

Thisse, Jacques-François

**Working Paper**

## How transport costs shape the spatial pattern of economic activity

OECD/ITF Joint Transport Research Centre Discussion Paper, No. 2009-13

**Provided in Cooperation with:**

International Transport Forum (ITF), OECD

*Suggested Citation:* Thisse, Jacques-François (2009) : How transport costs shape the spatial pattern of economic activity, OECD/ITF Joint Transport Research Centre Discussion Paper, No. 2009-13, Organisation for Economic Co-operation and Development (OECD), Joint Transport Research Centre (JTRC), Paris,  
<https://doi.org/10.1787/5kmmr3j65tg0-en>

This Version is available at:

<https://hdl.handle.net/10419/68751>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



JOINT TRANSPORT RESEARCH CENTRE

*Discussion Paper No. 2009-13  
December 2009*

# *How Transport Costs Shape the Spatial Pattern of Economic Activity*

Jacques-François THISSE  
CORE, Université catholique de Louvain, Belgium  
and École Nationale des Ponts et Chaussées, Paris

## HOW TRANSPORT COSTS SHAPE THE SPATIAL PATTERN OF ECONOMIC ACTIVITY

Jacques-François Thisse  
CORE, Université catholique de Louvain and Ecole nationale des Ponts et Chaussées

### SUMMARY

1. INTRODUCTION.....	3
2. THE TRADE-OFF BETWEEN INCREASING RETURNS AND TRANSPORT COSTS.....	5
2.1 The optimal number and size of firms .....	5
2.2 The optimal location of a firm .....	6
3. THE MOBILITY OF FIRMS AND WORKERS.....	8
3.1 The home-market effect .....	8
3.2 The emergence of a core-periphery structure.....	10
4. THE BELL-SHAPED CURVE OF SPATIAL DEVELOPMENT.....	13
4.1 Vertical linkages .....	13
4.2 Imperfect labour mobility .....	14
4.3 The spatial fragmentation of firms.....	15
5. THE TRADE-OFF BETWEEN COMMUTING COSTS WITHIN THE CITY AND TRANSPORT COSTS BETWEEN CITIES .....	16
5.1 The monocentric city .....	17
5.2 The polycentric city .....	18
6. CONCLUDING REMARKS .....	19
REFERENCES.....	23

## 1. INTRODUCTION

By its very nature, transport is linked to trade. Trade being one of the oldest human activities, the transport of commodities is, therefore, a fundamental ingredient of any society. People get involved in trade because they want to consume goods that are not produced within reach. The Silk Road provides evidence that shipping high-valued goods over long distances has been undertaken because of this very precise reason. But why is it that not all goods are produced everywhere? The reason is that regions are specialized in the production of certain products. The first explanation for *specialization* that comes to mind is that nature supplies specific environments needed to produce particular goods. According to Diamond (1997), spatial differences in edible plants, with abundant nutrients, and wild animals, capable of being domesticated to help man in his agricultural and transport activities, explain why only a few regions have become independent centers of food production. Though relevant for explaining the emergence of civilization in a few areas, we must go further to understand why, in the wake of the Industrial Revolution, interregional and international trade has grown so rapidly.

Goods are not ubiquitous because regions are endowed with a *comparative advantage*. Specifically, this advantage stems from the ability of a region to supply a particular good at a lower opportunity cost than other regions, sometimes because its inhabitants have learned how to produce it by means of technologies unknown to others. Spatial heterogeneities among regions, such as the uneven distribution of immobile resources (natural harbors) and amenities (climate), as well as differences in the access to major transshipment points (e.g. the Great Lakes in Canada and the United States), may also be at the origin of a variety of comparative advantages. Each region thus specializes in the production of goods for which it has a comparative advantage and trades with regions specialized in the production of other goods. However, the existence of transport costs renders a whole range of goods for which neither region has a sufficiently important productivity advantage non tradable. In other words, the production cost advantage is not sufficient to overcome the disadvantage linked to the value of transport costs. As the magnitude of transport costs decreases, the range of tradable goods widens. Even though exogenous comparative advantages are important, it is my belief that they cannot by themselves explain the formation of big agglomerations and large trade flows across regions and countries. Furthermore, some of these heterogeneities (think of the supply of transport infrastructure) are not given by nature and should be treated as being endogenous.

Modern trade theory has underscored the fact that specialization may also be the outcome of activities displaying *increasing returns* (Helpman and Krugman, 1985). To understand how this works, it is important to recognize that increasing returns may arise for a variety of reasons. First of all, scale economies are said to be *internal* to firms when the productive efficiency of firms increases with the size of their output. One major reason for this is that firms are able to adopt more efficient technologies once their sizes have reached a minimum threshold. Firms may also increase their productivity through learning-by-doing economies that emerge over the production process itself. Less known, perhaps, is the concept of scale economies *external* to firms whose origin lies in the socio-economic structure of their close environment (Duranton and Puga, 2004). This includes a wide range of factors such as the access to specialized business-to-business services, the formation of a specialized labour force, the production of new ideas, based on the accumulation of human capital and face-to-face communications, and the

availability of efficient and specialized infrastructure. Scale economies are the prime driver in the formation of cities where the division of labour and the specialization of tasks reach a level impossible to achieve with a dispersed population (Fujita and Thisse, 2002). It should then be clear that regions and cities get specialized in the production of specific goods because of the cost advantage generated by increasing returns, either internal or external to firms. Transport costs remain an impediment to trade, but market size matters here. Indeed, the existence of large local markets may overcome high transport costs through low average production costs.

Thus, we may safely conclude that the demand for the transport of commodities stems from the need to trade, which itself comes from the productive specialization of regions. All distance-related costs having dramatically decreased with technological advances in transportation and the development of the new communication technologies, it is easy to figure out why trade has grown at a fast pace. In addition, new and cheaper transport means impact on the location of firms and households. By changing the accessibility to input and output markets, lower transport and communication costs give them incentives to relocate. Therefore, it is legitimate to ask the question: *what is the impact of falling transport and communication costs on the location of economic activity?*

In order to say something relevant about the way a spatial economy is organized, it is necessary to assume that the production of goods involves increasing returns. If returns to scale are constant, allowing for the mobility of households and firms has a weird implication: all locations have the same relative prices and the same production structure. Indeed, in a world where activity can operate at arbitrarily small levels without efficiency losses, firms and households may reduce transport expenditures to zero by dispersing their activity across space. Every region then becomes an autarky, as it only needs to produce for its own domestic market. Hence, the standard economic paradigm combining constant returns and perfect competition is unable to account for the emergence and growth of big economic agglomerations and the existence of large shipments of goods.

Thus, the presence of increasing returns has a fundamental implication for the spatial structure of the economy: *not everything can be produced everywhere*. Therefore, it is no surprise that, in many real-world situations involving the location of large equipments, decision-makers face a trade-off between global efficiency and spatial equity (e.g. the proliferation of transport facilities is often the consequence of policies that put too much weight on spatial equity). Increasing returns have another major implication for the space-economy: lower transport costs may amplify or reduce the geographical advantage and disadvantage held by particular regions. Or, to put it differently, *a small exogenous comparative (dis)advantage can become a large endogenous comparative (dis)advantage*.

That said, what drives the location of firms and consumers is the existence of spatially dispersed markets. Accessibility is measured by all the costs generated by the various types of spatial frictions that economic agents face in the exchange process. Hence, it should be clear that *the way the space-economy is organized depends on the mutual interactions between mobility costs and scale economies*, the specification of which varies with the spatial scale (the world, the country or the city). In my opinion, the opportunity of developing interurban passenger transport must be evaluated within this framework because it strongly affects the type of mobility across cities that highly-skilled workers may choose.

The purpose of this paper is to discuss some of the main trade-offs at work at different spatial scales. Needless to say, within the format of this paper, I can cover only a few of the main ideas developed in economic geography and urban economics. The emphasis will be on the impact that falling transport costs have on microeconomic decisions on, and the resulting aggregate outcomes of, the location of firms and workers.

## 2. THE TRADE-OFF BETWEEN INCREASING RETURNS AND TRANSPORT COSTS

### 2.1 The optimal number and size of firms

The Industrial Revolution brought dramatically low transport costs as well as a huge increase in the size of production plants. The very first industrial plants had a very small optimal size. Indeed, as observed by Bairoch (1997): “In most manufacturing sectors, it was possible for a firm to have a competitive position with a very small size. The narrowness of the market, due to high transport costs, made it even easier to operate at a very low scale.” Things changed after the first half of the nineteenth century. The minimal size of a firm grew because of the use of increasingly diversified equipment, which then required many more workers. This growth in the size of firms was sustained by the expansion of markets areas, which in turn was possible because of the strong decline in transport costs. In brief, the interactions between these changes led to a gradual reduction in the number of firms, whose size increased. Take, for example, the case of Belgian steel enterprises: while their average workforce in 1845 was 26 people, it reached 446 people in 1930 (Bairoch, 1997). Hence, it is no surprise that the trade-off between increasing returns and transport costs is at the heart of location theory.

The trade-off between these two forces is easy to understand. First, as mentioned above, in the absence of increasing returns, one plant could be built in each consumption place so that there would be nothing to ship. Moreover, in the absence of transport costs, a single plant would be enough to satisfy the entire demand (except for the case where its marginal cost of production would increase). When transport costs increase with distance, this is formally equivalent to the case in which a fixed cost coexists with a growing marginal cost. Each plant supplies consumers located within a certain radius, the length of which depends on the relative level of the transport costs and the intensity of increasing returns, but those located beyond this radius are supplied by another unit.

The nature of this trade-off can be illustrated by considering the simple case of three spatially separated markets, **W**(est), **C**(enter) and **E**(ast), where the local demand for a given good is perfectly inelastic and normalized to **1**. Building one facility in a market requires **F** euro, while shipping one unit of the good between any two adjacent markets is equal to **T** euro. It is readily verified that the choice is between the following two options. First, building a facility in each market generates a total cost equal to **3F** since there is no shipping. Second, when a single facility is built, the optimal location is **C** and the corresponding cost **F + 2T**. The cost-minimizing solution, then, is to have a single facility if and only if

$$\mathbf{F + 2T < 3F \Leftrightarrow T < F.}$$

This inequality holds when **F** is high and **T** is low. Otherwise, it is optimal to have three facilities. This example is enough to understand that, on the one hand, high fixed costs favour the concentration of production in a small number of large units, as in modern developed economies; while, on the other hand, the situation in which high transport costs encourage the proliferation of small settlements across space characterizes preindustrial economies. Despite its simplicity, this example illustrates a very general principle: *strong scale economies in production* (large **F**), *low transport costs of commodities* (small **T**), or both foster the agglomeration of economic activities in a small number of areas.

By modifying slightly the example, it is possible to uncover another major principle of economic geography. Specifically, we assume that the common demand for the good is shifted upward from  $\mathbf{1}$  to  $\mathbf{D}$  units. The above inequality then becomes

$$\mathbf{F} + 2\mathbf{DT} < 3\mathbf{F} \Leftrightarrow \mathbf{DT} < \mathbf{F}.$$

Clearly, this ceases to hold when  $\mathbf{D}$  is sufficiently large. Hence, when local markets are large (large  $\mathbf{D}$ ), it is optimal to supply each of them from a facility set up there. In other words, even when unit transport costs are low (small  $\mathbf{T}$ ), *the proximity to large markets matters for the location of firms*.

## 2.2 The optimal location of a firm

The simplest firm-location problem is the one in which the firm, which cannot be subdivided in smaller units because of increasing returns, buys one input in one market ( $\mathbf{W}$ ) and sells its output in another ( $\mathbf{E}$ ), with a link connecting the two markets. The optimal location of the firm, which minimizes the sum of transport costs, can be viewed as the equilibrium point of a system governed by two forces generated by the need for proximity to the product market and the factor market. The intensity of these two forces depends, on the one hand, on the quantities shipped ( $\mathbf{w}_1 > \mathbf{w}_2$ ) and, on the other, on the marginal cost of transport with respect to distance.

Assuming that input and output are shipped by means of the same transport mode, the value of the elasticity of the unit transport cost function  $\mathbf{T}$  with respect to distance is an indicator of the degree of increasing returns in transportation. More precisely, a high value of this elasticity means that making the movement slightly longer increases its cost greatly. In this case, the value of transport costs is determined mainly by the distance covered when shipping goods. Such a situation describes quite well periods in which moving commodities was both dangerous and difficult, thus necessitating coaching inns for ground transport and coastal navigation for maritime transport. On the contrary, a low elasticity implies that the share of transport costs due to investments in infrastructure and equipment grows, so that distance matters less. Clearly, such a situation is characteristic of modern economies.

To start with, assume that the elasticity of the transport cost  $\mathbf{T}$  is larger than  $\mathbf{1}$ . In that case, the intensity of the pulling forces increases rapidly with distance, as illustrated in Figure 1a. Consequently, the system of forces is in equilibrium when the firm chooses the location where the marginal transport costs with respect to distance are equal: increasing the length of a trip is so costly that it is desirable for the firm to reduce the distance to the market with the higher marginal cost. This is why a place located in between the two markets is cost-minimizing. If the elasticity decreases to reach a value equal to  $\mathbf{1}$ , the firm chooses to establish itself in the market with the highest weight (see Figure 1b where the bold line takes its lowest value at  $\mathbf{W}$  since  $\mathbf{w}_1 > \mathbf{w}_2$ ). Because the intensity of the forces is now independent of the distances to the input and output markets, every intermediary location becomes suboptimal. This also holds when elasticity takes on values less than  $\mathbf{1}$ , as the marginal cost of transport decreases with distance.

Figure 1.

Figure 1a.

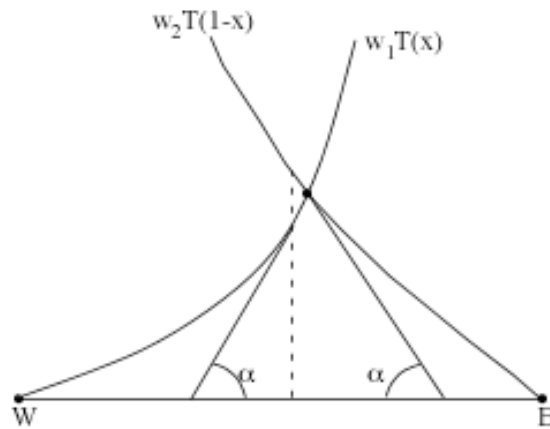


Figure 1.a:  $w_1T'(x) = -w_2T'(1-x) = \alpha$

Figure 1b.

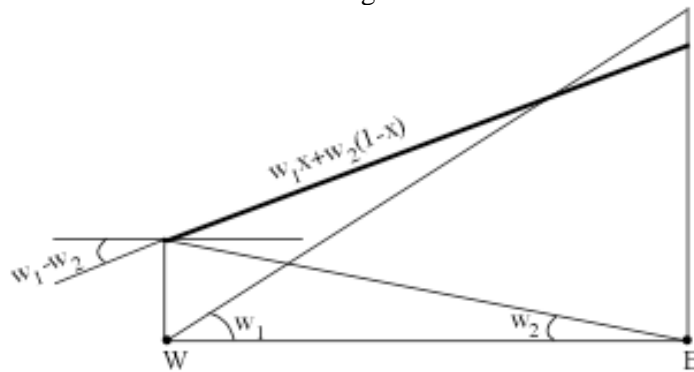


Figure 1.b:  $w_1 > w_2$

The way in which distance has affected transport costs over time may then be described succinctly as follows. The long period during which all movements were very costly and risky was followed by another during which, thanks to technological and organizational advances, ships could cross longer distances in one go, thus reducing their number of stops. On land, it was necessary to wait for the advent of the railroad for appreciable progress to occur, but the results were the same. In both cases, long-distance journeys became less expensive and no longer demanded the presence of relays or rest areas. Such an evolution in technologies has favoured places of origin and destination at the expense of intermediate places. As this argument may be extended to the case of any transport network having several nodes and markets, we may confidently assert that *increasing returns in transport explain why places situated between large markets and transport nodes have lost many of their activities*. Stated in a different way, the construction of new and large transport infrastructures will be beneficial to the main centers it connects, but not the regions it crosses. But if the global morphology of the network is changed through new and bigger nodes (e.g. Singapore or Chicago), these infrastructures may affect the location of economic activity.

To sum up, scale economies in production and transport activities have combined to lead to the spatial concentration of human activities. In particular, the development of new transport technologies



exhibiting a high degree of increasing returns strengthens the tendency toward more spatial polarization of high value-added activities.

### 3. THE MOBILITY OF FIRMS AND WORKERS

Countries and regions are affected not only by the growing mobility of commodities but also by that of production factors (e.g. capital and labour). What I want to stress here is that *lowering transport costs change firms' and workers' incentives to move*. It is, therefore, crucial to have a good understanding of how firms and workers react to these changes in order to assess the full impact of trade and transport policies. In this respect, it should be stressed that policy-makers often overlook the fact that their decisions impact on the location choices made by firms and households. These choices may lead to a new pattern of economic activity that vastly differs from the existing one. In particular, the economic geography approach to factor mobility highlights the fact that the mobility of factors need not reduce spatial inequality. It also stresses the fact that the mobility of firms and workers do not have the same impact on the global economy.

#### 3.1 The home-market effect

Both economists and geographers agree that a large market tends to increase the profitability of the firms established in it. More generally, the idea is that locations that have good access to several markets offer firms a greater profit. Hence, it is reasonable to expect that the firms that set up in large regions enjoy higher profits than the ones installed in small ones. In brief, firms would seek the locations with the highest market potential where demand is high and transport costs low (Redding and Venables, 2004). The core region should, therefore, attract new firms, thereby heightening the inequalities between the core regions and the others. Nevertheless, as firms set up in the core regions, competition there is also heightened, thereby holding back the tendency to agglomeration. Consequently, the interregional distribution of firms is governed by two forces pulling in opposite direction: the agglomeration force is generated by firms' desire for *market access*, while the dispersion force is generated by firms' desire to avoid *market crowding*.

This question has been studied in a standard two-region, two-sector, and two-factor economy (Helpman and Krugman, 1985). The industrial sector produces differentiated goods under increasing returns and imperfect competition, using capital and labour, whereas the traditional sector produces one good under constant returns and perfect competition, using labour only. This setting combines the mobility of both commodities and capital, while consumers/workers continue to be immobile. Furthermore, the mobility of goods is imperfect because their shipments incur positive transport costs. It is therefore tempting to conclude that the region with the larger market will always attract firms for the reason that this location minimizes total transport costs to both markets. However, as said above, this argument ignores the fact that when more firms locate within the same region, local competition is intensified and profits are lower.

When one region is larger in terms of population and purchasing power, this push and pull system reaches equilibrium when this region hosts *a more than proportionate share of firms*, a result that has been coined the "home market effect" (HME). Because of its comparative advantage in terms of size, it

seems natural that the larger region should attract more firms. What is less expected is that the share of firms exceeds the relative size of this region, thus implying that the initial advantage is magnified. This is because firms installed in the larger region have a better access to a bigger pool of consumers that allows them to produce at a lower average cost. Hence, contrary to general belief, capital does not necessarily flow from the regions where it is abundant to the regions where it is scarce.

Moreover, the HME is amplified by decreases in transport costs: *more firms choose to set up in the larger region when transport costs decrease*. This somewhat paradoxical result can be understood as follows. On the one hand, lower transport costs makes exports to the smaller market easier, which allows firms to exploit more intensively their scale economies; on the other hand, lower transport costs also reduces the advantages associated with geographical isolation in the smaller market where there is less competition. These two effects push toward more agglomeration of the industrial sector, thus implying that, as transport costs go down, the smaller region gets de-industrialized to the benefit of the larger one. The HME is thus liable to have unexpected implications for transport policy, such as that implemented by the European Commission in its cohesion program. By making the transport of goods cheaper in *both* directions, the construction of a new infrastructure permits an increase in both imports to, and exports from, the smaller region. As seen above, a transport cost-reducing policy is likely to induce some firms to pull out of the smaller region, thus failing to reduce regional disparities. To some extent, this explains the disillusion regarding the effectiveness of policies that aim for a more balanced distribution of activities across the European Union (Midelfart-Knarvik and Overman, 2002).

It is well documented that on average firms and workers tend to be more productive in larger markets (Syverson, 2004). Once it is recognized that firms are heterogeneous in productivity, location choices act as a selection device. Specifically, decreasing transport costs lead to the gradual agglomeration of low-cost firms in the larger region because these firms are able to survive in a more competitive environment. In contrast, high-cost firms seek protection against competition from the low-cost firms by establishing themselves in the smaller region. This implies a higher productivity level in large markets than in small markets. However, as the global economy gets more and more integrated, the selection effect is turned upside down, the market access effect stressed above becoming the dominant force. Consequently, as transport costs decline, interregional productivity differences first increase and then decrease. Note also that the least efficient firms go out of business because global competition is too tough for them to survive in either region.

The HME cannot be readily extended to multi-regional set-ups because there is no obvious benchmark against which to measure the “more than proportionate” share of firms. But why should one bother about the existence of many regions instead of two? The new fundamental ingredient that a multi-regional setting brings about is that the accessibility to spatially dispersed markets varies across regions. In other words, the *relative position* of a region within the network of exchanges (which also involves cultural, linguistic and political proximity) matters. Any global (local) change in this network such as market integration (the construction of a major transportation link) is likely to trigger complex effects that vary in non-trivial ways with the properties of the graph representing the network (Thomas, 2002). When there are only two regions, the overall impact can be captured through the sole variation in transport costs. On the contrary, when there are many regions, a change that directly affects two regions generates general equilibrium effects that are unlikely to leave the remaining regions unaffected. In particular, a multi-regional setting should make it possible to study how lowering transport costs amplify or reduce the geographical advantage and disadvantage held by different regions.

Unfortunately, economic geography and urban economics do not have much to say regarding those questions, although the evidence shows that accessibility strongly affects the potential of regions and cities for development (Collier, 2007). To illustrate, Limão and Venables (2001) show that, in

comparison with the median coastal country, the median landlocked country bears an additional transport cost of 55%, while its volume of trade at the same income level and distance decreases by 60%. Differences in accessibility have another facet which is often ignored: the level of human capital is higher in regions with a greater market access (Redding and Schott, 2003). With this in mind, it should be clear that accounting explicitly for a multi-regional economy with different transport costs is a critical issue (Behrens et al., 2010). Given the high analytical complexity of the problem, there is a need for computable and calibrated general spatial equilibrium models coping with several sectors and regions connected through a network having a specific design. In particular, what we have seen in section 2.2 shows that strategic choices on how to extend or reform transport networks is very likely to affect the location of firms in ways that should be carefully investigated through such models.

### 3.2 The emergence of a core-periphery structure

While firms bring with them the benefits of added production capability, the returns from physical capital need not be spent in the region where it is invested. By contrast, when human capital moves to a new region, workers bring with them both their production and consumption capabilities. As a result, their relocation simultaneously affects the size of labour and product markets in both the origin and the destination regions, expanding in the former and shrinking in the latter. Another major difference is that the mobility of capital is driven by differences in nominal returns, whereas workers move when there is a positive difference in real wages. Indeed, the gap in living costs matters to workers who consume in the region where they work, but not to capital-owners who consume their income in their region of residence, which need not be the region where their capital is invested. When some workers choose to migrate, their decisions change the relative attractiveness of both origin and destination regions. The resulting effects have the nature of externalities because workers do not account for them when making their decisions to move. Moreover, these externalities are pecuniary because prices fail to reflect the true social value of individual decisions when markets are imperfectly competitive.

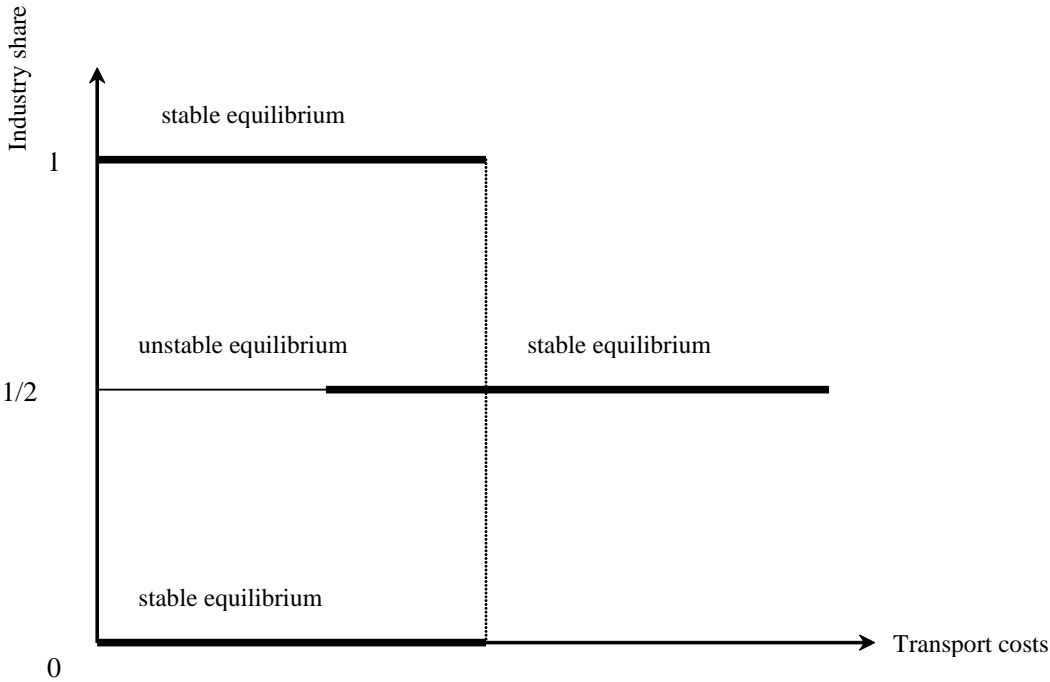
As in the foregoing, let us consider a two-region, two-sector, and two-factor economy. One production factor (unskilled labour) is spatially immobile and used as the input in the traditional sector; the second factor (skilled labour) is spatially mobile and used as the input in the industrial sector. In what has come to be known as the core-periphery model, two major effects are at work: one involves firms and the other workers. Assume that one region becomes slightly bigger than the other. First, a larger market size leads to a higher demand for the industrial goods. This generates a more than proportionate increase in the share of firms, which pushes nominal wages up. Second, the presence of more firms means a greater variety of local products as well as a lower local price index – a cost-of-living effect. Accordingly, real wages should rise, and this region should attract a new flow of workers. The combination of these two effects gives rise to a *cumulative causation* process that leads to the agglomeration of firms and skilled workers in a single region - the *core* of the economy, while the other region becomes the *periphery*.

Even though this process seems to generate inevitably a “snow ball” effect, it is not so clear that it will always develop according to that prediction. Indeed, the foregoing argument has ignored several key impacts of migration on the labour market. On the one hand, the increased supply of labour in the region of destination will tend to push wages down. On the other hand, the increase in local demand for industrial goods leads to a higher demand for labour. Thus, the final impact on nominal wages is hard to predict. Likewise, there is increased competition in the product market, which makes the region less attractive to firms. The combination of all those effects may lead to a “snowball meltdown”, which could result in the spatial dispersion of firms and workers.

Turning to the specific conditions for agglomeration or dispersion to arise, Krugman and others have shown that *the level of transport costs is the key-parameter* (Krugman, 1991; Fujita *et al.*, 1999). On the one hand, if transport costs are sufficiently high, interregional shipments of goods are discouraged, which strengthens the dispersion force. The economy then displays a symmetric regional pattern of production in which firms focus mainly on local markets. Because the distribution of workers is the same within each region, spatial disparities vanish in that there are no interregional price and wage differentials. On the other hand, if transport costs are sufficiently low, then all firms will concentrate into the core, while the periphery will retain the traditional sector only. In this way, firms are able to exploit increasing returns by selling more goods in the region benefiting from the market expansion effects sparked by the migration of skilled workers without losing much business in the smaller market. Thus, the mobility of skilled labour is likely to exacerbate the HME discussed in section 3.1, the reason being that the size of local markets changes with labour migration. Figure 2 shows how sudden and big is the shift in the interregional distribution of the industrial sector.

Capital mobility and labour mobility are, therefore, not equivalent for the spatial organization of the economy. While spatial inequalities in section 3.1 reflect the exogenous distribution of capital-ownership, in the core-periphery setting they stem from the endogenous redistribution of human capital.

Figure 2



Despite its extreme nature, the above prediction provides a fairly neat description of the spatial unevenness of economic development observed in different periods and different continents. To illustrate, consider Bairoch’s (1997) estimates of the GDP per capita over the period 1800-1913 across European countries. This corresponds to a period of intense technological progress that preceded a long series of political disturbances.

<b>Countries</b>	<b>1800</b>	<b>1830</b>	<b>1850</b>	<b>1870</b>	<b>1890</b>	<b>1900</b>	<b>1913</b>
Austria-Hungary	200	240	275	310	370	425	510
Belgium	200	240	335	450	55	650	815
Bulgaria	175	185	205	225	260	275	285
Denmark	205	225	280	365	525	655	885
Finland	180	190	230	300	370	430	525
France	205	275	345	450	525	610	670
Germany	200	240	305	425	540	645	790
Greece	190	195	220	255	300	310	335
Italy	220	240	260	300	315	345	455
Netherlands	270	320	385	470	570	610	740
Norway	185	225	285	340	430	475	615
Portugal	230	250	275	290	295	320	335
Romania	190	195	205	225	265	300	370
Russia	170	180	190	220	210	260	340
Serbia	185	200	215	235	260	270	300
Spain	210	250	295	315	325	365	400
Sweden	195	235	270	315	405	495	705
Switzerland	190	240	340	485	645	730	895
United Kingdom	240	355	470	650	815	915	1035
<b>Mean</b>	200	240	285	350	400	465	550
<b>Coefficient of variation</b>	0,12	0,18	0,23	0,31	0,38	0,39	0,42

Table 1: **Per capita GDP of European countries expressed in 1960 US do**

Source: Bairoch (1997)

Even if the numbers given in Table 1 must be used cautiously, they reveal clear tendencies. First, in 1800, most countries, except the Netherlands and, to a lesser extent, the United Kingdom, had fairly similar incomes per capita. As the Industrial Revolution developed and spread across the continent, each country experienced growth: the average GDP increases from 200 dollars in 1800 to 550 dollars in 1913. However, this process affected countries in a very unequal way. This is shown by the rise of the coefficient of variation that rose from 0.12 to 0.42, which confirms the existence of strongly rising spatial inequalities. Second, countries with the highest growth rates are those located close to the United Kingdom, which became the centre of the global economy of the nineteenth century. This is readily verified by means of a regression of the logarithm of the GDP per capita on the logarithm of the distance to the UK, which shows that the impact of this variable is significantly negative. Moreover, the absolute value of this regression coefficient, which has the meaning of elasticity, rises from 0.090 in 1800 and reaches a peak equal to 0.426 in 1890 (and remains stable afterwards). Stated differently, before the Industrial Revolution, a decrease of 10% in the distance to the UK is accompanied by an increase of the GDP per capita equal to 0.9%. By World War I, this elasticity had reached 4.4%, thus showing how far spatial inequalities had evolved during the 19th century.

It is worth stressing that the emergence of the European core-periphery structure arose while transport costs were falling at a historically unprecedented pace. According to Bairoch (1997), on the whole, between 1800 and 1910, the reduction in the real average prices of transportation was on the order of 10 to 1. Therefore, while the European economy experienced a rapid growth, *this phenomenal decrease in transport costs was accompanied with an increasingly unbalanced geographical distribution of wealth*. At the interregional level, Pollard (1981) similarly observes that “*the industrial regions*

*colonize their agricultural neighbours [and take] from them some of their most active and adaptable labour, and they encourage them to specialize in the supply of agricultural produces, sometimes at the expense of some pre-existing industry, running the risk thereby that this specialization would permanently divert the colonized areas from becoming industrial themselves.”*

Another important implication of the cumulative causation at work in the core-periphery model is the emergence of what can be called a *putty-clay geography*. Even though firms are a priori footloose, once the agglomeration process is set into motion, it keeps developing within the same region. Individual choices become more rigid because of the self-reinforcing nature of the agglomeration mechanism (the snowball effect mentioned above). In other words, the process of agglomeration sparks a *lock-in effect*. Hence, although firms and workers are (almost) freed from natural constraints, they are still connected through complex networks of interactions, which are probably more difficult to unearth than the old location factors related to the supply of natural resources.

#### 4. THE BELL-SHAPED CURVE OF SPATIAL DEVELOPMENT

The core-periphery model overlooks many costs whose origin lies in the space-economy (e.g. the various congestion costs generated by the emergence of an agglomeration). It also leads to a very extreme prediction that might not be robust against the introduction of additional parameters. This is what I want to cover in this section through a few suggestive examples.

##### 4.1 Vertical linkages

So far, agglomeration has been considered as the outcome of a cumulative causation process fed by the mobility of workers. However, agglomeration of economic activities also arises in contexts in which labour mobility is very low, as in most European countries. This underscores the need for alternative explanations of industrial agglomeration. One strong contender is the presence of input–output linkages between firms: the output of one firm can be an input for another, and vice versa. In such a case, the entry of a new firm in a region not only increases the intensity of competition between similar firms; it also increases the market of upstream firm–suppliers and decreases the costs of downstream firm–customers.

This is the starting point of Krugman and Venables (1995). Their idea is beautifully simple and suggestive: the agglomeration of the final sector in a particular region occurs because of the concentration of the intermediate industry in the same region, and conversely. Indeed, when firms belonging to the final sector are concentrated within a single region, the local demand for intermediate inputs is very high, thus making this region very attractive to firms producing intermediate goods. Conversely, because intermediate goods are made available at lower prices in the core region, firms producing final goods find that region very attractive. Thus, a cumulative process may still develop that leads to industrial agglomeration within the core region. In this alternative setting, new forces are at work. Indeed, if firms agglomerate in a region where the supply of labour is inelastic, then wages must surely rise. This in turn has two opposite effects. On the one hand, consumers' demand for the final product increases because they have a higher income. This is again a market expansion force, triggered now by higher incomes rather than larger populations. On the other hand, such wage increases also push

toward the re-dispersion of firms. Indeed, when the wage gap between the core and the periphery becomes sufficiently large, some firms will find it profitable to relocate in the periphery, even though the local demand for their output is lower than in the core. The agglomeration is thus self-defeating, especially when transport costs are low because demand asymmetries have a weak impact on profits.

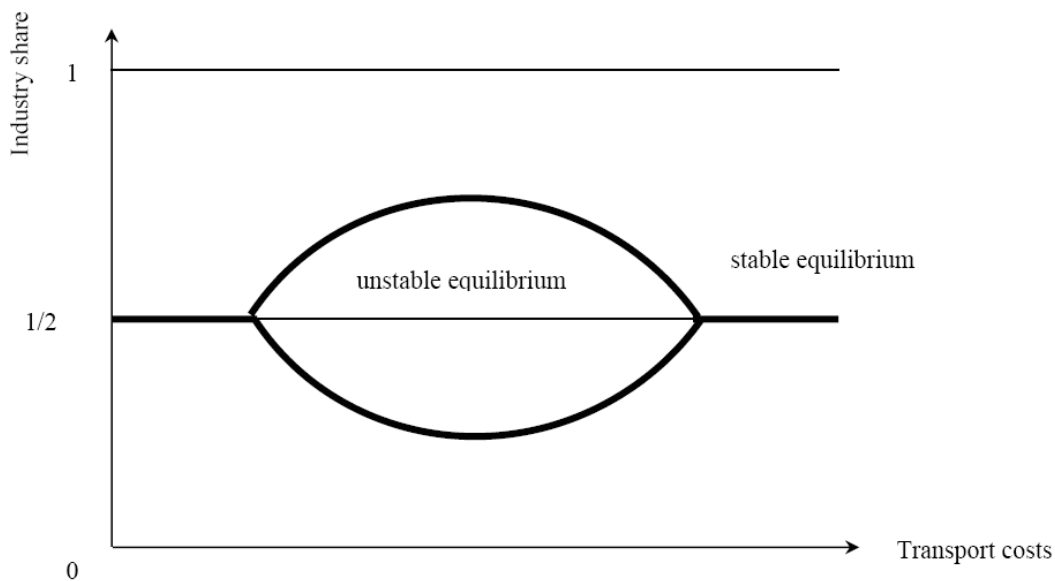
Thus, the set of equilibrium patterns obtained in the presence of vertical linkages is much richer than in the core-periphery model. In particular, if a deepening of economic integration triggers the concentration of industrial activities in one region, then beyond a certain threshold, an even deeper integration may lead to a reversal of this tendency. Some firms now relocate from the core to the periphery. In other words, the periphery experiences a process of reindustrialization. Simultaneously, the core might start losing firms, thus becoming de-industrialized. Therefore, *economic integration would yield a bell-shaped curve of spatial development*. By reducing the tension between the market outcome and the political concern for more spatial equity, the bell-shaped curve of spatial development lends support to a deeper integration of European economies.

## 4.2 Imperfect labour mobility

In the core-periphery model, workers are assumed to have the same preferences. It is highly implausible, however, that all individuals will react in the same way to a given real wage gap between regions. Some of them show a high degree of attachment to the region where they are born and will stay put even though they may guarantee to themselves higher living standards in another region. In the same spirit, lifetime considerations such as marriage, divorce and the like play an important role in the decision to migrate. Note also that regions are not similar and exhibit different natural and cultural features. Typically, individuals exhibit idiosyncratic tastes about such attributes, so that non-economic considerations matter to potentially mobile workers when they make their decision to move or not. In particular, as argued in hedonic theory of migration, once individual welfare levels get sufficiently high through the steady increase of income, workers tend to pay more attention to the non-market attributes of their environment.

Although individual migrations are difficult to model, it turns out to be possible to identify their aggregate impact on the spatial distribution of economic activities by using discrete choice theory. Recall that discrete choice models, which are widely used in transport analysis, aim at predicting the aggregate behaviour of individuals facing mutually exclusive opportunities such as modal choices. Using the logit model permits to assess the impact of heterogeneity in migration behaviour in that interregional migrations become sluggish (Tabuchi and Thisse, 2002). More precisely, as transport costs steadily decline, more and more skilled workers get agglomerated in one region for the reasons explained in the foregoing section, but the agglomeration process is now gradual and smooth. After having reached a peak in their spatial concentration, skilled workers gradually get re-dispersed. This is because the non-economic factors that drive the choice of a residential location become predominant and take over the economic forces stressed above, the intensity of which decreases with declining transport costs. As a result, the relationship between the degree of spatial concentration and the level of transport costs is bell-shaped (see Figure 3 for an illustration). Therefore, idiosyncratic factors in migration decisions act as a strong dispersion.

Figure 3.



Hence, *within the EU polarization should arise on a relatively small scale*. For example, the analysis developed by Crozet (2004) suggests that Lombardy should attract firms within a radius ranging from 95 to 150 km from its centre. Consequently, this region is not expected to threaten any other major Italian region, since the largest city closest to Milan, i.e., Turin, is situated 141 km away, while Genoa and Rome are 164 and 576 km away, respectively.

The sticky mobility of European workers also has an implication that has been overlooked by policy-makers: the relative dispersion of the industrial sector caused by the heterogeneity of preferences is likely to generate efficiency losses at the macroeconomic level. These stem from larger trade flows and insufficient exploitation of scale economies. If so, the low mobility of European workers thus presents two opposite facets: on the one hand, it corresponds to workers' greater attachment to their region or country as embedded in their individual preferences; on the other hand, it gives rise to some losses with respect to productive efficiency, and these are liable to dampen European economic growth.

### 4.3 The spatial fragmentation of firms

A growing number of firms choose to break down their production process into various stages spread across different places. Specifically, the modern firm organizes and performs its activities in distinct locations, which altogether form a *supply chain* starting at the conception of the product and ending at its delivery. This spatial fragmentation of production aims at taking advantage of differences in technologies, factor endowments, or factor prices across places (Feenstra, 1998). The most commonly observed pattern is such that firms relocate their production activities in low-wage regions or countries, while keeping their strategic functions (e.g. management, R&D, marketing and finance) concentrated in a few affluent urban regions where the high-skilled workers they need are available.

In such a context, the development of new communication technologies is a major force that should be accounted for. It goes hand in hand with the growing role of transportation firms in the global logistics. With this in mind, two types of spatial costs must then be considered, namely communication costs and transport costs. Low transport costs allow firms producing overseas to sell their output on their



home market at a low price. Equally important, but perhaps less recognized, is the fact that coordinating activities within a firm is more costly when headquarters and plants are physically separated because the transmission of information remains incomplete and imperfect (Leamer and Storper, 2001). However, lower communication costs make coordination easier and, therefore, facilitate the process of fragmentation. More precisely, in order to make low-wage areas more attractive for the set-up of their production, firms need both the development of new communication technologies and substantial decreases in transport costs.

Assume that each firm has two units, one headquarter and one plant. All headquarters are located in the same region and use skilled labour, whereas plants use headquarter-services together with unskilled labour. A firm is free to decentralize its production overseas by choosing distinct locations for its plant and headquarter. Two main scenarios are to be distinguished as they lead to very different patterns (Fujita and Thisse, 2006). When communication costs are high, all firms are national and established in the core region. Once communication costs steadily decrease, the industry moves toward a configuration in which some firms become multinational whereas others remain national. Eventually, when these costs have reached a sufficiently low level, the economy ends up with a de-industrialized core that retains only firms' strategic functions.

According to the value of communication costs, a fall in transport costs may lead to fairly contrasted patterns of production. When communication costs are high, reducing transport costs leads to a growing agglomeration of plants within the core, very much as in the core-periphery model. Hence, the core region attracts all activities. Things are totally different when communication costs are low. For high transport costs, most plants are still located within the core. However, once these costs fall below some threshold, the relocation process unfolds over a small range of transport cost values. This could explain why the process of de-industrialization of some developed regions seems, first, to be slow and, then, to proceed quickly, yielding a space-economy very different from the initial one. As suggested by the declining part of the bell-shaped curve, the welfare gap between the core and the periphery shrinks. Nevertheless, this catching-up process, which leads to a higher welfare level in the periphery, causes welfare losses in the core.

## **5. THE TRADE-OFF BETWEEN COMMUTING COSTS WITHIN THE CITY AND TRANSPORT COSTS BETWEEN CITIES**

Tradable goods do not account for a very large fraction of the GDP of rich countries. On the contrary, many consumption goods and services are produced locally and not traded between regions. The forces pushing toward factor price equalization within every region thus lead to additional costs generated by the agglomeration of firms and workers within the same region. This in turn increases the cost of living in the larger region and may induce some workers to change place. A natural way to capture this phenomenon is to focus on the housing market where competition gets tougher as more people establish themselves in the same area, thus raising housing and land costs.

As mentioned above, a human settlement of a sizeable scale almost inevitably takes on the form of a city. Typically, a city possesses one main employment centre that gathers together firms, while workers are distributed all around it. Workers seek to reduce their *commuting costs* by choosing a living place in the vicinity of their working place. However, because of the scarcity of land, everybody cannot live close

to the city centre. This in turn implies that workers must commute between the workplace and their living place. Competition for land among workers gives rise to a *land rent* that varies inversely with the distance to the city centre, thereby compensating workers living far from their workplace. In other words, there is a trade-off between commuting and housing costs: the former increasing with distance while the latter decrease (Fujita, 1989).

Land rent augmented by commuting costs defines what I call *urban costs*. In most developed countries, they stand for a large, and growing, share of households' budgets. In the United States, housing accounts on average for 20% of household budgets while 18% of total expenditures is spent on car purchases, gasoline, and other related expenses. The latter does not account for the cost of time spent in travelling, which keeps rising. We thus find it reasonable to claim that more than 30% of the income of US households is spent on urban costs. In France, between 1960 and 2000, housing and transportation expenses increased from 23% to 40% of household expenditures, which represents a growth of almost 75% despite an almost quadrupling of the real per capita income. Moreover, as predicted by urban economics, urban costs increase with city size. In the United States, urban costs are less than \$15,000 per year in cities like Pittsburgh, Baltimore and Kansas City, but rise to nearly \$20,000 per year in, e.g., San Francisco, Los Angeles and New York. Looking at French data reveals that, in 2000, urban costs represented more than 40% of individual incomes in Paris, but around 33% of individual incomes in medium-sized cities. Urban costs play a growing role in shaping the city, but we will see that they also have a strong impact on national urban systems and intercity trade flows.

## 5.1 The monocentric city

In the monocentric city, firms are agglomerated and form the central business district (CBD), inducing all households to commute between their working place and their residences. It is empirically well documented that firms seek proximity in order to enjoy the various types of benefits generated by the need for strategic information, such as knowledge spillovers, business communications and social interactions (Rosenthal and Strange, 2004). Knowledge, ideas and tacit information generate spillovers from one firm to another. Consequently, if economic agents possess different pieces of information, pooling them through informal communication channels can benefit everyone. Firms get agglomerated in a CBD when external economies are strong, commuting costs are low, or both. This is because firms are able to capitalize on the benefits generated by the various spatial externalities generated endogenously through non-market interactions among firms, without having to compensate workers for their high commuting costs. At the other extreme, firms and workers are mixed across locations, very much as in preindustrial cities endowed with poor urban transport systems. This configuration emerges as an equilibrium outcome when spatial externalities are weak, commuting costs are high, or both (Fujita and Thisse, 2002). In short, *high commuting costs fosters the dispersion of activities within the city, whereas low commuting costs leads to the specialization of land use between firms and households*. This is reminiscent of what we have seen in the core-periphery model in that lower mobility costs push toward more agglomeration.

But this is only one side of the coin. Let us return to the core-periphery setting discussed in section 3.2, and assume that a large share of the industrial sector is concentrated in a big city. If transport costs steadily decrease, the urban costs borne by workers within the core become too high to be compensated by a better access to the array of tradable goods. Therefore, dispersion arises once transport costs have reached a sufficiently low level by comparison with commuting costs. Lower urban costs in the periphery more than offset the additional transport costs to be paid for consuming the varieties produced in the core. Consequently, as the costs of shipping goods keep decreasing, the economy involves the following phases: dispersion, agglomeration, and re-dispersion. This is strikingly similar to the

bell-shaped curve discussed in section 4. What triggers the re-dispersion of firms and workers is now the crowding of the land market. The relocation of the manufacturing sector away from large metropolitan areas toward medium-sized cities illustrates the impact that high commuting costs and low transport costs may have on firms' locations.

It should be clear that the re-dispersion phase depends on the strength of the spatial externalities among firms as well as on the efficiency of the urban transport means used by workers. The spectacular drop in commuting costs sparked by the near-universal use of cars has facilitated the agglomeration of activities within large cities, and then has delayed the interregional re-deployment of activities. So it is the relative evolution of interregional transport costs and intra-urban commuting costs that determines the structure of the space-economy. Stated differently, what matters for the global economy is not just the evolution of transport costs between regions; what goes on inside the different regions is also crucial.

## 5.2 The polycentric city

The foregoing argument suggests that workers and firms get re-dispersed because urban costs become very high in the core region. However, once it is recognized that big cities may become polycentric through the development of secondary business centers (SBDs), the average commuting costs and land rent borne by those working in a SBD are lower than those paid by the individuals working in the CBD. Simultaneously, because fewer workers commute to the CBD, the corresponding workers also bear lower urban costs. In sum, workers' welfare becomes higher when the city becomes polycentric. By the same token, firms are able to pay lower wages and land rents while retaining most of the benefits generated by urban agglomerations. For example, Timothy and Wheaton (2001) report substantial 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). Thus, we may expect the escalation of urban costs in large cities to prompt the redeployment of activities in a polycentric pattern.

For this to happen, however, firms located in SBDs must be able to maintain very good access to the inner city, which provides highly specialized business-to-business services (Porter, 1995), which in turn requires low communication costs. Indeed, SBDs have not eliminated the importance of the CBD. This is confirmed by Schwartz (1993) who observes that about half of the business services consumed by US firms located in suburbia are supplied in city centers. In the case of New York, Los Angeles, Chicago and San Francisco, this figure even grows to 65%. The same is true of France, as can be seen from the distribution of higher-order metropolitan functions (executives, engineers, and business service company management jobs, research, commerce, banking and insurance, art). These are more common in city centers than in their periphery. For example, for the Paris urban area, they make up 19.3% of employment within Paris itself, 15.7% in the suburbs, and 6.6% in the outside belt (Julien, 2002). These higher-order functions seek out central positions and major city centers retain specific features relative to SBDs. This implies that firms in SBDs incur an access cost to the main centre when they resort to these higher urban functions. Even if this cost is likely to have sharply fallen with the reduction in communication costs, allowance still has to be made for it.

By introducing communication costs, we account for the fact that agglomeration and dispersion across space may take two quite separate forms because they are now compounded by centralization or decentralization of activities within the same city. When commuting and communication costs are high, the space-economy is likely to be formed by several small cities. In contrast, when communication costs reach low values while commuting costs take intermediate values, large polycentric cities are likely to emerge. Therefore, by facilitating the formation of SBDs, the development of new information and communication technologies slows down the redispersion process. Stated differently, *employment*

*decentralization within the metropolis allows the core regions to retain their primacy* (Cavailhès et al., 2007). Such results shed light on the interplay between different types of spatial friction affecting the location of economic activities between and within urban agglomerations. Historical evidence shows that both trade and commuting costs have been decreasing since the beginning of the Industrial Revolution. Once again, what matters for the organization of the space-economy is the relative evolution of these two costs.

Nevertheless, the emergence of a handful of large polycentric cities dominating the European economic space is not inevitable. High-speed rail (HSR) provides fast and convenient travel between large and medium-sized cities by reducing the opportunity cost of being located in one city rather than another, especially when urban costs are high. If HSR is sufficiently cheap and fast, one can think of this transport mode as stimulating the emergence of several interregional urban systems within the EU. In this case, HSR would stabilize prevailing conurbation patterns within Europe by putting a brake on firms' and skilled workers' tendencies to agglomerate in big cities. This is in line with the European cohesion policy objectives.

All of this draws attention to two facts that policy-makers often neglect: on the one hand, local factors may change the global organization of the economy and, on the other, global forces may affect the local organization of production and employment. Stated in a different way, the local and the global interact to shape the entire economy. This relationship calls for a better coordination of transport policies at the city and interregional levels. In doing so, one should also account for the changes in new information and communication technologies as these ones influence the way firms conduct their business across space.

## 6. CONCLUDING REMARKS

- (i) In 1885, Wilhelm Launhardt, a civil engineer who worked on the construction of transport infrastructures in Germany, noted that “the improvement of means of transport is dangerous for costly goods: these lose the most effective protection of all tariff protections, namely that provided by bad roads.” And indeed, we have seen that a policy that systematically aims at improving the accessibility of a small region to the global economy runs the risk of being ineffective in promoting the development of this region. *The cumulative nature of the agglomeration process makes the resulting imbalanced pattern of economic activity particularly robust to various types of shocks.* In other words, affluent regions enjoy the existence of agglomeration rents that single-minded policies cannot easily dissipate. Consequently, the objective of the European Commission being to foster a more balanced distribution of economic activities across European regions, it should add more instruments to its policy portfolio.
- (ii) However, we have also seen that the evolution of the space-economy depends on the interaction between several additional forces. The sluggish mobility of workers, the existence of non-tradable goods, the demand for intermediate goods, or the spatial fragmentation of firms, all suggest the existence of a bell-shaped curve linking regional disparities and spatial integration. Taking into account these new forces leads us to believe that a sufficiently extensive economic integration of the space-economy is likely to favour the development of

several large urban regions, which could be spread over the entire territory of the EU. Eventually, spatial inequalities at the interregional level would be (partially) reduced through the redispersion of the industrial sector, very much as in the US where this sector is mainly located within medium- or low-population density areas (Glaeser and Kohlhase, 2004). By substituting long-distance commuting for the migration of skilled workers, high-speed rail may play a major role in this process. However, for the HSR to have a significant impact on the location of activities, it is crucial to connect cities that have a high potential of interaction. It would be naive to expect the HSR to become by itself the engine of regional development. On the contrary, such a transport policy must be part of a broader and integrated portfolio of instruments. The European Commission and many national governments have spent enough money on building “cathédrales dans le desert.”

- (iii) During the last decade, the media have embraced the idea that we would be living in a world where the tyranny of distance, which weighed so heavily on human history, would be gone. The spectacular and steadily drop in transport costs since the mid-19th century, relayed by the retreat of protectionism and, more recently, by the near-disappearance of communication costs, is said to have freed economic agents from the need for proximity. In this way, technology and globalization would have joined together to make the traditional geography of activities obsolete, and transform yesterday’s world with its peaks and troughs into a “flat world”.

Recent empirical and theoretical work in economic geography shows a very different reality. While it is true that the importance of being close to natural resources has largely declined, thus giving firms and households more freedom, distance and location have not disappeared from economic life. For example, by showing that distance remains a major impediment to trade and interactions between spatially separated firms and consumers, the gravity model invalidates the idea that the tyranny of distance would be over (Head and Mayer, 2004). It is worth stressing, however, that market accessibility must be evaluated by *all* the costs generated by the various types of spatial frictions that firms and their customers face when trading goods. Such costs are called *trade costs*. Spulber (2007) refers to them as “the four Ts”:

- *Transaction* costs that result from doing business at a distance due to differences in customs, business practices, as well as political and legal climates;
- *Tariff and non-tariff* costs, such as different anti-pollution standards, anti-dumping practices, and the massive regulations that still restrict trade and investment;
- *Transport* costs *per se*, because goods have to reach their consumption place, while many services remain non-tradable; and
- *Time* costs, as, despite Internet and video-conferences, there are still communication impediments across dispersed distribution and manufacturing facilities that slow down reactions to changes in market conditions, while the time needed to ship certain types of goods has a high value.

Transport policies cannot ignore this multi-facet of trade costs, nor their mutual interactions.

- (iv) Despite more precise measurements of trade costs, economic geography still fails to provide an explicit description of the interactions between the transport and industrial sectors, or between carriers themselves. In particular, modelling explicitly the transport sector and the formation of freight rates through the strategic behaviour of carriers, as well as competition between

transport modes, should attract more attention (Behrens et al., 2009). If trucking may reasonably be approximated by perfect competition in the wake of the Motor Carrier Act of 1980, which abolished most entry barriers and fare controls in the US, railroads are characterized by a small number of firms. Railroads are subject to high fixed costs, as they require heavy infrastructure, thereby creating natural oligopolies that behave strategically.

Moreover, integrating variables specific to the transport activity, such as density economies, market segmentation in the supply of transport services, logistic features, and scheduling considerations should also be addressed. All in all, it should be clear that a more realistic description of the transport sector would make economic geography and urban economics more appealing and relevant to transportation economists. This entire area is strongly under-analyzed and deserves much more attention in the future research agenda.

- (v) Economic geography has chosen to focus on the historical trend of falling trade costs. Yet, one may wonder whether an increase in trade costs would bring the economy back to the initial situation. The answer is probably not. Even though the agglomeration process is not completely irreversible, the putty-clay nature of the space-economy and the existence of agglomeration rents imply a strong inertia in the location of economic activities. In this respect, it is also worth stressing that economic geography models often exhibit hysteresis in which a lag occurs between the application and the removal of lowering trade costs and its subsequent effect on the location of agents.
- (vi) How to design “optimal” transport policies remains the most difficult issue. Policy recommendations depend primarily on what decision-makers want to optimize: global efficiency, spatial equity, the ecological footprint, or a combination of all of them? Cities and industrial clusters are replete with different types of externalities, namely interactions that are not mediated by the market. Although the process of interaction goes both ways, individuals worry only about their role as “receivers” but neglect the fact that they are also “transmitters” to the others. As a result, the optimal distribution of firms is more concentrated than the equilibrium one (Fujita and Thisse, 2002). This may come as a surprise since the conventional wisdom is that market cities are too crowded in the vicinity of the centre. Note, however, that this conclusion does not take into account the various negative externalities generated by congestion and pollution. This makes the overall assessment of land-use patterns in cities especially hard. One clear recommendation emerges from theoretical and empirical studies: for the agglomeration economies to produce their effects, the intra-urban mobility is crucial. To avoid free-riding and coordination failures, the optimal governance of cities should cover the whole area under consideration in order to permit the internalization of all costs and benefits (Cheshire and Magrini, 2009).

At the interregional level, the reasons for over- or under-agglomeration have more to do with linkages between firms and consumers-workers, through product and labour markets. Pecuniary externalities are critical because firms and workers do not account for the impact that their decisions to move have on the well-being of those who stay put as well as on those who live in the region of destination. Consequently, when migration flows are substantial, one may expect the interregional economy to be inefficiently organized. Preliminary analysis suggests that the mobility of firms and workers may yield a pattern of activities which is too concentrated. When some share of skilled workers finds it individually desirable to move to the larger region, the impact on the other skilled workers may be negative because the fiercer competition sparked on the local market is not outweighed by the better penetration of the smaller region. Hence, very much as in a huge prisoner’s dilemma, the moving workers may end up being worse off after having moved than before moving. On the other hand, when

the spatial economy is sufficiently integrated, the gains stemming from a better exploitation of scale economies become predominant, making the agglomeration of the industrial sector globally efficient. Note also that the over-agglomeration result does not account for the fact that technological progress brings about new types of innovative activities that benefit from being agglomerated, such as the R&D sector. This in turn may boost the growth rate of the global economy (Fujita and Thisse, 2002).

Last, we have seen that global forces are likely to affect the local organization of production and employment, whereas local factors may well change the global organization of the economy. This calls for the integration of the various types of spatial friction acting at different spatial scales. Such a task is probably out of reach for the time being, but it should guide us in setting the research agenda in transport analysis and in designing more effective policies.

## REFERENCES

- Bairoch, P. (1997), *Victoires et déboires, Histoire économique et sociale du monde du XVI<sup>e</sup> siècle à nos jours*, Paris: Editions Gallimard.
- Behrens, K., C. Gaigné and J.-F. Thisse (2009), “Industry Location and Welfare When Transport Costs Are Endogenous”, *Journal of Urban Economics* 65, 195-208.
- Behrens, K., A.R. Lamorgese, G.I.P. Ottaviano and T. Tabuchi (2010), “Beyond the Home Market Effect: Market Size and Specialization in a Multi-country World”, *Journal of International Economics*, forthcoming.
- Cavallières, J., C. Gaigné, T. Tabuchi, and J.-F. Thisse (2007), “Trade and the Structure of Cities”, *Journal of Urban Economics* 62, 383-404.
- Cheshire, P. and S. Magrini (2009), “Urban growth drivers in a Europe of sticky people and implicit boundaries”, *Journal of Economic Geography* 9, 85-116.
- Collier, P. (2007), *The Bottom Billion, ‘Why the Poorest Countries Are Failing and What Can Be Done About It’*, Oxford: Oxford University Press.
- Crozet, M. (2004), “Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model”, *Journal of Economic Geography* 4, 439-458.
- Diamond, J. (1997), “Guns, Germs, and Steel: The Fate of Human Societies”, New York: W.W. Norton.
- Duranton, G. and D. Puga (2004), “Micro-foundations of Urban Increasing Returns: Theory”, In J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*. Volume 4. Amsterdam: North Holland, 2063-2117.
- Feenstra, R.C. (1998), “Integration of Trade and Disintegration of Production in the Global Economy”, *Journal of Economic Perspectives* 12 (4), 31-50.
- Fujita, M. (1989), “Urban Economic Theory, Land Use and City Size”, Cambridge: Cambridge University Press.
- Fujita, M., P. Krugman and A.J. Venables (1999), “The Spatial Economy. Cities, Regions and International Trade”. Cambridge, MA: The MIT Press.
- Fujita, M. and J.-F. Thisse (2002), “Economics of Agglomeration: Cities, Industrial Location and Regional Growth”, Cambridge: Cambridge University Press.



- Fujita, M. and J.-F. Thisse (2006), “Globalization and the Evolution of the Supply Chain: Who Gains and Who Loses?” *International Economic Review* 47, 811-836.
- Glaeser, E.L. and J.E. Kohlhase (2004), “Cities, Regions and the Decline of Transport Costs”, *Papers in Regional Science* 83, 197-228.
- Head, K. and T. Mayer (2004), “The Empirics of Agglomeration and Trade”. In J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*. Volume IV. Amsterdam: North-Holland, 2609-2669.
- Helpman, E. and P.R. Krugman (1985), “Market Structure and Foreign Trade”. Cambridge, MA: The MIT Press.
- Julien, P. (2002), “Onze fonctions pour qualifier les grandes villes”, INSEE Première N°840.
- Krugman, P.R. (1991), “Increasing Returns and Economic Geography”, *Journal of Political Economy* 99, 483-499.
- Krugman, P.R., and A.J. Venables (1995), “Globalization and the Inequality of Nations”, *Quarterly Journal of Economics* 110, 857-880.
- Leamer, E.E. and M. Storper (2001), “The Economic Geography of the Internet Age”, *Journal of International Business Studies* 32, 641-655.
- Limão, N. and A.J. Venables (2001), “Infrastructure, Geographical Disadvantage, Transport costs, and Trade”, *World Bank Economic Review* 15, 451-479.
- Midelfart-Knarvik, K.H. and H.G. Overman (2002), “Delocation and European Integration: Is Structural Spending Justified?” *Economic Policy* 35, 321-359.
- Pollard, S. (1981), “Peaceful Conquest. The Industrialization of Europe 1760-1970”, Oxford: Oxford University Press.
- Porter, M.E. (1995), “Competitive Advantage of the Inner City”, *Harvard Business Review*, May-June, 55-71.
- Redding, S. and P.K. Schott (2003), “Distance, Skill Deepening, and Development: Will Peripheral Countries Ever Get Rich?” *Journal of Development Economics* 72, 515-541.
- Redding, S. and A. Venables (2004), “Economic Geography and International Inequality”, *Journal of International Economics* 62, 53-82.
- Rosenthal, S. and W. Strange (2004), “Evidence of the nature and sources of agglomeration economies”. In J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*. Volume 4. Amsterdam: North Holland, 2119–2171.
- Schwartz, A. (1993), “Subservient Suburbia”, *Journal of the American Planning Association* 59, 288-305.
- Spulber, D.F. (2007), “Global Competitive Strategy”, Cambridge: Cambridge University Press.

- Syverson, C. (2004), “Market Structure and Productivity: A Concrete Example”, *Journal of Political Economy* 112, 1181-1222.
- Tabuchi, T. and J.-F. Thisse (2002), “Taste Heterogeneity, Labor Mobility and Economic Geography”, *Journal of Development Economics* 69, 155-177.
- Timothy, D. and W.C. Wheaton (2001), “Intra-urban Wage Variation, Employment Location and Commuting Times”, *Journal of Urban Economics* 50, 338-366.
- Thomas, I. (2002), “Transportation Networks and the Optimal Location of Human Activities: A Numerical Geography Approach”, Cheltenham, U.K.: Edward Elgar.