

Trade and the structure of cities

Jean Cavailhès^a, Carl Gaigné^{b,*}, Takatoshi Tabuchi^c,
Jacques-François Thisse^{d,e,f,g}

^a INRA-CESAER, Dijon, France

^b INRA, UR122, Economie Rurale, 4 Allée Adolphe Bobierre, F-35011 Rennes, France

^c Faculty of Economics, University of Tokyo, Japan

^d ECORE, Université Catholique de Louvain, Belgium

^e Institute of Economic Analysis, Kyoto University, Japan

^f PSE, France

^g CEPR, UK

Received 30 May 2006; revised 7 December 2006

Available online 28 December 2006

Abstract

Our purpose is to investigate how the interplay between trade, commuting and communication costs shapes the economy at both the interregional and intra-urban levels. Specifically, we study how economic integration affects the internal structure of cities and show how decentralizing the production and consumption of goods in secondary employment centers allows firms located in a large city to maintain their predominance. Several new results in both economic geography and urban economics are established, which all agree with empirical evidence.

© 2007 Elsevier Inc. All rights reserved.

JEL classification: F12; F22; R12; R14

Keywords: City structure; Secondary business centers; Commuting costs; Trade costs; Communication costs

1. Introduction

Cities are major actors in the process of trade. It is, therefore, fundamental to understand (i) how the intensity of trade is influenced by their size and structure and, conversely, (ii) how

* Corresponding author. Fax: 33 (0)2 23 48 53 80
E-mail address: gaigne@rennes.inra.fr (C. Gaigné).

economic integration affects the internal structure of cities. This is what we undertake in this paper by modeling the interplay between *trade costs*, *commuting costs* and *communication costs*. Our approach, which combines basic ingredients from urban economics and new economic geography, explains how decentralizing the production of goods in secondary employment centers may allow large cities to retain a large share of firms and jobs in an integrating world.

Our starting point is that firms' performances are affected by the level of housing and commuting costs, which we call "urban costs". This occurs through the land rent they pay to occupy central urban locations, and through the higher wages they have to pay to their workers to compensate them for their longer commutes and/or higher land rents. Hence, high urban costs render firms less competitive on local and foreign markets alike. As a result, despite scale economies arising from urban agglomeration (Duranton and Puga [5]), increasing urban costs could shift employment from large monocentric cities either to their suburbs or to distant and smaller cities, where these costs are lower, at least once trade costs have sufficiently declined to permit large-scale exports to distant markets. In other words, economic integration could well challenge the supremacy of large cities in favor of small cities. The main point we wish to stress in this paper is that *the emergence of subcenters within cities is a powerful strategy for large cities to maintain their attractiveness*.

Despite the many advantages provided by the inner city through a good access to highly specialized services (Porter [21]), firms or developers may choose to form secondary employment centers, enterprise zones, or edge cities (Henderson and Mitra [15]). In this way, firms are able to pay lower wages and land rents while retaining most of the benefits generated by large urban agglomerations. And, indeed, Timothy and Wheaton [25] report large variations in wages according to intra-urban location (15% higher in central Boston than in outlying work zones, 18% between central Minneapolis and the fringe counties). As they enjoy living on larger plots and/or move along with firms, workers may also want to live in suburbia (Glaeser and Kahn [10]). Consequently, the creation of subcenters within a city, i.e. the formation of a polycentric city, appears to be a natural way to alleviate the burden of urban costs. It is, therefore, no surprise that Anas et al. [2] observe that "polycentricity is an increasingly prominent feature of the landscape".¹

Thus, the escalation of urban costs in large cities seems to prompt a redeployment of activities in a polycentric pattern, while smaller cities retain their monocentric shape. However, for this to happen, firms set up in the secondary centers must maintain a very good access to the main urban center, which requires low communication costs. Indeed, as pointed out by Schwartz [22], about half of the business services consumed by US firms located in suburbia are supplied in city centers. As a result, by focusing on urban and communication costs, we recognize that both *agglomeration* and *dispersion* may take two quite separate forms because they are now compounded by *centralization* or *decentralization* of activities within the same city. Such a distinction is crucial for understanding the interaction between cities and trade.

To achieve our goal, we develop a two-region model where regions have a spatial extension that imposes commuting and communication costs whereas interregional shipments of commodities imply trade costs. Unlike Helpman [13], Tabuchi [23] and others, our framework allows cities to be polycentric. We will see how this seemingly minor change sheds light on several important concrete issues that have been pretty much overlooked until now. More precisely, we organize our main conclusions around two main ideas: (i) *local factors may change the global organi-*

¹ To illustrate, Giuliano and Small [8] identify 29 job centers in Los Angeles, McMillen and McDonald [16] find 15 in Chicago, and Creveso and Wu [4] count 22 for the San Francisco Bay Area.

zation of the economy, whereas (ii) global forces may affect the local organization of cities. This interaction arises because the global organization of production and employment may take different forms as either a single polycentric city or two monocentric cities may emerge, thus yielding very contrasted economic landscapes and trade patterns. Regarding the first idea, we focus on communication and commuting costs. When these costs are high, the economic landscape is likely to be formed by several small cities trading differentiated varieties. By contrast, when commuting and/or communication costs reach low values in each region, a large city emerges. In particular, by facilitating the formation of secondary centers, the development of New Information and Communication Technologies (in short NICTs) may prevent the re-dispersion of activities between regions that a deep economic integration is expected to trigger (Ottaviano and Thisse [20]).

Concerning the second idea, our thought experiment is about trade costs. When trade and commuting costs are large, agglomeration within a polycentric city is the market outcome. This is because the decentralization of jobs leads to lower land rents for workers and allows firms to pay them lower wages. This result agrees with the formation of megalopolises in which employment is decentralized in several centers that all belong to the same metropolitan area (MacMillen and Smith [17]). When trade costs fall below some threshold, the agglomeration becomes *partial* in that it loses jobs to the benefit of smaller cities because inter-city trade is cheaper. Yet, even for fairly low trade costs, the large city is able to maintain its prevalence because its polycentricity allows it to preserve its competitive position through lower wages and land rent.

In the sections that follow, we first describe our modeling strategy (Section 2). The intra-urban equilibrium is characterized in Section 3, whereas the subsequent section analyzes the urban system when cities have given structures. In Section 5, we study the impact of trade and communication costs on the size and structure of cities. Section 6 concludes.

2. The model

2.1. The spatial economy

Consider an economy with two regions, labeled $r = 1, 2$, separated by a given physical distance, one sector and two primary goods, labor and land. Each region can be urbanized by accommodating firms and workers within a city, and is formally described by a one-dimensional space X . Whenever a city exists, it has a central business district (in short CBD) located at the origin $0 \in X$. One would expect us to explain why this CBD exists as well as why firms leaving the CBD want to be together and form a *secondary business district* (in short SBD). Doing that would require the introduction of local spatial externalities and local public goods that would render the analysis much more involved from the technical point of view, without adding much to our results. Indeed, our model has nothing new to add to what is known in this domain. By contrast, we determine the sizes of the CBD and the SBDs, thus the structure of each city, in the presence of inter-city trade and factor mobility.²

² Thus, we differ from Fujita et al. [7] because cities have a spatial extension and an endogenous structure. Unlike them, however, the inter-city distance is given. We also differ from Henderson [14] who considers monocentric cities and zero transport costs between cities.

Firms are free to locate in the CBD or to set up in the suburbs of the metro where they form a SBD. Both the CBD and SBDs are assumed to be dimensionless.³ In what follows, the superscript *C* is used to describe variables related to the CBD, whereas *S* describes the variables associated with a SBD. Without loss of generality, we focus on the right-hand side of the city, the left-hand side being perfectly symmetrical. Distances and locations are expressed by the same variable x measured from the CBD located at $x = 0$ in city $r = 1, 2$ whereas the SBD, if any, is established at $x_r^S > 0$, which is endogenous.

Even though firms consume services supplied in each SBD, the higher-order functions (specific local public goods and non-tradeable business-to-business services such as marketing, banking, insurance) are still located in the CBDs (Schwartz [22]). Hence, for using such services, firms set up in a SBD must incur a *communication cost*, which is given by

$$\mathcal{K}(x_r^S) = K + kx_r^S \quad (1)$$

where K and k are two positive constants. Indeed, communicating requires the acquisition of specific facilities, thus explaining why communication costs have a fixed component. However, relationships between the CBD and a SBD also involves face-to-face communication. Therefore, some workers must go to the CBD, thus making communication costs dependent on the distance x_r^S between the CBD and the SBD. For simplicity, we assume that this cost is linear in distance, but this does not affect the nature of our results.

Both the CBD and the SBD are surrounded by residential areas occupied by workers. Furthermore, as the distance between the CBD and SBD is small compared to the intercity distance, we disregard the intra-urban transport cost of goods. Finally, for analytical convenience we restrict ourselves to the case where the CBD is surrounded by two SBDs. It should be clear, however, that our analysis can be extended to several subcenters as communication costs get smaller.

Under those various assumptions, the location and size of the SBDs as well as the size of the CBD are endogenously determined. In other words, apart from the assumed existence of the CBD, the internal structure of each city is endogenous.

2.2. Workers

The economy is endowed with L mobile workers. The welfare of a worker depends on her consumption of the following three goods. The first good is unproduced and homogeneous.⁴ It is assumed to be costlessly tradeable and chosen as the numéraire. The second good is produced as a continuum n of varieties of a horizontally differentiated good under monopolistic competition and increasing returns, using labor as the only input. Any variety of this good can be shipped from one city to the other at a unit cost of $\tau > 0$ units of the numéraire. The third good is land; without loss of generality, we set the opportunity cost of land to zero.

Each worker living in city r consumes a residential plot of fixed size chosen as the unit of area.⁵ The worker also chooses a quantity $q(i)$ of variety $i \in [0, n]$, and a quantity q_0 of the numéraire. She is endowed with one unit of labor and $\bar{q}_0 > 0$ units of the numéraire. The initial endowment \bar{q}_0 is supposed to be large enough for her consumption of the numéraire to be

³ In Cavailhès et al. [3], we suppose that firms consume land, thus implying that clusters have a spatial size. This makes the analytical treatment of the model more cumbersome without changing the nature of our results.

⁴ The model can easily be extended by introducing a second sector producing the homogeneous good under constant returns and perfect competition, using an immobile factor.

⁵ Allowing for a variable lot size makes the analysis much more involved without affecting the nature of our results. See Tabuchi [23] for a study of the monocentric-city case.

strictly positive at the market outcome. Each worker commutes to her employment center—without cross-commuting—and bears a unit commuting cost given by $t > 0$, so that for the worker located at x the commuting cost is either tx or $t|x - x_r^S|$ according to the employment center.

The budget constraint of an individual residing at $x \in X$ in city r and working in the corresponding CBD can then be written as follows:

$$\int_0^n p_r(i)q(i) di + q_0 + R_r^C(x) + tx = w_r^C + \bar{q}_0 \tag{2}$$

where $R_r^C(x)$ is the land rent prevailing at a distance x from the CBD. The budget constraint of an individual working in the SBD is obtained by replacing tx by $t|x - x_r^S|$, $R_r^C(x)$ by $R_r^S(x)$, and w_r^C by w_r^S . Thus, as in Ogawa and Fujita [18], commuting costs and wages are endogenously determined by the global distribution of firms and workers within the city.

Preferences over the differentiated product and the numéraire are identical across workers and represented by a quasi-linear utility encapsulating a quadratic sub-utility:

$$U(q_0; q(i), i \in [0, n]) = \alpha \int_0^n q(i) di - \frac{\beta - \gamma}{2} \int_0^n [q(i)]^2 di - \frac{\gamma}{2} \left[\int_0^n q(i) di \right]^2 + q_0 \tag{3}$$

where $\alpha > 0$ and $\beta > \gamma > 0$. The condition $\beta > \gamma$ implies that workers have a preference for variety.

2.3. Firms

Technology in manufacturing is such that producing $q(i)$ units of variety i requires a given number ϕ of labor units.⁶ There are no scope economies so that, due to increasing returns to scale, there is a one-to-one relationship between firms and varieties. Thus, the total number of firms is given by $n = L/\phi$. Labor market clearing implies that the number of firms located (or varieties produced) in city r is such that $n_r = \lambda_r n$, where λ_r stands for the share of workers residing in r .

Denote by Π_r^C (respectively Π_r^S) the profit of a firm set up in the CBD (respectively the SBD) of city r . Let θ_r be the share of firms located in the CBD of city r and, therefore, by $(1 - \theta_r)/2$ the share of firms in its right-hand side SBD. When the firm producing variety i is located in the CBD of city r , its profit function is given by:

$$\Pi_r^C(i) = I_r(i) - \phi w_r^C \tag{4}$$

where $I_r(i)$ stands for the firm’s revenue earned from local sales and from exports (see (14) below). When the firm sets up in the SBD of the same city, its profit function becomes:

$$\Pi_r^S(i) = I_r(i) - \phi w_r^S - \mathcal{K}(x_r^S) \tag{5}$$

where the firm’s revenue is the same as in the CBD because shipping varieties within the city is costless so that prices and outputs do not depend on firm’s location in the city. Those two expressions encapsulate the trade-off faced by firms located in city r : by locating at the SBD, firms are able to pay a lower wage to workers, but must incur the communication cost $\mathcal{K}(x_r^S)$.

⁶ When a second sector is considered, we may assume that the production of $q(i)$ units of variety i requires $m q(i)$ units of the immobile factor. Without loss of generality, we may then set $m = 0$ (Ottaviano et al. [19]).

2.4. Market structure

Solving the budget constraint for the numéraire consumption, plugging the corresponding expression into (3) and taking the first order condition with respect to $q(i)$ yields

$$\alpha - (\beta - \gamma)q(i) - \gamma \int_0^n q(j) dj = p(i) \quad i \in [0, n].$$

The demands for a variety i produced in city r by a worker living in city r and a worker living in city s can then be written, respectively, as follows:

$$q_{rr}(i) = a - (b + cn)p_{rr}(i) + cP_r, \tag{6}$$

$$q_{rs}(i) = a - (b + cn)p_{rs}(i) + cP_s \tag{7}$$

where $p_{rr}(i)$ (respectively $p_{rs}(i)$) denotes the price a variety- i firm located in city r charges to consumers living in city r (respectively city $s \neq r$) and P_r the average price (up to n) of all varieties in city r :

$$P_r \equiv \int_0^{n_r} p_{rr}(i) di + \int_0^{n_s} p_{sr}(i) di \quad s \neq r. \tag{8}$$

Furthermore, we have $a \equiv \alpha b$, $b \equiv 1/[\beta + (n - 1)\gamma]$ and $c \equiv [\gamma/(\beta - \gamma)]b$. Parameter a expresses the desirability of the differentiated product with respect to the numéraire and may, therefore, be viewed as a measure of the size of this market; b gives the link between individual and industry demands: when b rises, consumers become more sensitive to price differences. Finally, parameter c is an inverse measure of the degree of product differentiation between varieties; when $c \rightarrow \infty$, varieties are perfect substitutes, whereas they are independent for $c = 0$.

Firm i located in city r faces a downward sloping demand in city r and city $s \neq r$:

$$Q_{rr}(i) = \lambda_r Lq_{rr}(i), \quad Q_{rs}(i) = \lambda_s Lq_{rs}(i)$$

where $q_{rr}(i)$ and $q_{rs}(i)$ are given by (6) and (7), respectively.

As empirical evidence suggests that firms practice some form of spatial price discrimination (Greenhut [11], Engel and Rogers [6], Haskel and Wolf [12]), we assume that markets are spatially segmented, which means that each firm chooses a delivered price specific to the city in which its variety is sold. As the price of a variety does not vary within a city, the total revenue of firm i located in city r is given by

$$I_r(i) = p_{rr}(i)Q_{rr}(i) + [p_{rs}(i) - \tau]Q_{rs}(i).$$

Because there is a continuum of firms, each firm has a negligible impact on the market outcome in the sense that it may accurately ignore its influence on, and hence reactions from, other firms. However, aggregate market conditions of some kind (here the price index P_r) affect any single firm. This defines a setting in which individual firms are not competitive (in the classic economic sense of having infinite demand elasticity) but, at the same time, they have no strategic interactions with one another. Because varieties are symmetric, all firms located in the same city charge the same price. As shown by Ottaviano et al. [19], the equilibrium prices are as follows:

$$p_{rr}^* = \frac{1}{2} \frac{2a + c\tau(1 - \lambda_r)n}{2b + cn}, \tag{9}$$

$$p_{rs}^* = p_{ss}^* + \frac{\tau}{2} \quad s \neq r. \tag{10}$$

It thus appears that the equilibrium price prevailing in a city decreases with the number of firms located there, but increases with the level of trade costs. Finally, even though factor prices do not enter (9)–(10) because they have the nature of a fixed cost, they have a negative impact on the number of firms set up in city r , whence an indirect positive impact on equilibrium prices.

Substituting (10) into the demands (6)–(7) and using (8), the equilibrium consumption levels can be expressed as follows:

$$q_{rr}^* = a - bp_{rr}^* + cn_s \tau / 2, \tag{11}$$

$$q_{rs}^* = q_{ss}^* - (b + cn) \tau / 2. \tag{12}$$

Not surprisingly, high trade costs raise the local demand for each locally produced variety at the expense of varieties produced in the other city. This substitution effect decreases when varieties becomes more differentiated.

Hence, evaluated at the equilibrium prices (9)–(10), the consumer surplus is given by:

$$S_r^* = \frac{a^2 n}{2b} - a(n_r p_{rr}^* + n_s p_{sr}^*) + \frac{b + cn}{2} [n_r (p_{rr}^*)^2 + n_s (p_{sr}^*)^2] - \frac{c}{2} (n_r p_{rr}^* + n_s p_{sr}^*)^2 \tag{13}$$

while the equilibrium revenue of a firm located in r is may be expressed as follows:

$$I_r = (b + cn) [\lambda_r L p_{rr}^{*2} + \lambda_s L (p_{rs}^* - \tau)^2]. \tag{14}$$

Both (13) and (14) depend on the distribution of workers between the two cities.

It remains to determine the conditions to be imposed on τ for trade to occur between cities at the equilibrium prices regardless of the interregional distribution of workers. This is so if and only if the equilibrium demand q_{rs}^* is positive for any distribution of workers. It is readily verified that this condition is equivalent to:

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

which is assumed to hold throughout the paper. This condition also guarantees that it is always profitable for a firm to export to the other city ($p_{rs}^* - \tau > 0$).

To sum up, we consider a full-fledged general equilibrium model involving labor, land as well as a differentiated product and a homogeneous good. At the global level, increasing returns at the plant level are the agglomeration force whereas urban costs are the dispersion force. At the city level, communication costs act as the agglomeration force and commuting costs as the dispersion force. In the next section, we study the city equilibrium within one city before considering the case of an urban system in the subsequent section.

3. Decentralization within a city

A *city equilibrium* is such that each individual maximizes her utility subject to her budget constraint, each firm maximizes its profits, and markets clear. Individuals choose their workplace (CBD or SBD) and their residential location with respect to given wages and land rents. In each workplace (CBD or SBD), the equilibrium wages are determined by a bidding process in which firms compete for workers by offering them higher wages until no firm can profitably enter the market. Given such equilibrium wages and the location of workers, firms choose to locate either in the CBD or in the SBD. At the city equilibrium, no firm has an incentive to change place within the city, and no worker wants to change her working place and/or her residence. In this section, we analyze such an equilibrium, taking as fixed the number of workers. To ease the burden of notation, we drop the subscript r .

3.1. Land rents, wages and workplaces

Within each city, a worker chooses her location so as to maximize her utility (3) under the budget constraint (2). Let $\Psi^C(x)$ and $\Psi^S(x)$ be the bid rent at $x \in X$ of an individual working respectively in the CBD and the SBD. Land is allocated to the highest bidder.⁷ Because there is only one type of labor, at the city equilibrium it must be that

$$R(x) = \max\{\Psi^C(x), \Psi^S(x), 0\}.$$

Denote by y the right endpoint of the area formed by residents working in the CBD. Let z_1 be the endpoint of the residential area on the left-hand side of the SBD, and z_2 the symmetrical residential endpoint, which is also the outer limit of the city. Because communication costs to the CBD increase with distance, the two residential areas are adjacent when the city is polycentric, which implies $y = z_1$. Therefore, the critical points are as follows:

$$y = \frac{\theta l}{2}, \quad x^S = \frac{1 + \theta}{4} l, \quad z_2 = \frac{l}{2} \tag{15}$$

where l is the city size and θ the share of firms located in the CBD. Note that the bid rents at y and z_2 are equal to zero because the opportunity cost of land is zero. An illustration of the land rent profile is provided in Fig. 1.

Because of the fixed lot size assumption, at the city equilibrium the value of the equilibrium consumption of the nonspatial goods

$$E = \int_0^n p(i)q(i) di + q_0$$

is the same regardless of the worker’s location. Then, the budget constraint of an individual residing at x and working in the CBD implies that $w^C + \bar{q}_0 - R(x) - tx = E$, whereas the budget constraint of an individual working in the SBD is $w^S + \bar{q}_0 - R(x) - t|x - x^S| = E$. At the city equilibrium, the worker living at the right-endpoint y of the CBD residential area (or at

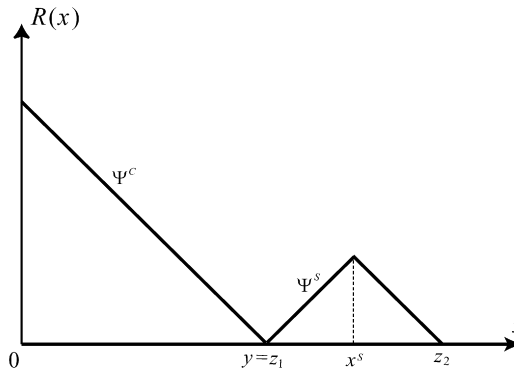


Fig. 1. Equilibrium land rent.

⁷ Utilities being quasi-linear, the structure of land ownership across individuals is immaterial for our analysis provided that the distribution is atomless.

the left-endpoint z_1 of the SBD residential area) is indifferent between working in the CBD or in the SBD, which implies

$$w^C - R(y) - ty = w^S - R(z_1) - t(x^S - z_1).$$

Because $y = z_1$ and $R(y) = R(z_1) = 0$, this equation becomes

$$w^C - w^S = t(2y - x^S) = t \frac{3\theta - 1}{4} l \tag{16}$$

where we have used the expressions of x^S and z_1 given in (15). Thus, *the difference in the wages paid in the CBD and in the SBD compensates exactly the worker for the difference in the corresponding commuting costs.* The wage wedge $w^C - w^S$ is positive as long as $\theta > 1/3$, thus implying that the size of the CBD exceeds the size of each SBD (recall that another SBD exists on the left-hand side of the CBD). Observe that a rise in the population size increases the wage wedge: as the average commuting cost rises, firms located in the CBD must pay a higher wage to their workers.

3.2. Equilibrium wages and the city structure

Regarding the labor markets, the equilibrium wages of workers are determined by the zero-profit condition. In other words, operating profits are completely absorbed by the wage bill. Hence, the equilibrium wage rates in the CBD and in the SBD must satisfy the conditions $\Pi^C(w^{C*}) = 0$ and $\Pi^S(w^{S*}) = 0$, respectively. Thus, setting (4) (respectively (5)) equal to zero, solving for w^{C*} (respectively w^{S*}), we get:

$$w^{C*} = \frac{I}{\phi}, \quad w^{S*} = \frac{I - \mathcal{K}(x^S)}{\phi}. \tag{17}$$

Hence $w^{C*} - w^{S*} = \mathcal{K}(x^S)/\phi$, which means that the equilibrium wage wedge is proportional to the level of the communication cost that prevails at the SBD.

Substituting (17) and (1) into (16) and solving with respect to θ yields:

$$\theta = \frac{4K + (t\phi + k)l}{(3t\phi - k)l}$$

which is positive and exceeds 1/3 if and only if $k/3\phi < t$, which means that commuting costs are large relative to distance-sensitive communication costs. When this condition is not satisfied, the city is monocentric regardless of the value of commuting costs.⁸ Assuming from now on that $k < 3t\phi$ holds, we have

$$\theta^* = \min \left\{ 1, \frac{4K + (t\phi + k)l}{(3t\phi - k)l} \right\}. \tag{18}$$

Observe first that $\theta^* = 1/3$ when $K = k = 0$ because the city is formed by three identical employment centers. Furthermore, when $\theta^* < 1$, increasing the population size leads to a decrease in the relative size of the CBD, though its absolute size rises, whereas both the relative and absolute sizes of the SBD rises. Indeed, increasing the size of the labor force leads to a more than proportionate increase in the wage rate prevailing in the CBD. This is because of the corresponding rise in the average commuting cost. The number of firms being fixed, this in turn implies that

⁸ In this case, the analysis provided in Tabuchi and Thisse [24] in which all workers are mobile applies.

more firms choose to set up in the SBD at the expense of the CBD. Last, as long as $\theta^* < 1$, the higher the communication cost (either K or k), the larger the CBD. In the same way, the lower the commuting cost, the larger the CBD size.

It is readily verified that the city is monocentric if and only if

$$t \leq \frac{2K + kl}{\phi l}. \tag{19}$$

Hence, a polycentric city is more likely to occur when commuting costs are large, communication costs are low, and the population size is large. This agrees with Anas et al. [2] who observe that by the end of the 19th century telephones have made it possible for US firms to decentralize, whereas NICTs play nowadays a comparable role. By contrast, a high degree of increasing returns favors the centralization of production.

We may summarize the main results of that analysis in the following proposition.

Proposition 1. *A city is monocentric if and only if $t \leq (2K + kl)/\phi l$. Otherwise, the city is polycentric.*

Finally, it is worth noting that the equilibrium land rents are given by

$$R(x) = \Psi^C(x) = t \left(\frac{\theta^* l}{2} - x \right) \quad \text{for } x < y \tag{20}$$

where we have used the expression of y and the condition $\Psi^C(y) = 0$ and by

$$R(x) = \Psi^S(x) = t \left(\frac{1 - \theta^*}{4} l + |x^S - x| \right) \quad \text{for } x > x^S. \tag{21}$$

Workers' bid rents around the SBD are maximized at x^S whereas $\Psi^S(z_1) = 0$. The gap $\Psi^C(x) - \Psi^S(|x - x^S|) > 0$ at any given x rises as the relative size of the CBD increases. Note also that the households' bid rents functions Ψ^C and Ψ^S in the CBD and the SBD are identical once the employment centers have the same size, that is, $\theta^* = 1/3$.

4. Urban system and inter-city trade

Consider now our two-city setting in which workers are free to choose the city in which they want to live. Let λ be the endogenous share of workers residing in city 1. A *global equilibrium* arises at $0 < \lambda^* < 1$ when the utility differential $\Delta V(\lambda^*) \equiv V_1(\lambda^*) - V_2(\lambda^*) = 0$, or at $\lambda^* = 1$ when $\Delta V(1) \geq 0$. An interior equilibrium is stable if and only if the slope of the indirect utility differential ΔV is strictly negative in a neighborhood of the equilibrium, i.e., $d\Delta V(\lambda)/d\lambda < 0$ at λ^* , whereas an agglomerated equilibrium is stable whenever it exists. What makes our analysis richer, but more complex, is the fact that *the utility differential ΔV is not uniquely defined in that it varies with the internal structure of cities*. Consequently, when we deal with a configuration involving only monocentric (respectively polycentric) cities, we must consider the possibility of a deviation toward a polycentric (respectively monocentric) city when studying the existence and stability of an equilibrium.

The indirect utility of an individual working in the CBD is given by

$$V^C(\lambda) = S^* + w^{C*} - C^C + \bar{q}_0 \tag{22}$$

where S^* is the consumer surplus given by (13) and C^C the urban costs borne by this individual. Using (20), it is readily verified that

$$C^C \equiv R^C + tx = t \frac{\theta^* \lambda L}{2}. \tag{23}$$

If she works in the SBD (if any), her indirect utility becomes

$$V^S(\lambda) = S^* + w^{S^*} - C^S + \bar{q}_0$$

where C^S now denotes the urban costs the individual bears. Using (21), we have

$$C^S \equiv R^S + t|x - x^S| = t \frac{(1 - \theta^*) \lambda L}{4}. \tag{24}$$

Two comments are in order. First, the equilibrium allocation of workers within each city depends on the global distribution of workers between cities through the value of λ . In particular, workers are distributed at the city equilibrium in a way such that

$$V^C(\lambda) = V^S(\lambda).$$

Likewise, when λn firms are established in city 1, firms are distributed at the city equilibrium such that $\Pi^C(\lambda) = \Pi^S(\lambda) = 0$ when SBDs exist. Second, when deciding whether or not to move from one city to the other, workers know whether the cities of origin and destination are monocentric and/or polycentric; they also know the land rent that prevails in each one of them.

In order to determine the stable configurations, we define two critical values of t by replacing l by λL and $(1 - \lambda)L$, respectively, in (19):

$$t_1 \equiv \frac{2K}{\phi \lambda L} + \frac{k}{\phi}, \quad t_2 \equiv \frac{2K}{\phi(1 - \lambda)L} + \frac{k}{\phi}. \tag{25}$$

Without loss of generality, we may assume that $\lambda \geq 1/2$, so that $t_1 \leq t_2$. Using Proposition 1, it is easily seen that the following three global patterns may emerge:

- (i) when $t < t_1$, both cities are monocentric,
- (ii) when $t_1 < t < t_2$, city 1 is polycentric and city 2 is monocentric, and
- (iii) when $t_2 < t$, both cities are polycentric.

Hence, under dispersion ($\lambda = 1/2$), we have $t_1 = t_2 = T_D$ where

$$T_D \equiv \frac{4K}{\phi} L + \frac{k}{\phi}$$

so that the two cities are monocentric if $t < T_D$ and polycentric if $t > T_D$. Similarly, under agglomeration ($\lambda = 1$), $t_1 = T_A$ where

$$T_A \equiv \frac{2K}{\phi} L + \frac{k}{\phi} < T_D$$

and $t_2 \rightarrow \infty$; thus, agglomeration arises in a monocentric city when $t < T_A$ or in a polycentric city when $t > T_A$.

In order to determine what a global equilibrium is, we must consider the utility differential corresponding to each of the three foregoing patterns. In what follows, we study the cases in which

- (i) no city is polycentric,
- (ii) one city is polycentric and the other monocentric, and
- (iii) no city is monocentric.

Because one region may be empty in the first and third cases, our analysis will allow us to identify all stable equilibria.

4.1. The monocentric case

Assume that $t < t_1$ so that no city is polycentric: $\theta^* = 1$ for all $\lambda \in [1/2, 1]$. As mentioned above, the equilibrium wages are given by a bidding process in which firms compete for workers by offering them higher wages until no firm can earn positive profits, given by (4), in the CBD of either city. Using (13) and the equilibrium wages given in Appendix A, it is readily verified that the utility differential with two monocentric cities (with subscript mm) is as follows (up to a positive and constant factor):

$$\Delta_{mm} V(\lambda) \equiv \delta_{mm}(\lambda - 1/2) \quad (26)$$

where

$$\delta_{mm} \equiv -\varepsilon_1 \tau^2 + \varepsilon_2 \tau - \varepsilon_3 t$$

with

$$\varepsilon_1 \equiv (b\phi + cL)(6b^2\phi^2 + 6b\phi cL + c^2L^2) > 0,$$

$$\varepsilon_2 \equiv 4a\phi(b\phi + cL)(3b\phi + 2cL) > 0,$$

$$\varepsilon_3 \equiv 2(2b\phi + cL)^2\phi^2 > 0.$$

When $t < T_D$, Proposition 1 implies that $\lambda = 1/2$ with two monocentric cities is feasible. Note that the condition $t < T_D$ also prevents a marginal deviation to a polycentric city to occur because, in the vicinity of $\lambda = 1/2$, the city in region 2 must be monocentric. The dispersed configuration with two monocentric cities is, therefore, a global equilibrium when $t < T_D$. It is stable if $\delta_{mm} < 0$ or, equivalently, $t > t_m$ with

$$t_m \equiv \frac{(-\varepsilon_1 \tau + \varepsilon_2) \tau}{\varepsilon_3}$$

which is positive for all admissible values of τ because $\tau_{\text{trade}} < \varepsilon_2/\varepsilon_1$. Thus, $\lambda^* = 1/2$ is stable when $t > t_m$ and $t < T_D$.

When $t < t_m$, dispersion between two identical monocentric cities is no longer stable. If $t < T_A$, $\lambda = 1$ with a monocentric city is feasible. Because $\Delta_{mm} V(1) > 0$ when $t < t_m$ whereas the city in region 2 is always monocentric in the vicinity of $\lambda = 1$, $\lambda^* = 1$ is a stable equilibrium when $t < t_m$ and $t < T_A$.

The discussion above may be summarized as follows.

Proposition 2. *If $t < t_m$ and $t < T_A$, there exists a stable global equilibrium in which the industry is agglomerated into a single monocentric city. If $t_m < t < T_D$, there exists a stable global equilibrium in which the industry is dispersed between two monocentric cities of equal size.*

Note, first, that communication costs have to be sufficiently large to fulfill the condition $t < T_A$. If communication costs are very low (formally $K = k = 0$), the global economy never involves monocentric cities. When they are large, agglomeration in a monocentric city may occur provided that commuting costs are sufficiently low ($t < t_m$). Once commuting costs get larger ($t > t_m$), the industry is dispersed between two monocentric cities. Hence, a pattern involving

two symmetric and monocentric cities is more likely to emerge when both commuting and communication costs are high.

However, for the dispersed pattern with two monocentric cities to arise, it must be that $t_m < T_D$. As ε_2 increases with a whereas ε_1 , ε_3 and T_D are independent of a , this condition is satisfied when the parameter a does not exceed the unique solution a_m to the equation

$$\varepsilon_2\tau = \varepsilon_1\tau^2 + \varepsilon_3T_D.$$

In other words, the size of the differentiated product market cannot be too large, $a < a_m$, for two monocentric cities to be a global equilibrium.

Otherwise, when $a > a_m$ —hence $t_m > T_D$ —the industry is agglomerated in a single monocentric city as long as $t < T_A$: there is both *agglomeration* and *centralization*. This form of extreme agglomeration arises because the intensification of price competition that such a global structure brings about is itself compensated by a sufficiently large market size effect ($a > a_m$).

4.2. The mixed case

We now come to the case in which cities have different internal structures. Without loss of generality, we assume that city 1 is polycentric whereas city 2 is monocentric: $\theta_1^* < 1$ and $\theta_2^* = 1$ or, equivalently, $t_1 < t < t_2$ where both t_1 and t_2 are unknown. Note that, when this condition holds, it implies $\lambda > 1/2$ so that the industry is split between the two regions in a way such that the polycentric city hosts the majority of workers.

Using the equilibrium wages given in Appendix A, it is readily verified that the equation of motion—with subscript pm —is now given (up to a positive and constant factor) by

$$\Delta_{pm} V(\lambda) \equiv \delta_1\lambda + \delta_2 = \delta_1 \left(\lambda - \frac{\delta_2}{-\delta_1} \right)$$

with

$$\delta_1 \equiv -\varepsilon_1\tau^2 + \varepsilon_2\tau - \varepsilon_3 \frac{2t^2\phi}{3t\phi - k}, \quad \delta_2 \equiv \frac{1}{2} \left[\varepsilon_1\tau^2 - \varepsilon_2\tau + \varepsilon_3t \left(1 - \frac{4K/L}{3t\phi - k} \right) \right].$$

Since $\Delta_{pm} V(\lambda)$ is linear, the intermediate value theorem implies that the interval $(1/2, 1)$ contains a unique equilibrium, given by $\lambda^* = -\delta_2/\delta_1 \in (1/2, 1)$, if and only if $\Delta_{pm} V(1/2)$ and $\Delta_{pm} V(1)$ have opposite signs. Furthermore, this equilibrium is stable when the two inequalities $\Delta_{pm} V(1/2) > 0$ and $\Delta_{pm} V(1) < 0$ hold because $\Delta_{pm} V(\lambda)$ is then monotone decreasing. This in turn implies that $\delta_1 < 0$. The first condition ($\Delta_{pm} V(1/2) > 0$) is equivalent to $t > T_D$, whereas the second ($\Delta_{pm} V(1) < 0$) amounts to

$$-\varepsilon_1\tau^2 + \varepsilon_2\tau - \varepsilon_3 \frac{t(t\phi + k + 4K/L)}{3t\phi - k} < 0$$

or

$$K > \bar{K} \equiv \frac{L}{4} \left[\frac{(-\varepsilon_1\tau^2 + \varepsilon_2\tau)(3t\phi - k)}{\varepsilon_3t} - t\phi - k \right].$$

Note that city 1 is polycentric and city 2 monocentric since the two conditions $\Delta_{pm} V(1/2) < 0 < \Delta_{pm} V(1)$ are met.

To summarize,

Proposition 3. *If $t > T_D$ and $K > \bar{K}$, there exists a stable global equilibrium in which the industry is split between a large polycentric city and a small monocentric city.*

4.3. The polycentric case

It remains to consider the case where $t > t_2$ so that no city is monocentric (with subscript pp): $\theta^* < 1$ for all $\lambda \in [1/2, 1]$. Using the equilibrium wages given in Appendix A, the equation of motion becomes

$$\Delta_{pp}V(\lambda) = \delta_{pp}(\lambda - 1/2) \quad (27)$$

where

$$\delta_{pp} \equiv -\varepsilon_1\tau^2 + \varepsilon_2\tau - \varepsilon_3 \frac{t(t\phi + k)}{3t\phi - k} = \frac{4\varepsilon_3t\bar{K}}{(3t\phi - k)L}. \quad (28)$$

As δ_{mm} and δ_{pp} differ only through the last term, the argument is similar to the one developed in Section 4.1.

When $t > T_D$, Proposition 1 implies $\lambda = 1/2$ with two polycentric cities is feasible. Note that $t > T_D$ also prevents a marginal deviation to a monocentric city to occur because the city in region 2 must be polycentric in the vicinity of $\lambda = 1/2$. Therefore, the dispersed configuration with two polycentric cities is a global equilibrium when $t > T_D$. It is stable if $\delta_{pp} < 0$ or, equivalently, $t > t_p$, which is the larger solution of $\delta_{pp} = 0$. This is because the smaller solution violates the condition $k < 3t\phi$. Note that t_p is positive for all admissible value of τ because $\tau_{\text{trade}} < \varepsilon_2/\varepsilon_1$. Thus, $\lambda^* = 1/2$ is stable if $t > t_p$ and $t > T_D$.

When $t < t_p$, dispersion between two identical polycentric cities ceases to be stable. If $t > T_A$, agglomeration within a polycentric city is feasible. Because $\Delta_{pp}V(1) > 0$ when $t > t_p$, $\lambda^* = 1$ is a stable equilibrium as long as $\Delta_{pm}V(1) > 0$, which is itself equivalent to $K < \bar{K}$.

To summarize,

Proposition 4. *If $t > T_A$ and $K < \bar{K}$, there exists a stable global equilibrium in which the industry is agglomerated into a single polycentric city. If $t > T_D$ and $t > t_p$, there exists a stable global equilibrium in which the industry is dispersed between two polycentric cities of equal size.*

Hence, high distance-sensitive communication costs can prevent the emergence of a single polycentric city because $\bar{K} < 0$, even when commuting costs are large ($t > T_A$). In addition, for agglomeration in a polycentric city to occur, it must be that $t > T_A$ and $K < \bar{K}$. It is easy to show that the latter condition holds if and only if $a > a_p$, where a_p is the unique solution to the equation

$$\varepsilon_2\tau = \varepsilon_1\tau^2 + \varepsilon_3 \frac{t(t\phi + k + 4K/L)}{3t\phi - k}.$$

Consequently, when the market size effect is strong ($a > a_p$), agglomeration within a polycentric city takes place. Because a_p is smaller than a_m , agglomeration is sustained for weaker market size effects in one polycentric city than in one monocentric city. This shows once more that dispersion is a less likely outcome when the size of the differentiated product market is large.

When commuting costs are large ($t > t_p$) and communication costs low (T_D is small), workers alleviate the burden of urban costs by having two polycentric cities: there is both *dispersion* and *decentralization*. In other words, the global organization of the production is associated with the lowest level of urban costs in the economy. Observe that a strong reduction in trade costs leads to low values of t_p , thus fostering the dispersion of the industry, which now takes the form of two polycentric cities. Because t_p exceeds t_m , agglomeration is sustainable over a larger set of t -values in the polycentric case than in the monocentric case.

Note, finally, that the configuration involving one polycentric city and one monocentric city is sustainable over a larger set of parameter values than the dispersed configuration with two polycentric cities. The reason is as follows. When a dispersed configuration with two polycentric cities is a stable equilibrium, the conditions $t > T_D$ and $t > t_p$ must hold. We know that $t > t_p$ is equivalent to $\delta_{pp} < 0$. If $\delta_{pp} < 0$, then $\bar{K} < 0$ also holds by (28). This implies the condition $K > \bar{K}$ is always met, so that Proposition 3 holds. In other words, partial agglomeration is also a stable outcome.

4.4. Summary results

Propositions 2 to 4 reveal the existence of five stable configurations:

- (A) a single monocentric city ($\lambda^* = 1$ and $\theta^* = 1$), denoted $(m, 0)$;
- (B) two identical monocentric cities ($\lambda^* = 1/2$ and $\theta^* = 1$), denoted (m, m) ;
- (C) a single polycentric city ($\lambda^* = 1$ and $\theta^* < 1$), denoted $(p, 0)$;
- (D) two identical polycentric cities ($\lambda^* = 1/2$ and $\theta^* < 1$), denoted (p, p) ; and
- (E) one large polycentric city and one small monocentric city ($\lambda^* \in (1/2, 1)$, $\theta_1^* < 1$ and $\theta_2^* = 1$), denoted by (p, m) .

Those five equilibria are illustrated in the (K, t) - and (τ, t) -planes in Figs. 2 and 3, respectively. It is readily verified that, in both figures, the positive orthant is fully covered by the set of conditions identified through Propositions 2 to 4. However, this covering does not define a partition because of the multiplicity of stable equilibria.

Interestingly, a partial agglomeration may emerge as an equilibrium outcome once it is recognized that SBDs can exist, while such a configuration never arises in standard NEG-models (Ottaviano and Thisse [20]). Another interesting feature of the partial agglomeration pattern is that it involves intra-industry and asymmetric trade in the differentiated product together with intersectoral trade.

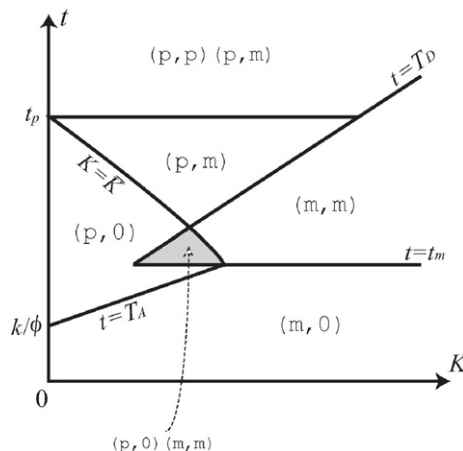


Fig. 2. Global equilibria in (K, t) -space ($a = 8, b = 3, c = 1, \phi = 1, L = 100, k = 1, \tau = \tau_{\text{trade}}/2$).

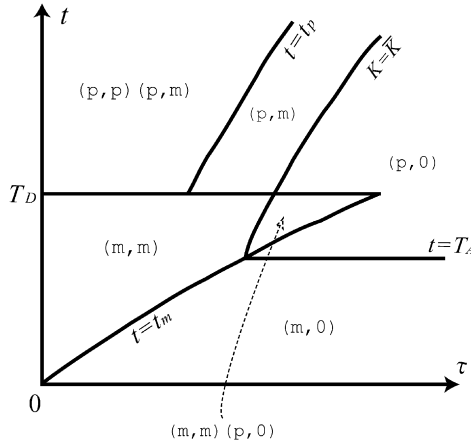


Fig. 3. Global equilibria in (τ, t) -space ($a = 8, b = 3, c = 1, \phi = 1, L = 100, k = 1, K = 50$).

5. Interaction between local and global forces

In this section, we study the impact of changes in commuting costs, communication costs and trade costs on the location of firms within and between cities. Our results are organized around the following two ideas:

- (i) local factors may well change the global organization of the economy, whereas
- (ii) global forces may affect the local/urban organization of production and employment.

From the historical point of view, we find it interesting to consider two cases. We first study how falling communication and commuting costs affects the space-economy. Next, we will consider the usual thought experiment of NEG, namely the impact of falling trade costs on the global distribution of firms and workers. The difference is that the structure of cities now depends on the evolution of the interregional distribution of activities, which affects itself the way cities are organized.

5.1. How the local affects the global

5.1.1. The impact of communication costs

Consider the case of communication costs decreasing from high to low values. For conciseness, we focus on the sole fixed component K . As shown by Fig. 2, which is drawn for the case where $t_m > k/\phi$, the evolution of the space-economy changes according to the value of commuting costs. Three cases appear to be relevant.

Assume, first, that t exceeds t_p . When K is large, the economy involves two monocentric cities because commuting costs are too large to afford the agglomeration of activities within a single city. Once K decreases and crosses the line $t = T_D$, the economy may follow two very different paths. Along the first one, dispersion still prevails but the two monocentric cities are now polycentric. This is because communication costs have sufficiently decreased to permit the decentralization of some activities within each city. However, SBDs remain small when K has not decreased by a sufficiently large amount. By contrast, when K is very small, each city is

formed by three areas having almost the same size. Between these two polar cases, as K decreases the CBD of each city shrinks whereas its SBDs grow, the whole process being smooth (see Proposition 1). Along the second path, there is a bifurcation at the crossing point as revealed by the emergence of a polycentric city whose additional population comes from the other city, which then remains monocentric. The economic space becomes asymmetric and, as communication costs decrease, the large city grows at the expense of the small city. In this case, falling communication costs increase the size of the polycentric city ($d\lambda^*/dK < 0$). This is because sufficiently low values of K trigger the decentralization of production. This in turn increases net wages

$$\frac{d(w_r^S - C_r^S)}{dK} = \frac{d(w_r^C - C_r^C)}{dK} = -\frac{2}{3\phi} < 0$$

and makes the agglomeration forces stronger. The entire process may then be summarized as follows:

$$(m, m) \rightarrow \begin{matrix} (p, p) & \text{(case 1.1)} \\ (p, m) & \text{(case 1.2)} \end{matrix}.$$

In the second case, we have $t_m < t < t_p$. Again, the economic space involves two monocentric cities when K is large. Ignoring for the moment the shaded area in which two stable equilibria exist, we see from Fig. 2 that, when K falls, city 1 expands gradually under the form of a polycentric structure, whereas the monocentric city 2 shrinks. Eventually, when K becomes smaller than \bar{K} , the large city accommodates all workers and city 2 vanishes. There is agglomeration within a single polycentric city. Again, as shown by Proposition 1, the size of city 1 increases gradually from $L/2$ to L . We thus have:

$$(m, m) \rightarrow (p, m) \rightarrow (p, 0) \quad \text{(case 2)}.$$

It remains to consider the third case in which $k/\phi < t < t_m$. For large values of K , the economy has a single monocentric city, which may correspond to an isolated “city-state”. When the line $t = T_A$ is crossed, the city-state ceases to be monocentric and two small SBDs are created. As K keeps decreasing, the city still retains the whole population of workers, but its CBD loose more and more activities at the benefit of its SBDs. Here also, Proposition 1 implies that the transition process is smooth. This is summarized as follows:

$$(m, 0) \rightarrow (p, 0) \quad \text{(case 3)}.$$

To sum up, when communication costs decrease, two contrasted pictures emerge. In the former (cases 1.1 and 3), the global organization of the economy remains the same ($\lambda^* = 1/2$ or $\lambda^* = 1$). The decrease in communications cost affects only the city structure, which becomes more decentralized. In the latter (cases 1.2 and 2), the reduction of communication costs leads to an imbalanced distribution of activities as well as to asymmetric trade ($1/2 < \lambda^* < 1$). In other words, the global economy, the structure of cities and the nature of trade are all affected. It should then be clear that any simple prediction made about the impact the development NICTs could have on the spatial organization of the economy is likely to be inaccurate.

Note also that in all, except one, cases, the spatial evolution of the economy is *smooth*, a result that vastly differs from what is known in standard NEG-models (Ottaviano and Thisse [20]). The only case that shows a strong discontinuity arises when the evolution path goes through the shaded area. Urban inertia suggests that the economy retains its previous (m, m) structure when

K and t belong to this area.⁹ When the path gets out of the shaded area, the economy shifted abruptly from dispersion with two monocentric cities to one polycentric city, that is, the most drastic transformation the spatial economy may experience.

The evolution of the urban landscape described above is to be contrasted to what happens when cities are, by assumption, monocentric. In this case, it follows from Fig. 2 that the global equilibrium is always given by $(m, 0)$ as long as t is smaller than t_m , whereas this equilibrium is given by (m, m) when t exceeds t_m . This shows that the (partial) agglomeration of activities may be sustained over a larger domain of parameters when cities can be polycentric.

5.1.2. The impact of commuting costs

Regarding commuting costs, the following remarks are worth making. First, whatever the value of K , agglomeration is always the single global equilibrium when commuting costs are low ($t < t_m$), but never when t is high ($t > t_p$). This may be explained as follows. When cities are monocentric, it is readily verified that

$$\left. \frac{dC^C}{dt} \right|_{\theta^*=1} > 0$$

where C^C is given by (23). When cities are polycentric, some tedious, but standard, calculations show that

$$\left. \frac{dC^j}{dt} \right|_{\theta^* \in [1/3, 1)} > 0 \quad j = C, S$$

where C^j is given by (24). Hence, regardless of the city structure, urban costs borne by workers decrease as commuting costs fall. Consequently, net wages increase regardless of workers' residential location. This implies that more workers are willing to choose to set up in a single city. This larger concentration of workers then makes the agglomeration forces stronger, which in turn increases workers' utility. Eventually, all workers end up living in the same city when commuting costs are sufficiently low.

Second, increasing commuting costs from low values of t does not necessarily induce the decentralization of production within cities. When communication costs are low, agglomeration still prevails and the city becomes polycentric instead of being monocentric. However, this ceases to be true when communication costs are large. In this case, workers get dispersed between two monocentric cities. Likewise, as shown by Fig. 2, a polycentric city may emerge even when commuting costs are low ($t < t_m$) provided that communication costs are low too. This is because low commuting costs foster agglomeration, thus raising, all else equal, land rents, and inducing, *in fine*, the decentralization of production within the same city.

Third, when communication costs are fairly low, a progressive decrease in commuting costs yields an interesting picture. Again, Fig. 2 will be our backbone. It shows that the economy follows the path described below:

$$(p, p) \rightarrow (p, m) \rightarrow (p, 0) \rightarrow (m, 0).$$

When commuting costs are large but communication costs low, the market forces reduce the burden of urban costs by dispersing workers across cities as well as employment in each city.

⁹ Urban inertia is strengthened by the durability of the housing stock, a variable not taken into account here (Glaeser and Gyourko [9]).

Below the threshold t_p , commuting costs can sustain a larger polycentric city, thus implying that city 2 becomes monocentric because its population is too low for this city to retain its polycentric structure. As t decreases further, city 1 keeps growing at the expense of the small one, which vanishes when t is sufficiently low. Indeed, both communication and commuting costs take low enough values to permit the agglomeration of activities within a single polycentric city. Finally, when commuting costs take extremely low values, city 1 becomes monocentric. This is because such a spatial arrangement allows firms and workers to save on communication costs as commuting is almost inexpensive.

Given the discussion above, we may safely conclude that *allowing for structural changes within cities affects both the global economy and the nature of trade.*

5.2. How the global affects the local

Figure 3 displays our main conditions in the (τ, t) -space. In what follows, we consider the case of high commuting costs ($t > T_D$). When trade costs are high, we see that the economy involves a single polycentric city. The fact that commuting and trade costs are high prevents both the existence of monocentric cities and the emergence of a dispersed pattern. Once trade costs cross the line $K = \bar{K}$, there is partial dispersion with the creation of a small city that allows workers to reduce the burden of urban costs in the large city. However, trade costs remain too large for that city to capture a large share of activities. If trade costs keep decreasing and cross the line $t = t_p$, two stable equilibria exist. In the first one, the small city remains monocentric but attracts a growing share of the labor force at the expense of the large city:

$$\frac{d(-\delta_2/\delta_1)}{d\tau} > 0 \quad \text{for all } \tau < \tau_{\text{trade}}.$$

In other words, as the global economy gets more integrated, less workers settle in the large city, the partial re-dispersion of the economic activity being gradual. Yet, *the large city keeps hosting a large share of mobile activities.* Stated differently, as trade costs keep decreasing, the regional gap shrinks but does not disappear because λ^* converges to

$$\frac{3}{4} - \frac{T_D}{t} \in \left(\frac{1}{2}, \frac{3}{4} \right)$$

when $\tau \rightarrow 0$. It is worth stressing the fact that this process, which is summarized below, is smooth in that it involves the progressive growth of the small city together with the gradual decline of the large one:

$$(p, 0) \rightarrow (p, m) \rightarrow \begin{pmatrix} p, m \\ p, p \end{pmatrix}.$$

In the second equilibrium, full dispersion with two polycentric cities prevails. Because trade costs are low enough, the dispersion of activities does not hamper inter-city trade, whereas high commuting costs leads to the decentralization of activities. Although there is multiplicity of equilibria here, urban inertia is likely to select the first equilibrium. In this case, the economy shifts from agglomeration (and no trade) to partial agglomeration (and asymmetric trade) when trade costs steadily decline. Such an outcome is reminiscent of the bell-shaped curve of spatial development—the sequence dispersion/agglomeration/re-dispersion—obtained in several NEG-models with urban costs (Ottaviano and Thisse [20]).

Again, it is worth comparing the first scheme of evolution to what happens when cities are assumed to be monocentric ($t < T_A$). In this case, Fig. 3 shows that agglomeration prevails below the line $t = t_m$; otherwise there is always dispersion. Therefore, when cities are allowed to become polycentric, *the larger city is able to maintain its primacy over a larger set of structural parameters*, thus preventing the complete re-dispersion of activities that would arise in the case of monocentric cities. Thus, there is a need to extend the analysis of urban systems to cope with polycentric cities, instead of focusing almost exclusively on monocentric cities as existing theories do (Abdel-Rahman and Anas [1]).

6. Concluding remarks

We have presented a simple model that uncovers how the interplay between different types of spatial friction affects the location of economic activities *between* and *within* cities. Historical evidence shows that both trade and commuting costs have been decreasing since the beginning of the Industrial Revolution. Ever since the end of the 19th century, the development of the new communication technologies has allowed firms to alleviate the burden of urban costs in large metropolitan areas, through the emergence of secondary employment centers. We have shown how these various technological changes have impacted on the way firms and workers locate. Our results agree with empirical evidence. Thus, we may safely conclude that what matters for the organization of the space-economy is the relative evolution of three types of costs: the commuting of workers, the transfer of information and the transport of commodities.

When cities are open to trade, the organization of the space-economy varies with the ability of cities to accommodate a small or a large population. In particular, our results show that, once agglomeration within a polycentric city has been achieved, the core maintains its primacy over a large range of trade cost values, thus confirming the idea that the polycentric structure fosters agglomeration. Our analysis also highlights the importance of local factors in the emergence of regional inequalities. For example, agglomeration is more likely to occur when commuting costs take very low values, but never arises when these costs are high.

The multiplicity of stable equilibria observed here has also an important implication that has been very much overlooked in the literature: *different types of spatial patterns may coexist under identical technological and economic conditions*. It should be no surprise, therefore, to observe a variety of urban systems in the real world.

Finally, our analysis also has an important policy implication. It is well known that quite a few American and European cities have lost employment and population for a few decades. This state of affair has led city managers to seek local policies that would prevent the decentralization of industrial activities toward small and/or remote places. However, most of them did fail. This paper suggests that urban decision-makers would have been better inspired *to foster the development of SBDs*, endowed with high-quality business-to-business services and consumption amenities that attract both firms and workers.

Acknowledgments

We wish to thank J. Brueckner and two referees for very useful comments and suggestions. We are also grateful to K. Behrens, V. Henderson, T. Mayer, and A. Venables for stimulating discussions.

Appendix A

The corresponding equilibrium wages (w_r^C and w_r^S for $r = 1, 2$) are such that all firms, located either in the CBD or in the SBD of each city, earn zero profits (given, respectively, by Eqs. (4) and (5)).

When city 1 is polycentric whereas city 2 is monocentric, the equilibrium wages (w_1^C , w_1^S and w_2^C) are such that no firm established in city 1 and located either in the CBD or in the SBD of this city, or established in the CBD of city 2 is able to make positive profits. More precisely, the equilibrium wages in city 1 are given by

$$w_1^C(\lambda) = \frac{I_1(\lambda)}{\phi}, \quad w_1^S(\lambda) = \frac{I_1(\lambda) - \mathcal{K}}{\phi} \quad (29)$$

where

$$I_1(\lambda) = \frac{(b\phi + cL)L}{4(2b\phi + cL)^2\phi} \left\{ [2a\phi + \tau cL(1 - \lambda)]^2 \lambda + [2a\phi - 2\tau b\phi - \tau cL(1 - \lambda)]^2 (1 - \lambda) \right\}$$

and the equilibrium wage in city 2 by

$$w_2^C(\lambda) = \frac{I_2(\lambda)}{\phi}$$

where

$$I_2(\lambda) = \frac{(b\phi + cL)L}{4(2b\phi + cL)^2\phi} \left\{ (2a\phi + \tau cL\lambda)^2 (1 - \lambda) + (2a\phi - 2\tau b\phi - \tau cL\lambda)^2 \lambda \right\}.$$

When both cities are polycentric, wages $w_r^C(\lambda)$ and $w_r^S(\lambda)$ are given by (29) for $r = 1, 2$.

References

- [1] H.M. Abdel-Rahman, A. Anas, Theories of systems of cities, in: J.V. Henderson, J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, 2004, pp. 2293–2339.
- [2] A. Anas, R. Arnott, K.A. Small, Urban spatial structure, *Journal of Economic Literature* 36 (1998) 1426–1464.
- [3] J. Cavailhès, C. Gaigné, J.-F. Thisse, Trade costs versus urban costs, Discussion paper No. 4440, CEPR, 2004.
- [4] R. Cerveso, K.L. Wu, Commuting and residential location in the San Francisco Bay area, *Environment and Planning A* 29 (1997) 865–886.
- [5] G. Duranton, D. Puga, Micro-foundations of urban increasing returns: Theory, in: J.V. Henderson, J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, 2004, pp. 2063–2117.
- [6] C. Engel, J. Rogers, How wide is the border? *American Economic Review* 86 (1996) 1112–1225.
- [7] M. Fujita, P. Krugman, T. Mori, On the evolution of hierarchical urban systems, *European Economic Review* 43 (1999) 209–251.
- [8] G. Giuliano, A. Small, Subcenters in the Los Angeles region, *Regional Science and Urban Economics* 21 (1991) 163–182.
- [9] E.L. Glaeser, J. Gyourko, Urban decline and durable housing, *Journal of Political Economy* 113 (2005) 345–375.
- [10] E.L. Glaeser, M.E. Kahn, Sprawl and urban growth, in: J.V. Henderson, J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, 2004, pp. 2481–2527.
- [11] M.L. Greenhut, Spatial pricing in the United States, West Germany and Japan, *Economica* 48 (1981) 79–86.
- [12] J. Haskel, H. Wolf, The law of one price—A case study, *Scandinavian Journal of Economics* 103 (2001) 545–558.
- [13] E. Helpman, The size of regions, in: D. Pines, E. Sadka, I. Zilcha (Eds.), *Topics in Public Economics. Theoretical and Applied Analysis*, Cambridge Univ. Press, Cambridge, 1998, pp. 33–54.
- [14] J.V. Henderson, The sizes and types of cities, *American Economic Review* 64 (1974) 640–656.

- [15] V. Henderson, A. Mitra, New urban landscape: Developers and edge cities, *Regional Science and Urban Economics* 26 (1996) 613–643.
- [16] D.P. McMillen, J.F. McDonald, Population density in suburban Chicago: A bid-rent approach, *Urban Studies* 35 (1998) 1119–1130.
- [17] D.P. MacMillen, S. Smith, The number of subcenters in large urban areas, *Journal of Urban Economics* 53 (2003) 321–338.
- [18] H. Ogawa, M. Fujita, Equilibrium land use patterns in a non-monocentric city, *Journal of Regional Science* 20 (1980) 455–475.
- [19] G.I.P. Ottaviano, T. Tabuchi, J.-F. Thisse, Agglomeration and trade revisited, *International Economic Review* 43 (2002) 409–436.
- [20] G.I.P. Ottaviano, J.-F. Thisse, Agglomeration and economic geography, in: J.V. Henderson, J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, 2004, pp. 2563–2608.
- [21] M.E. Porter, Competitive advantage of the inner city, *Harvard Business Review* (1995) 55–71, May–June.
- [22] A. Schwartz, Subserving suburbia, *Journal of the American Planning Association* 59 (1993) 288–305.
- [23] T. Tabuchi, Urban agglomeration and dispersion: a synthesis of Alonso and Krugman, *Journal of Urban Economics* 44 (1998) 333–351.
- [24] T. Tabuchi, J.-F. Thisse, Regional specialization, urban hierarchy, and commuting costs, *International Economic Review* 47 (2006) 1295–1317.
- [25] D. Timothy, W.C. Wheaton, Intra-urban wage variation, employment location and commuting times, *Journal of Urban Economics* 50 (2001) 338–366.