) by one when trade costs fall. Indeed, when $m > 2$, there always exist several stable equilibria for any value of τ. Insofar as there is no reason to choose a particular equilibrium, the evolutionary processes simulated by Krugman (1993) and Brakman *et al.* (2001) might well be the outcomes of some arbitrary selections implicitly used when several regions get more integrated.

6 Concluding remarks

This paper suggests that the secular fall in transport and communication costs should lead to a possibly strong concentration of mobile activities, which will eventually be followed by a re-dispersion of these activities. In other words, the general pattern of activities as trade costs fall would be more or less $\cap$-shaped. However, much work remains to be done in order to understand how regions evolve when trade costs take intermediate values, while it is also important to figure out how the medium regions react to decreasing trade costs.

Our model has dismissed the fact that commuting costs have decreased together with trade costs. So, it would be interesting to study the impact of their relative change on the spatial structure of industry.[9] Finally, it should be kept in mind that our model considers a given and fixed set of economic activities. In particular, the number of firms is the same regardless of the value of trade costs. In this respect, the observed decline of the industrial sector within big cities does not necessarily imply the economic and social decline of these areas. The continuous decrease in communication and transport costs gives rise to new economic activities that are typically information-oriented, and which, therefore, tend to grow in large metropolises. Thus, one task for future research is to investigate this question in a setting allowing firms and workers to locate "out in the burds".

[9] See Tabuchi and Thisse (2002) for a first attempt along these lines.

APPENDIX

A. Proof of Lemma 1

$$\sum_{j=1}^{n}[V_i(\boldsymbol{\lambda})-V_j(\boldsymbol{\lambda})] = -aN\left[np_{ii}-\sum_{j=1}^{n}p_{jj}-\frac{\tau}{2}\left(n\lambda_i-\sum_{j=1}^{n}\lambda_j\right)\right]$$
$$+\frac{(b+cN)N}{2}\left[np_{ii}^2-\sum_{j=1}^{n}p_{jj}^2+\tau n(1-\lambda_i)\left(p_{ii}+\frac{\tau}{4}\right)-\tau\sum_{j=1}^{n}(1-\lambda_j)\left(p_{jj}+\frac{\tau}{4}\right)\right]$$
$$-\frac{cN^2}{2}\left\{\sum_{j=1}^{n}\left[\left(p_{ii}+\frac{\tau(1-\lambda_i)}{2}\right)^2-\left(p_{jj}+\frac{\tau(1-\lambda_j)}{2}\right)^2\right]\right\}$$
$$+(b+cN)N\tau\left[n\left(p_{ii}-\frac{\tau}{4}\right)\lambda_i-\sum_{j=1}^{n}\left(p_{jj}-\frac{\tau}{4}\right)\lambda_j\right]+\frac{(b+cN)NA\tau}{nL}\left(np_{ii}-\sum_{j=1}^{n}p_{jj}\right)-n\theta_i(\lambda_i)+\sum_{j=1}^{n}\theta_j(\lambda_j).$$

However,

$$np_{ii}-\sum_{j=1}^{n}p_{jj}=\frac{cN\tau}{2(2b+cN)}(1-n\lambda_i),$$

$$np_{ii}^2-\sum_{j=1}^{n}p_{jj}^2=\frac{(2a+Nc\tau)Nc\tau}{2(2b+cN)^2}(1-n\lambda_i)+\frac{c^2N^2\tau^2}{4(2b+c)^2}\left(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2\right),$$

$$\tau n(1-\lambda_i)\left(p_{ii}+\frac{\tau}{4}\right)-\tau\sum_{j=1}^{n}(1-\lambda_j)\left(p_{jj}+\frac{\tau}{4}\right)=\tau\left[(np_{ii}-\sum_{j=1}^{n}p_{jj})-\left(n\lambda_ip_{ii}-\sum_{j=1}^{n}\lambda_jp_{jj}\right)-\frac{\tau}{4}(n\lambda_i-1)\right]$$
$$=\tau\left(\frac{\tau}{4}+\frac{a+cN\tau}{2b+cN}\right)(1-n\lambda_i)+\frac{cN\tau^2(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2)}{2(2b+cN)},$$

$$\sum_{j=1}^{n}\left\{\left[p_{ii}+\frac{\tau(1-\lambda_i)}{2}\right]^2-\left[p_{jj}+\frac{\tau(1-\lambda_j)}{2}\right]^2\right\}=\sum_{j=1}^{n}\left[p_{ii}+p_{jj}+\tau-\frac{\tau}{2}(\lambda_i+\lambda_j)\right]\left(p_{ii}-p_{jj}+\frac{\lambda_j-\lambda_i}{2}\tau\right)$$
$$=\frac{b+cN}{2b+cN}\tau\left[\frac{2(a+b\tau+N\tau c)}{2b+cN}(1-n\lambda_i)+\frac{b+cN}{2b+cN}\tau\left(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2\right)\right],$$

$$n\left(p_{ii}-\frac{\tau}{4}\right)\lambda_i-\sum_{j=1}^{n}\left(p_{jj}-\frac{\tau}{4}\right)\lambda_j=-\frac{(2a+cN\tau)(1-n\lambda_i)-cN\tau(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2)}{2(2b+cN)}+\frac{\tau}{4}(1-n\lambda_i)$$
$$=\left[\frac{\tau}{4}-\frac{2a+cN\tau}{2(2b+cN)}\right](1-n\lambda_i)-\frac{cN\tau}{2(2b+cN)}\left(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2\right).$$

Then,

$$
\begin{aligned}
\sum_{j=1}^{n}[V_i(\boldsymbol{\lambda})-V_j(\boldsymbol{\lambda})] = & -\frac{aN\tau(b+cN)}{2b+cN}(1-n\lambda_i)+\frac{(b+cN)N}{2}\left\{\left[\frac{Nc\tau(2a+Nc\tau)}{2(2b+cN)^2}+\tau\left(\frac{\tau}{4}+\frac{a+cN\tau}{2b+cN}\right)\right](1-n\lambda_i)\right.\\
& \left.+\left[\frac{cN\tau^2}{2(2b+cN)}+\frac{c^2N^2\tau^2}{4(2b+c)^2}\right]\left(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2\right)\right\}\\
& -\frac{cN^2\tau(b+cN)}{2(2b+cN)}\left[\frac{2(a+b\tau+N\tau c)}{2b+cN}(1-n\lambda_i)+\frac{b+cN}{2b+cN}\tau(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2)\right]\\
& +(b+cN)N\tau\left\{\left[\frac{\tau}{4}-\frac{2a+cN\tau}{2(2b+cN)}\right](1-n\lambda_i)-\frac{cN\tau}{2(2b+cN)}(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2)\right\}\\
& +\frac{(b+cN)NA\tau}{nL}\frac{cN\tau}{2(2b+cN)}(1-n\lambda_i)-n\theta_i(\lambda_i)+\sum_{j=1}^{n}\theta_j(\lambda_j)\\
= & (1-n\lambda_i)\left\{-\frac{aN\tau(b+cN)}{2b+cN}+\frac{(b+cN)N^2c\tau(2a+Nc\tau)}{4(2b+cN)^2}+\frac{(b+cN)N\tau}{2}\left(\frac{\tau}{4}+\frac{a+cN\tau}{2b+cN}\right)\right.\\
& \left.-\frac{cN^2\tau(b+cN)(a+b\tau+N\tau c)}{(2b+cN)^2}+(b+cN)N\tau\left[\frac{\tau}{4}-\frac{2a+cN\tau}{2(2b+cN)}\right]+\frac{c(b+cN)N^2A\tau^2}{2nL(2b+cN)}\right\}\\
& +(n\lambda_i^2-\sum_{j=1}^{n}\lambda_j^2)\left\{\frac{(b+cN)N}{2}\left[\frac{cN\tau^2}{2(2b+cN)}+\frac{c^2N^2\tau^2}{4(2b+cN)^2}\right]\right.\\
& \left.-\frac{cN^2\tau^2(b+cN)^2}{2(2b+cN)^2}-\frac{(b+cN)N^2\tau^2c}{2(2b+cN)}\right\}-n\theta_i(\lambda_i)+\sum_{j=1}^{n}\theta_j(\lambda_j)\\
= & n(-C_1\tau+C_2\tau^2)\left(\frac{1}{n}-\lambda_i\right)+C_3\tau^2\left(\sum_{j=1}^{n}\lambda_j^2-n\lambda_i^2\right)-n\theta_i(\lambda_i)+\sum_{j=1}^{n}\theta_j(\lambda_j)
\end{aligned}
$$

which is identical to (12).

B. Lemma B

Let D_{ij} be the submatrix of D obtained by deleting the i-th row and j-th column. We then have:

Lemma B. *If (15) holds at the equilibrium* $\boldsymbol{\lambda}^*$*, then*

$$|D|>0$$

$$|D_{ii}|=\frac{1}{m}\left(1+\sum_{\substack{j=1\\ j\neq i}}^{m}\frac{S'_m}{S'_j}\right)\prod_{\substack{\ell=1\\ \ell\neq i}}^{m-1}(-mS'_\ell)\qquad i=1,\ldots,m-1 \tag{23}$$

$$|D_{ij}|=(-1)^{i-j+1}(S'_i-S'_m)\prod_{\substack{\ell=1\\ \ell\neq i,\ \ell\neq j}}^{m-1}(-mS'_\ell)\quad i,j=1,\ldots,m-1,\ i\neq j \tag{24}$$

Proof.(i) Using some basic properties of determinants, we obtain

$$
\begin{aligned}
|D| &= \begin{vmatrix} -(m-1)S_1' - S_m' & S_2' - S_m' & S_3' - S_m' & \cdots & S_{m-1}' - S_m' \\ mS_1' & -mS_2' & 0 & \cdots & 0 \\ mS_1' & 0 & -mS_3' & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ mS_1' & 0 & 0 & \cdots & mS_{m-1}' \end{vmatrix} \\
&= \left[-(m-1)S_1' - S_m' + \sum_{\ell=2}^{m-1} \frac{S_1'(S_\ell' - S_m')}{S_\ell'} \right] \prod_{\ell=2}^{m-1} (-mS_\ell') \\
&= \frac{1}{m} \left(1 + \sum_{\ell=1}^{m-1} \frac{S_m'}{S_\ell'} \right) \prod_{\ell=1}^{m-1} (-mS_\ell') > 0,
\end{aligned}
$$

where the inequality follows from the stability condition (15).

(ii) We first consider the case of $i > 1$. By definition of D_{ii} and some properties of determinants, we have

$$
\begin{aligned}
|D_{ii}| &= \begin{vmatrix} -(m-1)S_1' - S_m' & \cdots & S_{i-1}' - S_m' & S_{i+1}' - S_m' & \cdots & S_{m-1}' - S_m' \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_1' - S_m' & \cdots & -(m-1)S_{i-1}' - S_m' & S_{i+1}' - S_m' & \cdots & S_{m-1}' - S_m' \\ S_1' - S_m' & \cdots & S_{i-1}' - S_m' & -(m-1)S_{i+1}' - S_m' & \cdots & S_{m-1}' - S_m' \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_1' - S_m' & \cdots & S_{i-1}' - S_m' & S_{i+1}' - S_m' & \cdots & -(m-1)S_{m-1}' - S_m' \end{vmatrix} \\
&= \begin{vmatrix} -(m-1)S_1' - S_m' & S_2' - S_m' & \cdots & S_{i-1}' - S_m' & S_{i+1}' - S_m' & \cdots & S_{m-1}' - S_m' \\ mS_1' & -mS_2' & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ mS_1' & 0 & \cdots & -mS_{i-1}' & 0 & \cdots & 0 \\ mS_1' & 0 & \cdots & 0 & -mS_{i+1}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ mS_1' & 0 & \cdots & 0 & 0 & \cdots & -mS_{m-1}' \end{vmatrix} \\
&= \Bigg[-(m-1)S_1' - S_m' + \sum_{\substack{j=2 \\ j \neq i}}^{m-1} \frac{S_1'(S_j' - S_m')}{S_j'} \Bigg] \prod_{\substack{\ell=2 \\ \ell \neq i}}^{m-1} (-mS_\ell') \\
&= \frac{1}{m} \Bigg(1 + \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{S_m'}{S_j'} \Bigg) \prod_{\substack{\ell=1 \\ \ell \neq i}}^{m-1} (-mS_\ell').
\end{aligned}
$$

Next, for $i = 1$, we have

$$
\begin{aligned}
|D_{11}| &= \begin{vmatrix} -(m-1)S'_2 - S'_m & S'_3 - S'_m & S'_4 - S'_m & \cdots & S'_{m-1} - S'_m \\ mS'_2 & -mS'_3 & 0 & \cdots & 0 \\ mS'_2 & 0 & -mS'_4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ mS'_2 & 0 & 0 & \cdots & -mS'_{m-1} \end{vmatrix} \\
&= \left[-(m-1)S'_2 - S'_m + \sum_{j=3}^{m-1} \frac{S'_j - S'_m}{S'_j} S'_2 \right] \prod_{\ell=3}^{m-1} (-mS'_\ell) \\
&= \frac{1}{m} \left(1 + \sum_{j=2}^{m} \frac{S'_m}{S'_j} \right) \prod_{\ell=2}^{m-1} (-mS'_\ell)
\end{aligned}
$$

(iii) We consider only the case where $j < i$. By straightforward calculation, we know

$$
\begin{aligned}
&|D_{ij}| \\
&= \begin{vmatrix}
-(m-1)S'_1 - S'_m & S'_2 - S'_m & \cdots & S'_{j-1} - S'_m & S'_{j+1} - S'_m & \cdots & S'_{i-1} - S'_m & S'_i - S'_m & S'_{i+1} - S'_m & \cdots & S'_{m-1} - S'_m \\
mS'_1 & -mS'_2 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
mS'_1 & 0 & \cdots & -mS'_{j-1} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
mS'_1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
mS'_1 & 0 & \cdots & 0 & -mS'_{j+1} & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
mS'_1 & 0 & \cdots & 0 & 0 & \cdots & -mS'_{i-1} & 0 & 0 & \cdots & 0 \\
mS'_1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & -mS'_{i+1} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
mS'_1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -mS'_{m-1}
\end{vmatrix} \\
&= (-1)^{i-j+1}(S'_i - S'_m) \prod_{\substack{\ell=1 \\ \ell \neq i,\ \ell \neq j}}^{m-1} (-mS'_\ell)
\end{aligned}
$$

C. Proof of Lemma 2

Let D_i be the matrix obtained from D by replacing the i-th column with

$$
\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m-1} \end{pmatrix} = (C_1 - 2C_2\tau) \begin{pmatrix} m\lambda^*_1 - 1 \\ m\lambda^*_2 - 1 \\ \vdots \\ m\lambda^*_{m-1} - 1 \end{pmatrix} - 2C_3\tau \begin{pmatrix} m(\lambda^*_1)^2 - \sum_{j=1}^m \left(\lambda^*_j\right)^2 \\ m(\lambda^*_2)^2 - \sum_{j=1}^m \left(\lambda^*_j\right)^2 \\ \vdots \\ m(\lambda^*_{m-1})^2 - \sum_{j=1}^m \left(\lambda^*_j\right)^2 \end{pmatrix}
$$

Since

$$\frac{d\lambda_m^*(\tau)}{d\tau} = -\sum_{\ell=1}^{m-1} \frac{d\lambda_i^*(\tau)}{d\tau}$$

it is readily verified that the solution to the system (19) is given by

$$\begin{aligned}\frac{d\lambda_i^*(\tau)}{d\tau} &= \frac{|D_i|}{|D|} \\ &= \frac{(-1)^i}{|D|}\sum_{j=1}^{m-1}(-1)^j z_j |D_{ji}| \qquad i = 1, \dots, m-1\end{aligned}$$

for $i = 1, \dots, m-1$. Using (23) and (24) in Appendix B, this expression becomes

$$\frac{(-1)^i}{|D|} \sum_{\substack{j=1 \\ j \neq i}}^{m-1} z_j (-1)^{-i+1}(S'_j - S'_m) \prod_{\substack{\ell=1 \\ \ell \neq i \\ \ell \neq j}}^{m-1} (-mS'_\ell) + \frac{z_i}{m|D|}\left(1 + \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{S'_m}{S'_j}\right) \prod_{\substack{\ell=1 \\ \ell \neq i}}^{m-1} (-mS'_\ell)$$

which is equal to

$$\begin{aligned}
&\frac{1}{m|D|} \prod_{\substack{\ell=1 \\ \ell \neq i}}^{m-1} (-mS'_\ell) \left[\frac{m(-1)^i \sum_{\substack{j=1 \\ j \neq i}}^{m-1} z_j(-1)^{-i+1}(S'_j - S'_m)}{-mS'_j} + z_i\left(1 + \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{S'_m}{S'_j}\right)\right] \\
&= \frac{1}{m|D|} \prod_{\substack{\ell=1 \\ \ell \neq i}}^{m-1} (-mS'_\ell) \left[\sum_{\substack{j=1 \\ j \neq i}}^{m-1} z_j - \sum_{\substack{j=1 \\ j \neq i}}^{m-1} z_j \frac{S'_m}{S'_j} + z_i + z_i \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{S'_m}{S'_j}\right] \\
&= \frac{1}{m|D|} \prod_{\substack{\ell=1 \\ \ell \neq i}}^{m-1} (-mS'_\ell) \left[\sum_{\substack{j=1 \\ j \neq i}}^{m-1} z_j - \sum_{\substack{j=1 \\ j \neq i}}^{m} z_j \frac{S'_m}{S'_j} + z_i + z_m + z_i \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{S'_m}{S'_j}\right] \\
&= \frac{1}{m|D|} \prod_{\substack{\ell=1 \\ \ell \neq i}}^{m-1} (-mS'_\ell) \left[\sum_{j=1}^{m} z_j + \sum_{j=1}^{m} \frac{S'_m}{S'_j}(z_i - z_j)\right] \\
&= \frac{1}{m|D|} \prod_{\substack{\ell=1 \\ \ell \neq i}}^{m-1} (-mS'_\ell) \sum_{j=1}^{m} \frac{S'_m}{S'_j}(z_i - z_j)
\end{aligned}$$

If $S_s' \neq 0$, then the above expression is equal to

$$\begin{aligned}\frac{d\lambda_i^*(\tau)}{d\tau} &= \frac{1}{m}\sum_{j=1}^{m}\left(\frac{d\lambda_i^*(\tau)}{d\tau} - \frac{d\lambda_j^*(\tau)}{d\tau}\right) \\ &= \frac{1}{(-S_i')\, m^3|D|}\prod_{\ell=1}^{m}(-mS_\ell')\sum_{j=1}^{m}\frac{1}{(-S_j')}\left[z_i - \frac{\sum_{j=1}^{m} z_j/\left(-S_j'\right)}{\sum_{j=1}^{m} 1/\left(-S_j'\right)}\right] \qquad \text{for } i = 1,\dots,m\end{aligned} \tag{25}$$

If $S_s' = 0$, then the expression (25) for region $i \neq s$ is equal to

$$\frac{d\lambda_i^*(\tau)}{d\tau} = \frac{1}{(-S_i')m^2|D|}\prod_{\substack{\ell=1\\ \ell\neq s}}^{m}(-mS_\ell')\,(z_i - z_s)$$

For region s, (25) is

$$\frac{d\lambda_s^*(\tau)}{d\tau} = -\sum_{\substack{j=1\\ j\neq s}}^{m}\frac{d\lambda_j^*(\tau)}{d\tau} = \frac{1}{m^2|D|}\prod_{\substack{\ell=1\\ \ell\neq s}}^{m}(-mS_\ell')\sum_{\substack{j=1\\ j\neq s}}^{m}\frac{z_s - z_j}{\left(-S_j'\right)}$$

□

D. Proof of Lemma 3

If $C_1 - 2[C_2 + C_3(\lambda_i^* + \lambda_{i+1}^*)]\tau < 0$ for all $i = 1,\dots,m-1$, then

$$z_i - z_{i+1} = m(\lambda_i^* - \lambda_{i+1}^*)\{C_1 - 2[C_2 + C_3(\lambda_i^* + \lambda_{i+1}^*)]\tau\}$$

is non-negative for all i since $\lambda_i^* \geq \lambda_{i+1}^*$. In this case, we let $k = m$. Otherwise, let

$$k = \min\{i|\ C_1 - 2[C_2 + C_3(\lambda_i^* + \lambda_{i+1}^*)]\tau \geq 0\}$$

If $i < k$, then $C_1 - 2[C_2 + C_3(\lambda_i^* + \lambda_{i+1}^*)\tau] < 0$, which implies $z_i \leq z_{i+1}$. On the other hand, if $i \geq k$, then $\lambda_i^* + \lambda_{i+1}^* \leq \lambda_k^* + \lambda_{k+1}^*$, and hence

$$C_1 - 2[C_2 + C_3(\lambda_i^* + \lambda_{i+1}^*)]\tau \geq C_1 - 2[C_2 + C_3(\lambda_k^* + \lambda_{k+1}^*)]\tau \geq 0$$

which means $z_i \geq z_{i+1}$. □

E. Proof of Lemma 4

Consider first the case where $S'_s \neq 0$.

(i) For $\tau \geq \tau_2^* > 0$, we have $k = k_+ = m$ and z_i is monotone increasing in i (i.e., $z_i \leq z_{i+1}$). From the assumption of asymmetric equilibrium, there is at least one region with a population increase $d\lambda_i^*/d\tau < 0$, it must be that $k_- > 1$. Hence, regions $1, \ldots, k_- - 1$ become larger, while all the other regions become smaller.

(ii) For $\tau_1^* < \tau < \tau_2^*$, z_i is either monotone or single-peaked. Since there is at least one region with a population increase either $1 < k_-$ or $k_+ < m$ holds. In the former case, $1 < k_- \leq k_+ \leq m$ holds, and in the latter case $1 \leq k_- \leq k_+ < m$ holds. By the definition of regions k_- and k_+, we know that regions $1, \ldots, k_- - 1, k_+ + 1, \ldots, m$ become larger. From Lemma 3 with $k_- \leq k \leq k_+$, all the other regions $k_-, \ldots, k_+$ become smaller.

(iii) For $\tau \leq \tau_1^*$, $k_- = k = 1$, and z_i is monotone decreasing in i. Since $k_+ > 1$ due to the assumption of asymmetric equilibrium, regions $1, \ldots, k_+$ become smaller, while all the other regions become larger.

Consider now the case where $S'_s = 0$.

(i') For $\tau \geq \tau_2^* > 0$, we have $z_1 \leq \ldots \leq z_{s-1} < z_s \leq z_m$ from (22). From the first equation of (21), regions $1, \ldots, s-1$ become larger and regions $s+1, \ldots, m$ become smaller, and hence $k_+ = m$. If region s is becomes smaller, then $k_- = s$ If region s becomes larger, then $k_- = s + 1$.

(ii') For $\tau_1^* < \tau < \tau_2^*$. If regions $s-1$ and $s+1$ are both increasing, then $z_{s-1} > z_s$ and $z_{s+1} > z_s$. Thus,

$$C_1 - 2[C_2 + C_3(\lambda_{s-1}^* + \lambda_s^*)]\tau > 0 \qquad \text{and} \qquad C_1 - 2[C_2 + C_3(\lambda_s^* + \lambda_{s+1}^*)]\tau < 0$$

hold from (18) and (22). By subtraction, we have $2C_3(\lambda_{s+1}^* - \lambda_{s-1}^*)\tau > 0$, which contradicts (18). Similarly, we can show that both regions $s+1$ and $s-1$ cannot become smaller. Therefore, s should be either one of $k_-, k_- - 1, k_+, k_+ + 1$. If region s becomes larger, then s should be either $k_- - 1$ or $k_+ + 1$. If region s becomes smaller, then s should be either k_- or k_+, and hence (ii) also again.

(iii') For $\tau \leq \tau_1^*$, the similar argument as (i') applies. □

F. Proof of Lemma 5

Since the proof of Lemma 4 can be applied for regions $i \neq s$, we prove it only for region $s = m$ excluding the case of $d\lambda_s^*(\tau)/d\tau > 0$ with $\tau_1^* < \tau < \tau_2^*$.

(i) For $\tau \geq \tau_2^* > 0$, $z_m - z_j > 0$ holds for all $j = 1, \ldots, m-1$ from (22). Therefore,

$$z_m - \bar{z} = \frac{1}{\sum_{j=1}^{m} 1/(-S'_j)} \sum_{j=1}^{m-1} \frac{z_m - z_j}{-S'_j} < 0 \tag{26}$$

and hence, $d\lambda_m^*(\tau)/d\tau > 0$. Since $\bar{z} > z_{m-1} \geq \ldots \geq z_1$ holds from $\bar{z} > z_m$, $d\lambda_i^*(\tau)/d\tau < 0$ holds for all $i = 1, \ldots, m-1$. Thus, Lemma 4 (i) holds with $k_- = m$.

(ii) For $\tau_1^* < \tau < \tau_2^*$, $z_m \geq \bar{z}$ holds if $d\lambda_m^*(\tau)/d\tau \leq 0$. Since $z_{k_-} \geq \bar{z}$, by definition of k_-, $z_j \geq \bar{z}$ holds for all $j = k_- + 1, \ldots, m-1$. Otherwise, $z_j < \bar{z}$ holds for some j, which implies $z_j < z_{k_-}$ and $z_j < z_m$. Hence,

$$C_1 - 2[C_2 + C_3(\lambda_m^* + \lambda_j^*)\tau] < 0 \quad \text{and} \quad C_1 - 2[C_2 + C_3(\lambda_{k_-}^* + \lambda_j^*)\tau] > 0.$$

from the fact that $\lambda_m \leq \lambda_j \leq \lambda_{k_-}$. By subtraction, we obtain $C_3(\lambda_{k_-}^* - \lambda_m^*) < 0$ which is a contradiction. Therefore, regions $k_-, \ldots, m-1$ become smaller and (ii) of Lemma 4 holds with $k_+ = m-1$.

(iii) For $\tau \leq \tau_1^*$, $z_m - z_j < 0$ holds for all $j = 1, \ldots, m-1$, and hence $d\lambda_m^*(\tau)/d\tau < 0$ from the equality of (26). Since $\bar{z} < z_m < z_{m-1} \leq \cdots \leq z_1$ holds, we have $d\lambda_i^*(\tau)/d\tau > 0$ for all $i = 1, \ldots, m-1$. Thus, Lemma 4 (iii) holds with $k_+ = m-1$ □

G. Proof of Lemma 6

For convenience, we set $S_i(\lambda) = S(\lambda)$. Because $S(\lambda)$ is concave and $\lim_{\lambda\to\infty} S(\lambda) < 0 = S(0)$, $S(\lambda)$ has a maximizer denoted by λ^p. If $\lambda^p \leq 0$, there is no irregular equilibrium. Thus, consider the case where $\lambda^p > 0$. For all $x \in [0, \lambda^p]$, let

$$f(x) \equiv S(\lambda^p + x) - S(\lambda^p - x)$$

then

$$\begin{aligned}
f'(x) &= S'(\lambda^p + x) + S'(\lambda^p - x) \\
&= 2\left(C_1\tau - C_2\tau^2\right) - 4C_3\tau^2\lambda^p - \theta'(\lambda^p - x) - \theta'(\lambda^p + x) \\
&= 2\theta'(\lambda^p) - \theta'(\lambda^p - x) - \theta'(\lambda^p + x) \\
&= [\theta'(\lambda^p) - \theta'(\lambda^p - x)] - [\theta'(\lambda^p + x) - \theta'(\lambda^p)] \\
&= x\left[\theta''(\lambda^p - x_1) - \theta''(\lambda^p + x_2)\right] \\
&\geq 0
\end{aligned}$$

where $x_1, x_2 \in (0, x)$. The third equality follows from $S'(\lambda^p) = 0$, the fifth equality from the mean value theorem, whereas the inequality holds because $\theta'''(\lambda) \leq 0$. Since $f(0) = 0$, we have $f(x) \geq 0$ for all $x \in [0, \lambda^p]$.

Assume that there exists a 2-city equilibrium with different sizes such that

$$S(\lambda^p + y) = S(\lambda^p - x) \qquad x, y > 0 \tag{27}$$

From $f(x) \geq 0$ and (27), we get

$$S(\lambda^p + x) \geq S(\lambda^p - x) = S(\lambda^p + y)$$

implying that $y \geq x$ holds from $S'(\lambda) \leq 0$ for $\lambda \geq \lambda^p$.

Equation (27) implies that

$$\begin{aligned} 0 &= \int_0^{\lambda^p + y} S'(\lambda) d\lambda - \int_0^{\lambda^p - x} S'(\lambda) d\lambda \\ &= \int_{\lambda^p}^{\lambda^p + y} S'(\lambda) d\lambda + \int_{\lambda^p - x}^{\lambda^p} S'(\lambda) d\lambda \\ &\leq y \left[S'(\lambda^p + y) - S'(\lambda^p) \right] /2 + x \left[S'(\lambda^p - x) - S'(\lambda^p) \right] /2 \\ &\leq x \left[S'(\lambda^p + y) + S'(\lambda^p - x) \right] /2 \end{aligned}$$

where the first inequality holds because $S'(\lambda^p + y) < S'(\lambda^p) = 0 < S'(\lambda^p - x)$ and $S'(\lambda)$ is convex, whereas the second inequality follows from $y \geq x$ and $S'(\lambda^p + y) \leq 0$. Thus, we have $S'(\lambda^p + y) + S'(\lambda^p - x) \geq 0$, which violates the stability condition (15). Hence, there is no irregular equilibrium with 2 cities. It can be similarly verified that any equilibrium with more than 2 cities having different sizes violates the stability condition (15). □

H. Proof of Proposition 1

The concavity of $S_i(\lambda_i)$ implies that $S_i(\lambda_i) = 0$ has a solution at $\lambda_i = 0$ and at most one solution in $(0, \infty)$. Denote this one by λ_i^o if it exists. Since urban costs are the same across regions, the surplus function S_i is independent of i so that $\lambda_i^o = \lambda^o$. We first show that λ^o is quasi-concave with respect to τ. From

$$\frac{d\lambda^o}{d\tau} = -\frac{\partial S(\lambda^o, \tau)/\partial \tau}{\partial S(\lambda^o, \tau)/\partial \lambda} = \frac{\lambda^o}{-S'(\lambda^o)} \left[C_1 - 2(C_2 + C_3 \lambda^o)\tau \right]$$

and the fact that $-S'(\lambda^o) > 0$ when $\lambda^o > 0$, it follows that

$$\text{sgn}\,(d\lambda^o/d\tau) = \text{sgn}\,[C_1 - 2(C_2 + C_3\lambda^o)\tau]$$

holds for any $\lambda^o > 0$.

If λ^o is monotone, then it is quasi-concave. If it is not monotone, then there exists τ^* satisfying

$$\left.\frac{d\lambda^o}{d\tau}\right|_{\tau=\tau^*} = 0 \tag{28}$$

The second derivative evaluated at τ^* is given by

$$\begin{aligned}\left.\frac{d^2\lambda^o}{d\tau^2}\right|_{\tau=\tau^*} &= \frac{1}{-S'(\lambda^o)}\left.\frac{\partial^2 S(\lambda^o,\tau)}{\partial\tau^2}\right|_{\tau=\tau^*} \\ &= \frac{\lambda^o}{-S'(\lambda^o)}\,[-2(C_2 + C_3\lambda^o)\lambda^o] < 0\end{aligned}$$

Thus, τ^* satisfying (28) must be unique. As a consequence, λ^o is increasing in $(0, \tau^*)$ and decreasing in (τ^*, ∞), and hence quasi-concave.

Because urban costs are identical, it is readily verified that an equilibrium with n cities is stable if $n\lambda^p < 1$, and an equilibrium with $m < n$ cities is stable if and only if $m\lambda^o > 1$ and $m\lambda^p < 1$.[10] Therefore, for decreasing τ with $\tau > \tau^*$ (resp. $\tau < \tau^*$), if an equilibrium with m cities ceases to exist, then $m\lambda^p = 1$ (resp. $m\lambda^o = 1$) so that any equilibrium with $m' > m$ (resp. $m' < m$) cities is unstable. Hence, the number of cities must decrease (resp. increase) as τ falls. □

References

[1] Alonso W. (1980) Five bell shapes in development, *Papers of the Regional Science Association* 45, 5-16.

[2] Bairoch P. (1988) *Cities and Economic Development: From the Dawn of History to the Present*, Chicago, University of Chicago Press.

[3] Berry B.J. (1976) *Urbanization and Counterurbanization*, London, Sage Publications.

[10] We exclude a finite number of τ-values for which $m\lambda^o = 1$ or $m\lambda^p = 1$ holds.

[4] Brakman S., H. Garretsen and C. van Marrewijk (2001) *An Introduction to Geographical Economics*, Cambridge, Cambridge University Press.

[5] Champion A.G. (1994) Population change and migration in Britain since 1981: evidence for continuing deconcentration, *Environment and Planning A* 26, 1501-1520.

[6] Fujita M. (1989) *Urban Economic Theory. Land Use and City Size*, Cambridge, Cambridge University Press.

[7] Fujita M. and P. Krugman (1995) When is the economy monocentric? von Thünen and Chamberlin unified. *Regional Science and Urban Economics* 25, 505-28.

[8] Fujita M., P. Krugman and T. Mori (1999) On the evolution of hierarchical urban systems. *European Economic Review* 43, 209-51

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

[10] Fujita M. and J.-F. Thisse (2002) *Economics of Agglomeration. Cities, Industrial Location and Regional Growth*, Cambridge, Cambridge University Press.

[11] Geyer H.S. and T.M. Kontuly (1996) *Differential Urbanization: Integrating Spatial Models*, London, Arnold.

[12] Ginsburgh V., Y.Y. Papageorgiou and J.-F. Thisse (1985) On existence and stability of spatial equilibria and steady-states. *Regional Science and Urban Economics* 15, 149-158.

[13] Head K. and T. Mayer (2000) Non-Europe, The magnitude and causes of market fragmentation in the EU. *Weltwirtschaftliches Archiv* 136, 284-314.

[14] 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.

[15] Henderson J.V. (1974) The sizes and types of cities, *American Economic Review* 64, 640-656.

[16] Henderson J.V. (1988) *Urban Development. Theory, Fact and Illusion*, Oxford, Oxford University Press.

[17] Krugman P. (1991) Increasing returns and economic geography, *Journal of Political Economy* 99, 483-499.

[18] Krugman P. (1993) On the number and location of cities, *European Economic Review* 37, 293-298.

[19] McCallum J. (1995) National borders matter: Canada-US regional trade patterns. *American Economic Review* 85, 615-623.

[20] Ottaviano G., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited, *International Economic Review* 43, 101-127.

[21] Papageorgiou Y.Y. and D. Pines (1999) *An Essay in Urban Economic Theory*, Dordrecht, Kluwer Academic Publishers.

[22] Pollard S. (1981) *Peaceful Conquest. The Industrialization of Europe 1760-1970*, Oxford, Oxford University Press.

[23] Tabuchi T. (1986) Existence and stability of city-size distribution in the gravity and logit models, *Environment and Planning A* 18, 1375–1389.

[24] Tabuchi T. (1998) Agglomeration and dispersion: a synthesis of Alonso and Krugman, *Journal of Urban Economics* 44, 333-51.

[25] Tabuchi T. and J.-F. Thisse (2002) Regional specialization and transport costs, mimeographed.

[26] Tabuchi T. and D.-Z. Zeng (2001) Stability of spatial equilibrium, Faculty of Economics, University of Tokyo, http://www.e.u-tokyo.ac.jp/˜ttabuchi.

[27] Vining D.R. and T. Kontuly (1978) Population dispersal from major metropolitan regions: an international comparison, *International Regional Science Review* 3, 49-73.

[28] Zeng D.-Z. (2002) Equilibrium stability for a migration model, *Regional Science and Urban Economics* 32, 123-138.

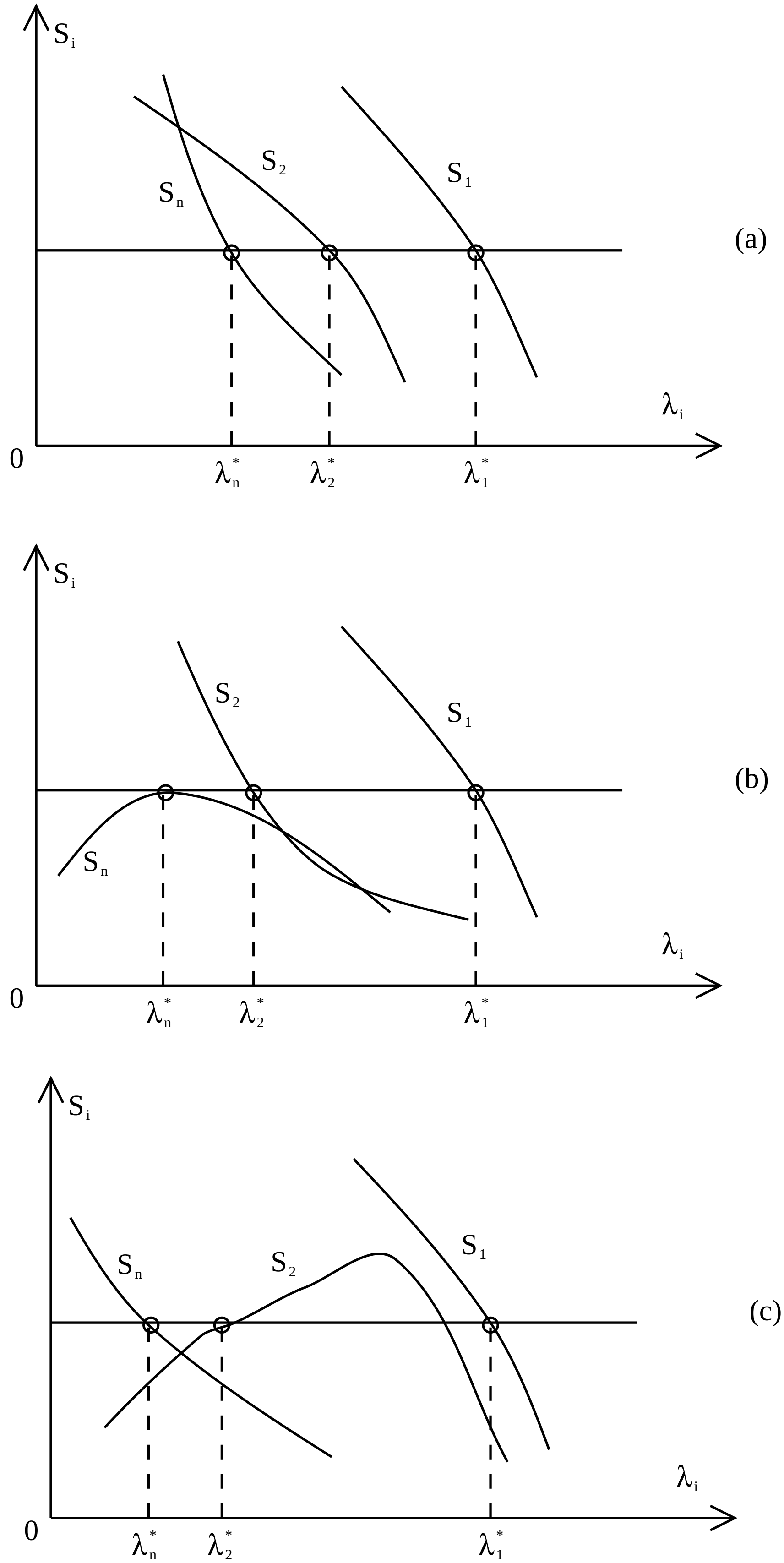

Figure 1: Regular and irregular equilibria

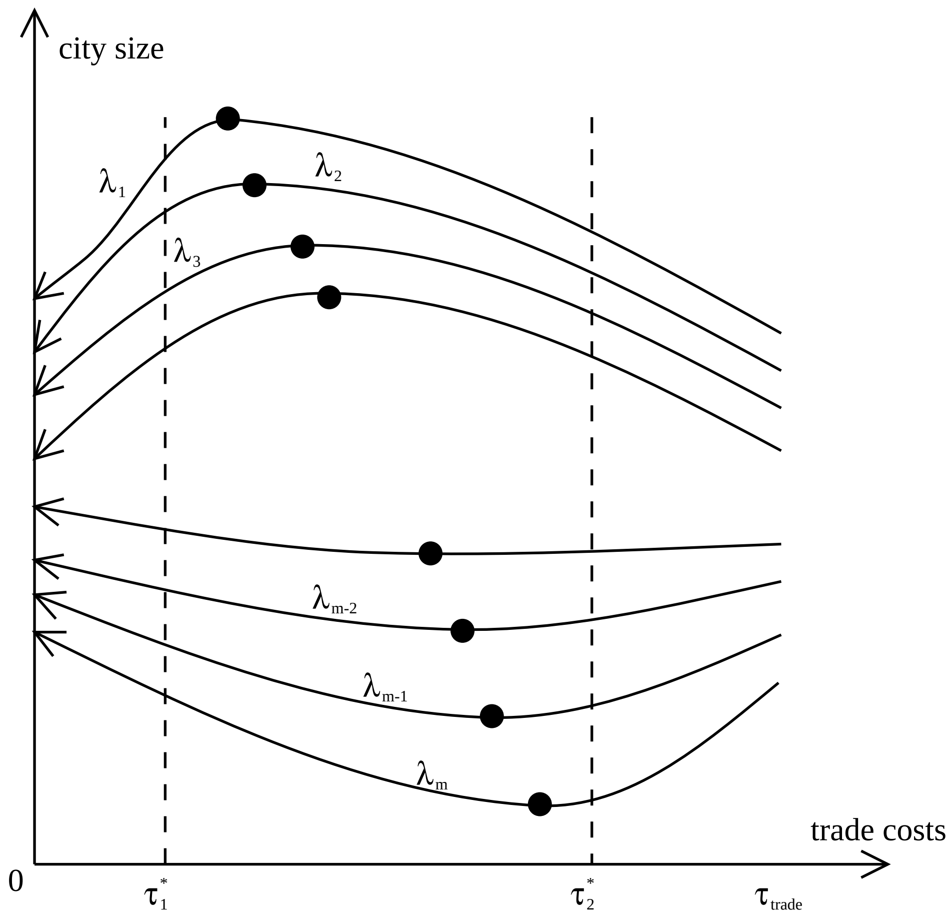

Figure 2: Evolution of city size distributions

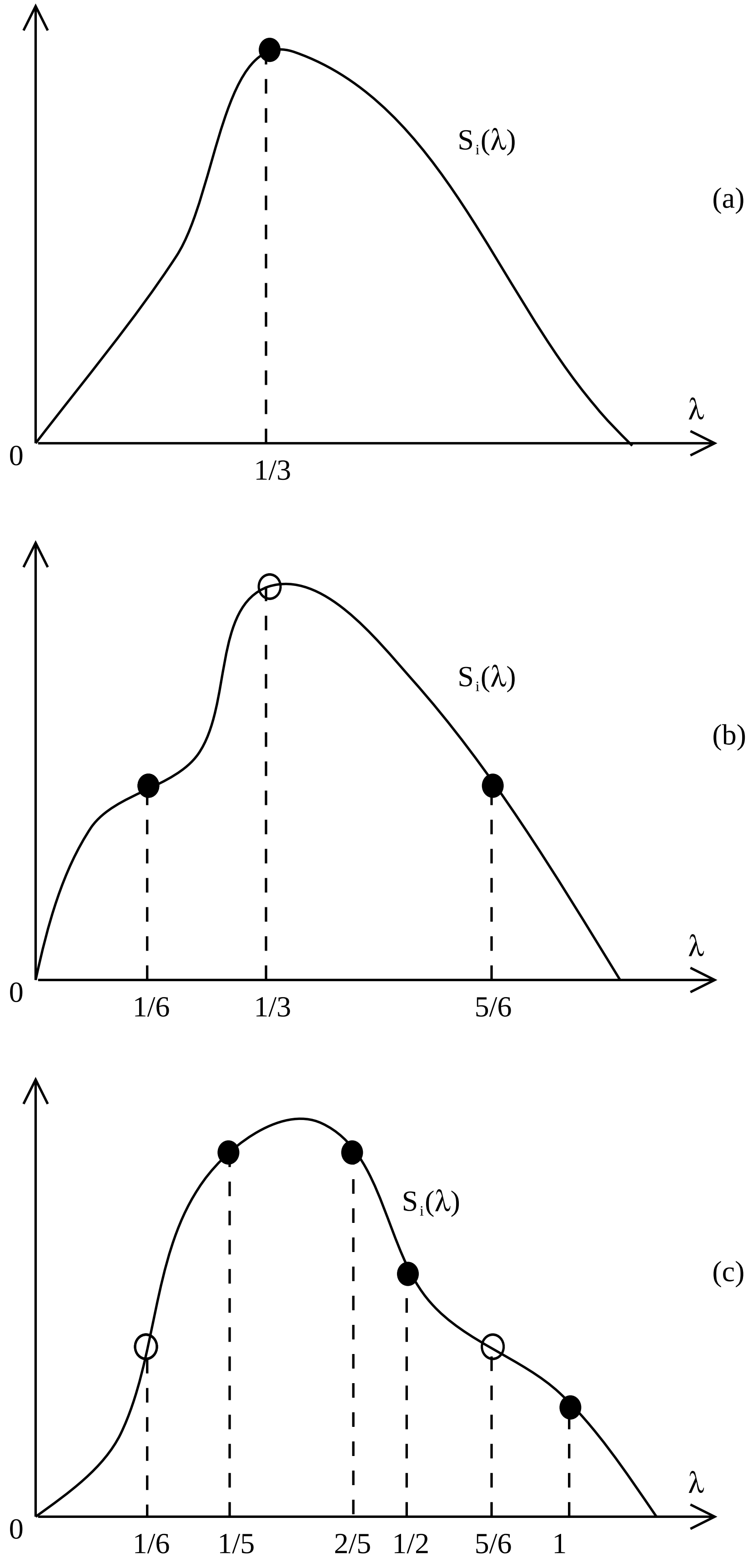

Figure 3: Multiple equilibria in the number of cities