ABSTRACT

This essay provides an overview of the role of infrastructure on economic geography in the light of both theoretical and empirical findings. *Two main lessons stand out. First, infrastructural improvements affect the geographical distribution of economic activities. Second, even when localized, infrastructure investment generates externalities that may diffuse quite far across the economy. These two lessons have two far-reaching policy implications. First, effective infrastructure projects require knowledge on their impacts on the spatial distribution of economic activities. These impacts depend crucially on the specific details of the projects and the specific sources of agglomeration economies they affect. Second, regions need to coordinate not only in terms of interregional infrastructure projects but also in terms of intraregional ones if they want to avoid beggar-thy-neighbour and self-defeating outcomes.*

Gianmarco I.P. Ottaviano (gianmarco.ottaviano@unibo.it) is Professor of Economics at the Università di Bologna, Dipartimento di Scienze Economiche.

Infrastructure and economic geography: An overview of theory and evidence

1. Introduction

Gianmarco Ottaviano

How important is infrastructure for regional development? What is its impact on the geographical distribution of economic activities? Does it promote agglomeration or dispersion?

The aim of this essay is to provide an overview of the role of infrastructure in affecting economic geography. In so doing, it summarises both theoretical and empirical findings pertaining to why, how and how much infrastructure contributes to the agglomeration or the dispersion of economic activities. It also discusses the implications of those findings for policy design.

Along the years countless studies have been devoted to uncovering the economic mechanisms driving the spatial distribution of economic activities. The availability of adequate analytical tools has been the main constraint on this endeavour. Indeed, for quite some time inadequate analytical tools kept space away from mainstream economics by classifying as irrelevant issues that were instead simply intractable. So, while what is really puzzling is the sheer extent of geographical concentration found at any scale of observation, "economists understood why economic activity spreads out, not why it becomes concentrated – and thus the central model of spatial economics became one that deals only with the way competition for land drives economic activities away from a central market" (Krugman 1995, p. 12).

The missing tools were models able to deal with increasing returns to scale and imperfect competition, possibly in a general equilibrium framework. In the last decades, however, these tools having become available, spatial economics has made a giant leap forward and regained its status in mainstream economics (see Fujita and Thisse 2002 for a detailed account). More practically, it has started to generate new insights to inform the design of economic policies (Baldwin *et al.* 2003).

Surveys of spatial economics abound covering both theoretical and empirical aspects (see, for example, all the chapters in Henderson and Thisse 2004). Rather than adding the *n*-th survey to an already long list, the present essay prefers to distil the key insights of spatial economics by means of a simple unified theoretical framework that provides an internally consistent and parsimonious way to relate the evolution of the economic landscape to a restricted number of parameters on which infrastructure is shown to act.

The essay in organized in five more sections. Section 2 presents the theoretical framework. Section 3 discusses how the framework can be interpreted to allow for the various sources of agglomeration economies highlighted in the literature. Section 4 uses the framework to discuss the role of infrastructure in regional development. Section 5 highlights specific empirical findings that are consistent with the implication of the theoretical framework. Section 6 concludes drawing some policy implications.

2. The location choice

Theoretical explanations of the geographical distribution of economic activities rest on modelling the location decision of firms. This decision is not trivial when two things are true. First, goods

and factors can be transported across space only at some cost. Second, the fragmentation of the production process reduces its efficiency, which happens when returns to scale are increasing at the plant level. Without transport costs space would be immaterial. Without plant-level scale economies, when faced with dispersed customers, firms would use the geographical fragmentation of production to circumvent transport costs by adapting to scattered demand and intermediate supply through many small local plants ('backyard capitalism'). In other words, "without recognizing indivisibilities – in human person, in residences, plants, equipment, and in transportation – urban location problems, down to those of the smallest village, cannot be understood" (Koopmans 1957, p. 154).

Both transport costs and scale economies are therefore necessary for a location problem to arise: Costly transportation gives physical substance to the concept of geography and increasing returns generate an economic trade-off between the 'proximity' to customers and the 'concentration' of production in as few plants as possible.[1]

2.1 Market seeking and cost saving attraction

The centrality of scale economies has important implications in terms of market structure. Indeed, since plant-level returns to scale are necessarily associated with market power, imperfect competition is inherent to the location problem.[2]

A simple monopolistic setup is enough to see how the relationship between proximity and concentration generates a location problem for any firm no matter whether involved in any interaction with other firms. Specifically, consider a situation in which there are two locations, H and F.[3] The former offers a larger local market for the monopolist's product and cheaper costs of production. Shipments between locations are hampered by trade barriers and technological constraints are such that the monopolist can profitably run one plant only. This indivisibility and costly shipments give substance to the monopolist's location problem.

Firms are attracted to locations with larger markets – production costs become more important as trade barriers fall.

Where will a profit maximizing monopolist place the plant? The answer is: In location H. Indeed, locating its plant in H allows the monopolist to minimize production costs ('cost saving attraction'). Moreover, it also allows the monopolist to minimize trade costs as, once in H, it can serve a larger share of overall demand locally, thus foregoing costly shipments ('market seeking attraction').

Since the monopolist would always locate in the place offering lower production costs and larger local demand these two factors act as agglomeration forces. The strength of these forces depends, however, on the level of trade barriers. In particular, lower trade barriers increase the relative importance of cost saving compared to market seeking. The reason is that, as trade barriers fall, firm location becomes increasingly immaterial in terms of access to customers.

2.2 Competition

A firm's location decision becomes more complex when it faces competitors. The reason is that geographical positioning can be used by the firm to relax competitive pressures and enhance its market power. This is the case both when competitors are a 'small group' of oligopolistic firms

1 Scotchmer and Thisse (1992) call this the 'folk theorem of spatial economics'.
2 This is the essence of the 'spatial impossibility theorem' proved by Starrett (1978).
3 See Behrens *et al.* (2007a) for a formal presentation of the theoretical framework underlying this section.

selling homogeneous products and when they are a 'large group' of monopolistically competitive firms selling differentiated products *à la* Chamberlin (1933). In both cases, location turns out to be a crucial decision variable for profit maximization.

To see this, let us introduce a competitor in the simple monopolistic setup. In particular, let us now assume the existence of two firms such that firm 1 is more efficient than firm 2. Each firm supplies one variety of a horizontally differentiated good. As before, technological constraints are such that each firm can profitably run one plant only and shipments between locations face trade barriers. Firms compete in two stages. First they decide where to locate, and then they choose how much to sell in the two markets.

Where will the profit maximizing duopolists place their plants? Under reasonable conditions, the strategic interaction between firms generates only two possible equilibrium outcomes. In the first, the firms co-locate in the low-cost/large-size location H ('agglomeration'). In the second, they are based in different locations with the low-cost firm serving the low-cost/large-size location ('sorting').

In order to understand under which conditions an outcome prevails over the other, it is useful to start from a situation in which there are no differences between locations, in terms of both market size and production costs, and between firms, in terms of efficiency. In this case, if firms chose the same location, they would face strong competition in their local market and weak competition in their distant market. Therefore, as the two locations are identical, one of the two firms would always find it profitable to relocate. The more so, the higher the substitutability between the products of the two firms and the higher the trade barriers protecting the distant market. Accordingly, competition promotes dispersion, especially in the presence of limited product differentiation and high trade costs.

Let us now introduce market size and production cost differences between locations, which reinstates the agglomeration forces described in the case of monopoly. This implies that, notwithstanding the dispersion force due to competition, the two firms may still prefer to co-locate in H provided that its size and cost advantages are large enough. This is more likely when product differentiation is pronounced and when trade costs are low as under these conditions competitive repulsion is weak.

Finally, let us see how efficiency differences between firms interact with differences in production cost and market size. While the focus on efficiency stresses the technological dimension of firm heterogeneity, it is worthwhile pointing out that similar results in terms of location choice can be obtained by focusing instead on the quality of the products offered by the firms. In this respect, while the elasticity of substitution between firms' products measures 'horizontal' product differentiation, the difference in firm efficiency measures 'vertical' product differentiation.

When efficiency differences between firms are introduced together with market size and production cost differences between locations, the less efficient firm 2 has a stronger incentive to avoid the fierce competition associated with agglomeration in the advantaged region H. As a result, firm heterogeneity acts as an additional dispersion force, the more so the higher the trade costs and the smaller the differentiation of products.

Less efficient firms dislike locations with strong competition – even more so with higher trade barriers.

Hence, firm heterogeneity fosters dispersion when matched with market size and production cost differences between locations. In other words, while horizontal differentiation favours the spatial concentration of economic activities, vertical differentiation hampers it.

3. Agglomeration economies

The simple theoretical framework discussed in Section 2 suggests that asymmetries (*i.e.*, differences) in market size and production cost between locations promote the geographical concentration of economic activities.

There are two different ways of looking at those asymmetries. In a first perspective, asymmetries in market size and production cost are taken as exogenously given. As such, they reflect respectively the relative advantage that location H has in terms of local production and consumption amenities deriving from climatic conditions, natural resources and natural means of communication.

While places do have different abundance of natural resources, proximity to natural means of communication, and climatic conditions, these features (also known as 'first nature') provide only a partial explanation of the pronounced differences in development existing even between areas that are fairly similar in terms of such exogenous characteristics. For this reason it has been argued that regional imbalances have to be driven also by additional, 'second nature' forces that are inherent to the functioning of economic interactions and are able to cause uneven development even across ex-ante identical places.[4] In this second perspective, market size differences and production cost differences are endogenously determined as increasing functions of the relative scale of economic activities taking place in one of the locations (H in our example).

Through the years a rich list of 'second nature' forces has been proposed by geographers, regional scientists and urban economists.[5] These forces are also called 'agglomeration economies' and exist as long as the scale of the local economy adds to the performance of local firms. They are 'external economies' as long as the benefits of localized interactions are not fully reflected in the prices of market transactions.

Firms' and workers' location decisions determine market size and production cost asymmetries, and vice versa.

An important common implication of agglomeration economies is that they are able to generate self-sustaining clustering insofar as the movements of firms and workers, attracted to places with larger local markets and lower production costs, end up reinforcing these differences and thus spatial imbalances ('cumulative causation'). In this respect, agglomeration economies give strength to 'second nature' against 'first nature', detaching the emerging economic landscape from the physical attributes of its underlying geography. Thus, while there is *a priori* great flexibility on where particular activities locate, once the agglomeration process has started, spatial differences take shape and become quite rigid ('putty clay geography', see Fujita and Thisse 1996).

While sharing this common characteristic of making the spatial concentration of economic activities self-sustaining, agglomeration economies nonetheless differ substantially from one another in terms of two crucial features: Their 'scope' and their 'source'. The taxonomy below follows Rosenthal and Strange (2004).

3.1 Scope

Agglomeration economies may extend along three main dimensions. The notion of 'scope' refers to the extent to which each dimension is developed.

First, agglomeration economies may extend across industries in a certain location or be confined inside them. This difference in 'industrial scope' is the most familiar one in urban economics where

4 The distinction between 'first nature' and 'second nature' is due to Cronon (1991).

5 See Fujita and Thisse (2002) for a thorough assessment of the relative merits of the different approaches.

specific names have been assigned to the two polar cases. When agglomeration economies spread across industries, they are called 'urbanization economies'. When they are confined inside industries, they are called 'localization economies'.

Agglomeration economies have an industry, geographical, and time dimension.

The second dimension along which agglomeration economies may extend is the spatial one. This is the 'geographic scope' and refers to the extent to which external economies depend on the proximity between firms and thus decay with distance.

Finally, there is the 'temporal scope' that refers to the external economies generated by interactions among firms located in the same place but occurring at different times. A typical example is knowledge creation through cumulated local learning.

3.2 Sources

The notion of 'source' refers to a micro-founded explanation of the existence of agglomeration economies. Four explanations have attracted special attention.[6]

3.2.1 Marshall

Three explanations are known as the 'Marshallian triad': 'Knowledge spillovers', 'labour market pooling', and 'input sharing' (Marshall 1890).

'Knowledge spillovers' arise when knowledge is transferred between agents thanks to sheer physical proximity irrespective of market transactions between them. Knowledge, ideas and, above all, tacit information, can be considered as impure public goods that generate spillover effects from one firm or institution to another. Consequently, if economic agents possess different pieces of information, pooling them through informal communication channels can benefit everyone, hence the importance of proximity (Feldman 1994). In this perspective, agents co-locate to take advantage of knowledge that is someway 'in the air', which makes them more efficient. Accordingly, in terms of the theoretical framework of Section 2, the cost advantage of location H becomes an increasing function of the relative number of its resident agents.

'Labour market pooling' refers to two related phenomena that arise when firms and workers face search and matching frictions. First, the spatial concentration of workers with different skills and firms with different needs increases the likelihood of good matches and, hence, the expected quality of a match (Helsley and Strange 1990). Second, if employer-employee matches face an idiosyncratic risk of being undone, spatial concentration reduces the duration of unemployment spells and unfilled vacancies. Co-location allows firms and workers to benefit from both opportunities. Through these channels, in terms of the framework of Section 2, both the cost production advantage and the market size advantage of location H become increasing functions of the relative number of its resident agents.

'Input sharing' generates agglomeration economies when the production of intermediate inputs faces increasing returns to scale and their transportability is limited. When this is the case, the input producing sector is able to reach an efficient scale of production only when its local market is large enough, which requires the spatial concentration of downstream customers. Accordingly, in terms of the framework of Section 2, the cost advantage of location H becomes an increasing function of the relative number of agents residing in that location.

6 See Duranton and Puga (2004) for a theoretical survey of the microfoundations of agglomeration economies.

'Urban consumption opportunities' are sometimes considered as a fourth explanation of urban primacy. They are, however, partly a variation on the theme of input sharing and partly a variation on the theme of knowledge spillovers. As to the former, when the supply of final goods and services faces increasing returns to scale and their transportability is limited, large local demand associated with the spatial concentration of people allows final production to achieve an efficient scale. As to the latter, the spatial concentration of people fosters social interactions that may be valuable *per se* even in the absence of knowledge transmission. A similar argument is readily applied to the provision of all sorts of goods and facilities characterized by some degree of indivisibility (road, schools, *etc.*). As long as some of these are publicly provided through local funds, the spatial concentration of economic activities generates the tax base needed to finance them ('fiscal externality'). In terms of the framework of Section 2, the market size advantage of location H becomes an increasing function of the relative number of its resident agents.

Knowledge spillovers, labour market pooling, input sharing, and consumption amenities foster firm performance.

To summarise, in a Marshallian perspective the effects of spatial concentration on market size and production cost asymmetries may be due to:

- Knowledge spillovers that increase the production cost asymmetry;
- Labour market pooling that increases both market size and production cost asymmetries;
- Input sharing that increases the production cost asymmetry; and
- Consumption amenities that increase the market size asymmetry.

3.2.2 Krugman

While Marshallian explanations have a long standing tradition, more recently a specific approach, the so called 'new economic geography' (henceforth, NEG), has gained momentum in mainstream economics. The distinguishing feature of NEG with respect to alternative approaches is the focus on market rather than non-market interactions in a 'general equilibrium' framework stressing the endogenous determination of good and factor prices as well as the importance of economy-wide budget constraints.[7]

One way to relate NEG to the 'Marshallian triad' is to think of its contribution as bringing the idea of input sharing and urban consumption opportunities one step further by adding a richer understanding of industrial organization and its implication for factor prices. In doing so, imperfect competition plays centre stage, differently from Marshallian explanations that traditionally rely on economies of scale at the industry level and, hence, are compatible with the assumption of perfectly competitive product markets (as in Henderson 1974; Fujita and Ogawa 1982; Helsley and Strange 1990).

Two scenarios have received particular attention. Both stress the impact of firms' location decisions on other firms' profits. The first scenario considers the effect of firm relocation when matched by labour migration ('demand linkage', see Krugman 1991 as well as Ottaviano *et al.* 2002). In this case, as the firm moves, it reduces demand in the place of origin while increasing it in the place of destination. As profits rise with demand, the firm's relocation harms competitors in the place of origin and benefits competitors in the place of destination. In terms of the theoretical framework of Section 2, the market size advantage of location H becomes an increasing function of the relative number of agents residing in that location.[8]

---

7 The advantage of the general equilibrium approach is nicely summarized by Fujita and Krugman (2004, p. 141): "[…] you want a *general-equilibrium* story, in which it is clear where the money comes from and where it goes".

8 When agglomeration fosters capital accumulation, the outcome is similar to the one with migration since both expand the local market through the additional income they generate (Baldwin 1999 as well as Baldwin *et al.* 2001).

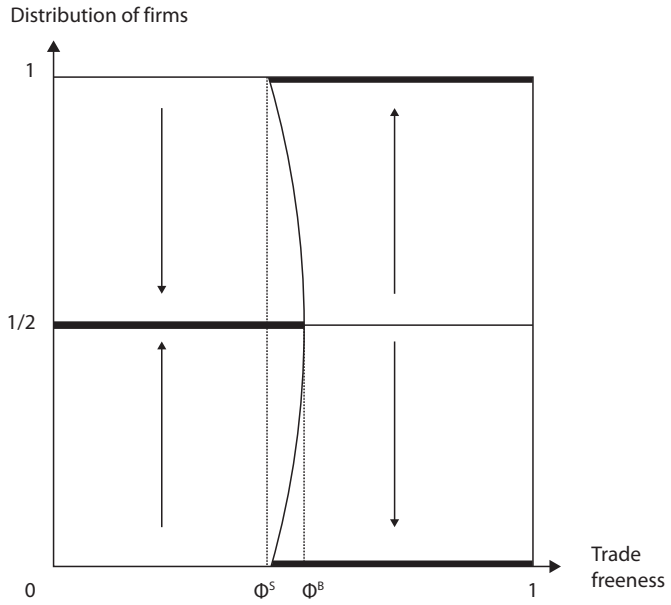
In the second scenario, firms have input-output linkages between each other: What is output for one firm is input for the other and *vice versa* ('cost linkage', see Krugman and Venables 1995; Venables 1996). When a firm relocates, it depresses both final demand and intermediate supply in the country of origin, whereas it reinforces them in the country of destination. Accordingly, other firms' profits suffer in the former and thrive in the latter. In terms of the theoretical framework of Section 2, the production cost advantage of location H becomes an increasing function of the relative number of its resident agents.

Demand and cost linkages determine the scale of the local economy – and firm performance.

The fact that, due to demand and cost linkages, the market size and production cost advantages of H grow with the local concentration of economic activities once more supports self-sustaining agglomeration. As already discussed, this is a common implication of agglomeration economies whatever their sources may be. The crucial contribution of NEG is that its solid microeconomic foundations allow the evolution of the spatial landscape to be related to observable microeconomic parameters such as the level of trade obstacles and the strength of firms' market power.

The following example illustrates how an understanding of the microeconomic foundations of NEG allows gaining sometimes unexpected insights. Consider the impact of trade barriers on clustering. The relation between these barriers and the spatial distribution of economic activities in simple NEG models is summarized in Figure 1.

Figure 1. The role of trade freeness



This figure portrays the possible long-run spatial configurations of our stylized two-location economy. It considers the special case in which there are neither market-size nor production cost asymmetries so that locations are identical in terms of 'first nature'. The extent of trade freeness is measured along the horizontal axis whereas the share of firms located in region H appears along the vertical one. The trade freeness index is an inverse measure of trade costs. When shipments are impossible, the trade freeness index is equal to its minimum value 0; when shipments are costless, the trade freeness index reaches its maximum at 1. A share of firms in H equal to 0 or 1 corresponds to agglomeration in locations F or H, respectively. Heavy solid lines indicate long-run outcomes, *i.e.*, geographical distributions of firms towards which the economy evolves as pointed out by the

vertical arrows. Figure 1 shows that, for low trade freeness (*i.e.*, values smaller than the threshold value φ^S), a dispersed geographical distribution of firms is the only long-run outcome. For high trade freeness (*i.e.*, values larger than the other threshold value φ^B), agglomeration in either location is the only long-run outcome. For intermediate values of trade freeness (*i.e.*, values between the two thresholds φ^S and φ^B) both dispersion and agglomeration can emerge in the long run. The values φ^B and φ^S are called 'break point' and 'sustain point' respectively: As trade freeness crosses φ^S from below agglomeration becomes 'sustainable' as a long-run outcome; as it crosses φ^B from below, the symmetric dispersion is 'broken'. The reason is that for large trade costs, the perfect symmetry of 'first nature' drives the location decisions of firms as these find it hard to serve both locations from a single production site. As trade costs fall, this becomes possible and 'second nature' leads to agglomeration.

Low trade costs lead to agglomeration when the supply of non-tradable goods and factors can easily adjust to demand.

Figure 1 is associated with models in which agglomeration does not affect the prices of non-traded goods and factors, such as land or, to some extent, unskilled labour. However, as firms cluster, the markets of non-tradables may be put under pressure on the demand side and their prices may therefore rise. When this happens, in terms of the framework developed in Section 2, the agglomerated location suffers a production cost disadvantage that increases in the share of its resident agents. If the price increase of non-tradables is strong enough, agglomeration may then be unwound. Whether this happens or not depends, among other things, on the level of trade barriers (Puga 1999).

To see this point, consider, for example, the implications of labour immobility. When workers are immobile, they play a double role. They give rise to localized labour supply, and their expenditures also generate localized final demand. For both reasons, as long as they are geographically dispersed, immobile workers lure firms away from congested areas. More generally, the anti-agglomeration effect of labour immobility is stronger the larger the share of immobile workers in the labour force (Krugman 1991).

Now enter trade barriers. Low trade barriers make it cheaper for firms to reach dispersed demand without producing locally. This weakens the anti-agglomeration effect of dispersed final demand. On the other hand, the level of trade barriers has no influence on the anti-agglomeration effect of dispersed labour supply, whose services are non-tradable by assumption. Accordingly, for high trade barriers, the clustering of supply in location H is hampered by the incentive that some firms still find in locating close to customers in location F. For low trade barriers, it is labour market pressure in location H that makes F an attractive production site. Accordingly, agglomeration is more likely to be sustainable for intermediate levels of trade barriers.

To summarise, NEG supplements the Marshallian perspective by suggesting that the market size and production cost effects of spatial concentration can be also due to:

- Demand linkages that increase the market size asymmetry; and

- Cost linkages that increase the production cost asymmetry.

In so doing, NEG highlights the complex interactions between goods and factors characterized by different degrees of tradability. In particular, rising demand for local non-tradables hampers agglomeration. If the supply of non-tradables is elastic enough to absorb any relevant impact of rising demand on their prices, agglomeration emerges when locations are highly integrated (low trade costs). In contrast, if higher demand maps into higher non-tradable prices, agglomeration emerges when locations are neither too isolated nor too highly integrated.

3.3 Welfare

Being driven by external economies, the economic landscape is inherently inefficient: The prices, on which consumers and firms base their consumption, production and location decisions, do not fully reflect the corresponding social values. Thus, proximity generates 'side effects' for which no quid-pro-quo is paid.

In the case of NEG, such side effects are associated with market transactions in an imperfectly competitive environment. Being intertwined with money transfers, they are called 'pecuniary externalities'. Marshallian sources belong instead to 'technological externalities'.[9] These are independent from any market interaction as they materialize through sheer physical proximity (Marshall 1890; Henderson 1978; Ciccone and Hall 1996). An example of positive pecuniary externality for a localized downstream industry is the fall in intermediate input prices due to the increase in upstream competition triggered by the entry of a new technologically advanced supplier; an example of positive technological externality is the increase in productivity that other upstream suppliers may experience through informal knowledge transmission generated by their proximity to the new technologically advanced rival.

Hence, whatever the sources of agglomeration economies, the implied geographical distribution of economic activities is generally inefficient from a social point of view.

4. The role of infrastructure

Different types of infrastructure may act on different sources of market size and production cost asymmetries.

In the theoretical framework developed in Section 2 the spatial concentration of economic activities is fostered by more pronounced market size asymmetries ('market seeking') and production costs asymmetries ('cost saving'). Concentration is also promoted by easier tradability of products and factor services between locations as well as by larger differentiation of firms in terms of products ('horizontal differentiation') and smaller differentiation in terms of quality/productivity ('vertical differentiation'). Once agglomeration economies are at work, Section 3 has argued that market size and production cost asymmetries become endogenous as they not only determine but are also determined by firms' location decisions. This section discusses how different types of infrastructure act on those asymmetries, thus affecting the evolution of the economic landscape.

4.1 Attraction and accessibility

To understand the impacts of different types of infrastructure on market size and production cost asymmetries one has to figure out the specific sources of agglomeration economies they affect. In so doing, one can exploit the fact that different sources naturally map into different geographical scopes.

An important distinction between the foregoing Marshallian and NEG sources is in terms of their geographical scope. In particular, the relative relevance of the two types of forces depends on the scale of the analysis (Ottaviano and Thisse 2001). Cities are replete with technological externalities (Anas *et al.* 1998). The same holds in local production systems (Pyke *et al.* 1990). Thus, to explain geographical clusters of somewhat limited spatial dimension such as cities and industrial districts, it seems reasonable to appeal to technological externalities that are the hallmark of Marshallian sources. However, when one turns to a larger geographical scale, it seems reasonable to think that direct physical contact provides a weaker explanation of interregional agglomerations such as the

9 The distinction between pecuniary and technological externalities is due to Scitovsky (1954).

'Manufacturing Belt' in the US and the 'Hot Banana' in Europe. This is the realm of NEG pecuniary externalities that arise from imperfect competition in the presence of market-mediated linkages between firms and consumers/workers.

Hence, while in the Marshallian perspective market size and production cost asymmetries only reflect the economic scale of local economic activity, in the NEG perspective those parameters also reflect the economic scale of all other connected locations from which inputs can be sourced and to which products can be sold. Two concepts can be usefully borrowed from spatial interaction theory to clarify this point (Smith 1975). In particular, if one visualizes the spatial economy as a network of interconnected markets, then according to Behrens *et al.* (2007b, 2008) the appeal of a market as a production site for firms depends on both its relative size ('attraction') and its relative centrality in the network of trading markets ('accessibility'). In this respect, the market size asymmetry and the production cost asymmetry would respectively measure the 'market seeking' and 'cost saving' dimensions of both attraction and accessibility.[10]

'Market potential' reflects that a location's appeal depends on own and other locations' market size and costs.

These two dimensions are embedded in a location's 'market potential'. This notion, introduced by Harris (1954) and recently refined by Head and Mayer (2004), has a nominal and a real definition. Whereas the 'nominal market potential' (henceforth, NMP) is a measure of customer proximity, the 'real market potential' (henceforth, RMP) is a combined measure of customer and competitor proximity. Formally, assume that location H is now only one of many alternative locations. The nominal market potential of H is then the weighted average of expenditures across all locations that plants can tap if located in H. In turn, the real market potential of H is the weighted average of real expenditures ('purchasing power') across all locations that plants can tap if located in H. In both cases, the weight of each location is a decreasing function of its distance from H due to trade barriers. The underlying idea is that NMP is a good proxy for total sales that plants can expect to make on average if located in H while RMP is a good proxy of the profits an average firm can make. In the long run, since firms can freely pick plant locations, profits should reach the same normal level everywhere. Therefore, in the long run RMP differences should eventually vanish as NMP differentials are capitalized in local price differences of non-tradables. Accordingly, short-run RMP differences should predict the future evolution of the economic landscape as firms are attracted towards areas temporarily boasting higher RMP.

4.2 Types of infrastructure

The distinction between attraction and accessibility is useful to classify the effects of different types of infrastructure on the differences in market size and production costs between locations.

A general definition of infrastructures is provided by the U.S. National Research Council (NRC), which adopts the term "public works infrastructures" to include "both specific functional modes - highways, streets, roads, and bridges; mass transit; airports and airways; water supply and water resources; wastewater management; solid-waste treatment and disposal; electric power generation and transmission; telecommunications; and hazardous waste management - and the combined system these modal elements comprise. A comprehension of infrastructure spans not only these public works facilities, but also the operating procedures, management practices, and development policies that interact together with societal demand and the physical world to facilitate the transport of people and goods, provision of water for drinking and a variety of other uses, safe disposal of

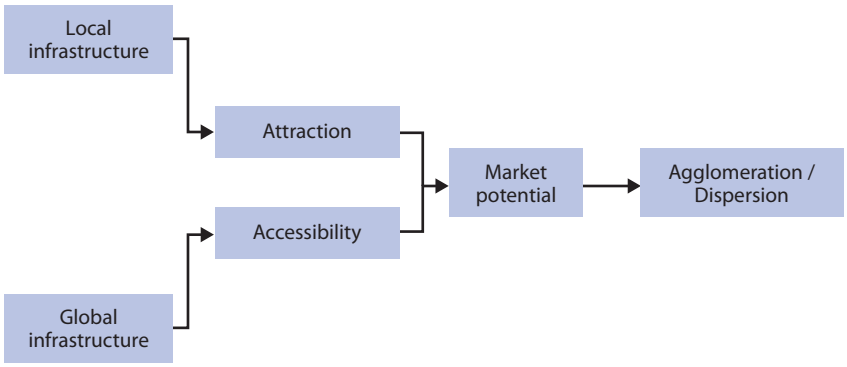
10 Attraction and accessibility are also the main ingredients of gravitational models of international trade (see *e.g.* Head and Mayer 2004).

society's waste products, provision of energy where it is needed, and transmission of information within and between communities" (NRC 1987, p. 4).

Such different types of infrastructure affect the strength of agglomeration economies differently depending on the corresponding sources. A crucial distinction is between 'local infrastructure' that mainly affects short-distance interactions and 'global infrastructure' that mainly affects long-distance interactions. The former mainly alters attraction whereas the latter mainly alters accessibility. While any classification is someway arbitrary, it is tempting to consider the provision of water for drinking and a variety of other uses, the safe disposal of society's waste products, the provision of energy where it is needed, the formation of human capital *etc*. as pertaining to local infrastructure, while the transportation of goods and people as well as the transmission of information as pertaining to both local and global infrastructure.

The foregoing considerations are synthetized in Figure 2, which summarizes the channels through which infrastructure affects the economic landscape in the economic geography literature.

Figure 2. Infrastructure and economic geography



4.3 Infrastructure and geographical disparities

Infrastructural changes may affect the geographical distribution of firms and workers between locations ('external geography') or within locations ('internal geography') (Martin and Rogers 1995).

4.3.1 External geography

When the focus is on the geographical distribution of firms and workers between locations ('external geography'), as far as NEG sources are concerned, the discussion in Section 4.1 implies that only infrastructural changes improving the market potential of a certain location are able to attract economic activities towards that location.

This has important (and someway unexpected) consequences. As a first consequence, improved global transport infrastructure between a developed location enjoying a market size advatange and a less developed one can decrease the attractiveness of the latter. This is called the 'straw effect' because economic activities migrate to developed locations through new infrastructure as juice in a glass is sucked up by a straw (Behrens *et al.* 2007b, 2008). The reason is easily understood in the wake of the simple two-location theoretical framework of Section 2, which shows that, as trade barriers fall, the dispersion push of competition weakens faster than the agglomeration pull of size differences. In other words, unless the prices of non-tradables are much higher in the developed

Better global infrastructure between rich and less developed locations can decrease the latter's attractiveness.

region, better transportation improves its market potential more than it improves the market potential of its less developed trading partner.

Improved local infrastructure does not always make a location more attractive.

Another, even more unexpected, consequence is captured by the 'shadow effect', according to which improved local transportation does not necessarily make a location more attractive. To see this, one has to move away from the simple two-location model (Martin 1999a, b; and Behrens *et al*. 2007b, 2008). For example, in the presence of a third location T, if this is large (*i.e*., it has strong 'attraction') and well connected to both H and F (*i.e*., it has good 'accessibility'), an increase in H's market size or a decrease in H's production costs due to better local transportation may well map into a decrease in its output share. The reason is that the improved local infrastructure of H is disproportionately used for shipments to and from T, which 'casts a shadow' on H's attractiveness. That would be the case, for instance, if location T were a transport 'hub' or 'gate'. A 'hub' is a location with better accessibility to all other locations; a 'gate' is a location through which goods mostly flow in and out of a region (Behrens *et al*. 2006). Favourable demand or cost shocks to any other location could result in supply expanding in the hub or in the gate and contracting elsewhere. Hence, clustering is more likely to take place in the presence of and close to hubs and gates (Krugman 1993; Behrens *et al*. 2006).

Better global infrastructure may nonetheless reduce geographical disparities. That would happen in three leading cases. First, as already discussed, if the prices of non-tradables are much lower in less developed locations, improved transport connections with developed locations result in firms and workers relocating to the less developed location (Puga 1999). Second, if better global infrastructure allows for long-distance commuting, the concentration of firms in developed regions is partly detached from local market size as workers spend their income elsewhere. This favours some dispersion of economic activities (Borck *et al.* 2007). Third, better global means of communication (*e.g.* improved ICT) foster the diffusion of local knowledge to distant places. Whenever knowledge spillovers are the main source of agglomeration economies, production cost asymmetries fall, thus promoting a more even economic geography (Baldwin *et al.* 2001).

To summarize, improved global infrastructure supports a more even distribution of economic activities when the prices of non-tradables are much lower in less developed locations, when it promotes long-distance commuting and when it is conducive to knowledge transmission from developed to less developed locations.

4.3.2 Internal geography

When the focus of the analysis is on the geographical distribution of firms and workers within locations ('internal geography'), one has, of course, to introduce some internal spatial dimension in the simple theoretical framework of Section 2. The simplest way is to think of a location as composed of many different sites. In this case, the presence of gates (*i.e*., sites within locations with a 'geographical advantage' in terms of better access to other locations) makes the internal geographies of locations interdependent. In particular, agglomeration in one location reduces the occurrence of agglomeration in the other one.

In addition, remoteness need not be a geographical disadvantage since a landlocked region may well be the location that attracts the larger share of firms. This is so when internal transport costs are high and, therefore, act as a barrier to competition from abroad (Behrens *et al*. 2006). Finally, when locations comprise many sites, one has also to keep in mind that labour mobility is easier within than between them. In this more complex scenario, agglomeration within locations is mainly shaped by internal trade impediments. *Vice versa*, agglomeration between locations is mainly shaped by

external trade barriers (Krugman and Livas 1996; Monfort and Nicolini 2000; Paluzie 2001; Crozet and Koenig-Soubeyran 2002; Behrens *et al*. 2006).[11]

To summarize, when improved local infrastructure affects the economic landscape within a region, this can indirectly change the internal landscape in other locations.

5. Empirical evidence

It is hard to tell whether infrastructure is a cause or an effect of local economic development.

The impact of infrastructure on economic performance is hard to disentangle from other concurrent effects. In particular, it is hard to figure out whether quality infrastructure is a cause or rather an effect of local economic development. Moreover, direct tests of the effects of infrastructure do not take explicitly into account the role of agglomeration economies, not to mention the specificities of their scope and sources.

This implies that what is currently available is a wide array of empirical findings that provide only partial pieces of information on the relevance of the theoretical framework discussed in the foregoing. Against this background, the present section reviews some of the most relevant findings. These, when taken together, seem to support the structure and the implications of the theoretical framework (Ottaviano and Pinelli 2005).

5.1 Growth and inequality

The empirical relevance of infrastructure for global and local economic development can be hardly overstated (World Bank 1994). The role of infrastructure has been stressed along two main dimensions: Effects on economic growth and effects on income inequality (Calderon and Serven 2004). As to the effects on growth, most studies focus on the impact of infrastructure on aggregate output and find this impact to be positive.[12] In particular, they identify positive and significant impacts on output of three types of infrastructure (telecommunications, transport, and energy) and show that such impacts are significantly higher than those of non-infrastructure capital (Calderon and Serven 2003).

The link between infrastructure and long-run growth is much less explored. Some studies find that public expenditure on transport and communications foster growth (Easterly and Rebelo 1993). This finding is also confirmed in the case of physical infrastructure (Sanchez-Robles 1998) and in the case of communications, measured by telephone density (Easterly 2001; Loayza *et al.* 2003). On the other hand, it is argued that sometimes the inefficiency of infrastructure provision can curb and even reverse the sign of the long-run growth impact (Devarajan *et al*. 1996; Hulten 1996; Esfahani and Ramirez 2003).

Turning to income inequality, the issue is whether infrastructure has a disproportionately positive impact on the income and welfare of the poor (World Bank 2003). The presence of a disproportionately positive impact finds some support in the existing evidence (see Brenneman and Kerf 2002 for a survey). Several studies point at the effects of infrastructure on human capital accumulation: Better transportation and safer roads promote school attendance; electricity allows more time for study and the use of computers; access to water and sanitation reduces child and maternal mortality.

11 Similar results hold true in the absence of interregional migration whenever firms are linked by strong input-output ties (Puga and Venables 1997; Monfort and Van Ypersele 2003).

12 This is highlighted in a seminal contribution by Aschauer (1989), who finds that the stock of public infrastructure capital is a significant driver of aggregate TFP. Even though subsequent efforts question Aschauer's quantitative assessment, overall his qualitative insight survives more sophisticated econometric scrutiny (see *e.g.* Gramlich 1994; Röller and Waverman 2001).

Infrastructure also connects poor people in underdeveloped areas to core economic activities, thus expanding their employment opportunities (Estache 2003). Finally, better infrastructure in poorer regions reduces production and transaction costs (Gannon and Liu 1997).

Efficient supply of the right infrastructure in the right place is more important for growth than the amounts spent.

Overall, existing studies show that infrastructure is an important determinant of economic growth and income inequality. The exact impact may depend, however, on the type of infrastructure. All in all, there is a broad consensus that: "Infrastructure is a necessary but not sufficient ingredient of economic growth, and (…) the efficient supply of the right kind of infrastructure (material and institutional) in the right place is more important than the amount of money disbursed or the pure quantitative infrastructure capacities created" (Sugolov *et al*. 2003, p. 3).

This is consistent with the discussion in Section 4 insofar as the impacts of different types of infrastructure depend on the specific scope and sources of agglomeration economies.

5.2 Spillovers

Knowledge spillovers are the Marshallian source that has by far received most attention. Two main research strategies have been implemented to assess their relevance. A first strategy exploits information that can be indirectly extracted from wage and price variations across locations. A second strategy measures the presence of spillovers directly in terms of knowledge creation.

5.2.1 Wage and rent gradients

Localized spillovers make firms and workers more productive when geographically clustered. As firms and workers move to take advantage of differential productivities, these end up being entirely capitalized in local price differences that exactly match the geographical variation of wages due to productivity differences. Therefore, if wages and prices were higher in areas with a higher density of firms and workers, there would be evidence of a productivity-enhancing spillover. Moreover, if wages and prices were found to be positively associated with the density of human capital, there would be specific evidence of a productivity-enhancing knowledge spillover.

Both skilled and unskilled wages do tend to be higher in locations where the labour force is more educated, and the quantitative effect is not negligible. A one-year increase in local average education raises the average wage by 3 to 5 percent. A one percentage point rise in the local share of college educated workers increases the average wage by 0.6 to 1.2 percent. At the same time, the presence of more educated workers is associated with higher local prices. As discussed above, this is supportive evidence for the presence of productivity-enhancing knowledge spillovers (Rauch 1993; Moretti 2004).

Some understanding of the channels through which knowledge transmission happens can be obtained from the behaviours of young and old workers. Even though the former earn less than the latter in denser areas such as cities, they tend to be over-represented in those areas. As argued by Peri (2002), a possible explanation of the fact that young workers accept lower wages in denser areas is that they value the learning opportunities density offers. Differently, as people get older, the expected return to learning decreases, giving more weight to the congestion costs associated with density. This causes younger workers to move to denser areas and older workers to move to less dense areas (selection effects).

The study of wage gradients also allows understanding why the empirical literature devotes special attention to knowledge spillovers when compared with other Marshallian sources. In decreasing

order of importance, learning spillovers, better matching between firms and workers, and selection effects are all responsible of the wage structure observed in denser areas (Glaeser and Maré 2001; Combes *et al*. 2004).

Thus, wages and rents are positively correlated with the geographical variation in the density of economic activities, which can be interpreted as the result of productivity-enhancing spillovers in denser areas.

Finally, some studies have also investigated the geographical scope of spillovers. Conley *et al.* (2003) reach the conclusion that knowledge transmission between two individuals vanishes starting from 90-minute-trip distances.

5.2.2 Knowledge creation

The second strategy to spillover measurement focuses directly on the process of knowledge creation. This is modelled through a 'knowledge production function' that explains the output of innovation such as patents in terms of knowledge inputs such as R&D spending and human capital (see Audretsch and Feldman 2004 for a detailed survey). When brought to data, this simple mechanism works at the level of areas and industries, but it does not work at the level of the individual firm. This is particularly evident in the case of small firms that are able to generate innovative output with negligible amounts of R&D. Such phenomenon is interpreted as evidence of the existence of knowledge spillovers that allow the firms in an area to benefit from research carried out by other institutions (universities or firms) located in the same area (Acs *et al.* 1994).

Using this research strategy, the study of the location pattern of patent families (*i.e.*, patents that reference or cite each other) allows one to establish the geographical scope of knowledge spillovers. The corresponding findings are in line with those obtained in the case of wage and rent gradients, revealing that the positive impact of spillovers fades away quite rapidly with distance. As the probability of cross-citation is much higher when inventors come from the same area, cross-fertilization is highly localized (Jaffe *et al.* 1993; Jaffe and Trajtenberg 2002). Thus, knowledge spillovers fade away quite rapidly with distance.

Knowledge spillovers fade away quite rapidly with distance.

Finally, some studies measure the overall impact of knowledge spillovers on plant productivity. According to the benchmark estimate, each year the contribution of spillovers to aggregate output growth is 0.1 percent. This estimate is essentially driven by high-tech plants since it is virtually zero for low-tech plants (Moretti 2002).

To summarize, the overall evidence is quite supportive of Marshallian sources.

5.3 Market potential

Turning to NEG sources, the discussion in Section 4.1 has highlighted the key role of market potential. In particular, it argued that demand and costs linkages attract firms and workers to places with higher 'market potential'.[13] As in the case of knowledge spillovers, the associated geographical advantage is capitalized in local price differences that exactly compensate the spatial variation of

13 Head and Mayer (2005) compare alternative measures of real and nominal market potential. Complex measures lead to results that are essentially the same as the ones associated with the simple measure devised by Harris (1954), *i.e.*, distance-weighted average expenditures of connected markets. That is why, to alleviate the reading, the distinction between NMP and RMP is dropped.

wages due to differential production and consumption amenities. This suggests two natural tests of the empirical validity of NEG arguments (see Head and Mayer 2004 for a survey). On the price side, higher market potential should be associated with higher wages and higher prices. On the quantity side, higher market potential should attract both firms and workers.

5.3.1 Price effects

Market potential variations explain about 35 percent of the cross-country income variation.

The price predictions have been tested at both international and interregional levels. In cross-country studies market potential variations explain around 35 percent of the cross-country income variation. This result does not depend on institutions, natural resources, and physical geography. Interestingly, a country's access to the coast raises the local nominal wage by over 20 percent, which reveals the dominant role of gate regions (see *e.g.* Redding and Venables 2000).

The prediction that higher market potential should be associated with both higher wages and higher land rents finds clear support in cross-region studies for the US (*e.g.* Hanson 1998). These also highlight the dominant role of transport hubs and gates. For example, a 10 percent increase of the distance from them reduces the nominal wage by 1-2 percent in Mexico (Hanson 1997).

To sum up, wages and rents are higher in locations with higher market potential, which can be interpreted as showing that demand and cost linkages enhance local productivity.

5.3.2 Quantity effects

The quantity predictions stem from the idea that local shocks to final demand or intermediate supply generate compensating movements by firms and workers. As for firms, most studies focus on foreign direct investment as this is considered the relatively footloose part of their activities.[14] A general conclusion is that foreign firms indeed favour locations with higher market potential: A 10 percent rise in market potential yields a 10.5 percent rise in the probability that a location is chosen by foreign investors.

As to workers, studies measuring the impact of customer and supplier proximity are scarce. The few existing ones suggest that migrants do respond to market potential differentials in the predicted way. Their response is nonetheless limited by distance, which may signal the presence of mobility costs and migration barriers. For example, in a study for European regions Crozet (2000) shows that a region with 100 km radius attracts workers within a radius of no more than 120 kilometres.

To summarize, the empirical literature that closely matches the theoretical predictions based on market potential is still quite thin. Nonetheless, the existing results lend support to the implications of NEG.

5.4 Trade barriers

NEG arguments imply a non-linear relation between trade barriers and the geographical concentration of economic activities: When trade costs are either high or low, economic activities are dispersed; when trade costs are intermediate, economic activities cluster (see Section 3.2).

---

14 Coughlin *et al.* (1991) study the location decision of all foreign investors across US states. Head *et al.* (1999) concentrate on Japanese firms only. Head and Mayer (2002) analyze the behaviour of Japanese firms across European regions.

Trade costs have declined over time due to both improvements in the transport technology and, after World War II, reductions in tariffs. Some scholars have, therefore, tried to infer the impact of trade costs on agglomeration from the evolution of industrial location over time. For example, the geographical concentration of manufacturing across US states fell until 1900, then rose to a climax around 1927, and finally fell again until 1987 when it reached its level in 1860 (Kim 1995). The geographical concentration of manufacturing across EU countries rose sharply between 1972 and 1996 but slowed down after the implementation of the Single Market Programme in 1986 (Brülhart 2001). While these patterns are broadly consistent with NEG predictions, they can hardly be interpreted as evidence of any clear-cut impact of trade costs on agglomeration since many other variables are likely to have affected industry location over time.

A more direct approach is implemented by a so-called 'concentration regression', which regresses alternative indices of geographical concentration on various measures of 'trade costs' such as administrative barriers, geographical size (with larger areas implying greater average distance), expenditure on transport and communication as well as road/railway/communication density.[15] Existing analyses are typically cross-country. Studies on the effects of external trade barriers on cross-country agglomeration are inconclusive and their results somewhat contradictory (Combes and Overman 2004). Studies on the effects of internal and external trade barriers on within-country agglomeration have produced more clear-cut results, showing that agglomeration is more pronounced when both external and internal interactions are more costly. This is consistent with NEG as long as the average integration of the sampled countries is low enough (Ades and Glaeser 1995; Rosenthal and Strange 2001).

Empirically, agglomeration is more pronounced for intermediate than for high or low trade costs.

Clearly, the main challenge in assessing whether there exists a non-linear relation between trade barriers and agglomeration comes from the fact that it is hard to tell whether the observed level of trade barriers is 'high', 'low' or 'intermediate'. An interesting way of circumventing this problem has been recently proposed in a study that represents the first attempt to explicitly investigate the impact of transport policy on industry location within a NEG framework (Teixeira 2006). The proposed approach is a mixture of regression and simulation analyses. In particular, using data on Portugal over the period 1985 to 1998, regressions confirm the empirical relevance of the underlying NEG theoretical framework over different periods of time.[16] They also find that Portuguese transport policy has not contributed to spatial equity. However, the simulation of a further planned expansion of the transport network shows that, if transport costs are lowered sufficiently, industry will eventually spread. This suggests a bell-shaped relationship between transport costs and agglomeration, as suggested by theory.

The existing empirical evidence on the impact of infrastructure on economic performance may be summarised with the words of Rosenthal and Strange (2004, p. 2160): "Taking all of these results together, an interesting pattern emerges, with industry attributes sensitive to shipping costs (reliance on manufactured inputs, reliance on natural resource inputs, marketing of perishable products) influencing agglomeration at the state level, knowledge spillovers impacting highly localized agglomeration, and labour impacting agglomeration at all levels of geography".

15 Additional controls are introduced for the potential impact of other variables (such as development stage, industrial composition, and institutions).

16 Holl (2004) reaches a similar conclusion in the case of Spain.

5.5. An illustration: Portuguese public investment in transport infrastructure

The following analysis of the Portuguese case, drawing on a study by Teixeira (2006),[17] provides a convincing and concrete example of how the theoretical framework discussed in the present essay can be effectively used to evaluate and forecast the effects of transport policies on the spatial distribution of economic activities.

The Portuguese case exhibits several features of an ideal natural experiment. From 1985 to 1998 Portuguese public investment in transport infrastructure represents more than 1.9 percent of GDP. More than 70 percent of such investment is related to road networks. Moreover, in those years transport infrastructure absorbs 13.5 percent of the European structural funds received by Portugal, 62.2 percent of which allocated to road networks, increasing the motorway network from 234 to 1393 kilometres.

The analysis implements an empirical NEG model of the Portuguese economy, with a breakdown of 18 regions and 25 sectors, with the aim of explaining the evolution of employment across regions and industries. It estimates the importance of transport costs for interregional trade by industry and relates their variation to sectoral indexes of spatial concentration of employment. Reliance on very detailed data accounting for several types of congestion (such as the percentage of long vehicles in traffic) is used to compute the total distance and time costs between districts. As a result, one finds that Portuguese transport costs have fallen dramatically (45 percent on average) among the Portuguese districts over the reference period.

Massive infrastructure investment in Portugal has not reduced interregional income inequalities.

Contrary to the Portuguese authorities' expectations, however, such a massive decrease in transport costs has not reduced regional imbalances. The reason lies in the specific details of the infrastructural project and the specific sources of agglomeration economies. These, being of the NEG type, give a strong 'network character' to the Portuguese economic space.

That NEG forces are at work is suggested by the sectoral pattern of agglomeration. From 1985 to 1998, while transport costs fall in all sectors and spatial concentration rises for Portugal as a whole, only 12 out of 25 sectors experience increased agglomeration. In particular, as predicted by NEG, these are sectors with significant increasing returns to scale and inter-industry linkages, such as the high-tech industries (*e.g.* Medical-surgical and optical, Electronic machinery, Equipment for treatment of information) and some capital goods industries (*e.g.* Basic metallurgy, Metal products, and Machinery industries). On the contrary, sectors with limited returns to scale and weak inter-industry linkages tend to become more dispersed (*e.g.* Textile, Clothing, and Leather and footwear).

That the specific details of the project (including the original status quo) matter is testified by the simulated effects of further plans by the Portuguese Ministry of Transports for the transport road network design in 2010. Again the planned investment is huge implying overall transport cost savings of about 42 percent from 1998 to 2010. Nevertheless, the analysis predicts that this time transport cost reductions would lead to substantial spatial dispersion of economic activity with 21 out of 25 sectors reducing their agglomeration.

17 A related study is that of Venables and Gasiorek (1999) who use a calibrated general equilibrium (CGE) model to investigate the effects of road projects in Cohesion countries.

6. Conclusion

This essay has provided an overview of the role of infrastructure in affecting economic geography in the light of both theoretical and empirical findings. As simple as it may sound, the single most important lesson is that infrastructural improvements affect the geographical distribution of economic activities. This casts a shadow on traditional cost-benefit analysis as this usually does not try to assess the impact of infrastructure projects on regional economic development (Puga 2002).

Infrastructure affects economic geography in various (and sometimes unexpected) ways. Most naturally, interregional infrastructure affects interregional location. However, it also affects intraregional location. For example, with the development of a higher order transport network, such as the Trans-European Networks (TENs), intraregional reallocations are becoming increasingly pronounced depending on differential access to the new networks (Vickerman 1995). Analogously, intraregional infrastructure affects both intraregional and interregional location. In particular, infrastructural changes in a certain region affect not only its internal geography but, precisely because they affect own internal geography, they may also alter the internal geography of connected regions.

The reason is the 'network character' of the space economy itself, which implies that the attractiveness of locating in a certain site depends not only on its market size and production cost advantages but also on its centrality with respect to alternative sites (no matter whether they belong to the same region or not). It is this network feature of attractiveness that underlies the concept of 'market potential'.

The market potential is a rather successful tool to predict the reallocations of economic activities triggered by infrastructural investment. This is true even in the case of its most unexpected implications such as the 'straw effect'. For example, in France there is some informal evidence that the construction of the Paris-Lyon high-speed rail line led to the relocation of headquarters from Lyon to Paris (Puga 2002). In Spain there are concerns that the Madrid-Barcelona high-speed rail line may reinforce the process of headquarters relocation towards Madrid (Vives 2001). In Italy it has been argued that the reduction in transport costs between North and South in the 1950s accelerated the deindustrialization process of the *Mezzogiorno* (Faini 1983).

A second important lesson is, therefore, that infrastructure investment generates externalities that may diffuse quite far. As this essay has shown, this may be true even for localised infrastructure projects.

Infrastructure investment generates far-diffusing externalities, hence the need for interregional coordination.

For example, there is evidence that, as economic activities relocate, the benefits that a region obtains from improved own intraregional infrastructure come at the expense of competing regions (*e.g.* see Boarnet 1998 for California). Similar evidence is found in the case of interregional infrastructure as earnings rise in regions receiving new national highways and fall in adjacent regions.[18]

These two lessons have two far reaching policy implications. First, effective infrastructural projects require knowledge on their impacts on the spatial distribution of economic activities. These impacts depend crucially on the specific details of the projects and the specific sources of agglomeration economies they affect. Second, regions need to coordinate not only interregional infrastructure projects but also intraregional ones to avoid beggar-thy-neighbour and self-defeating outcomes.

18 See, for example, Chandra and Thompson (2000) for a study on earnings impacts of interstate highways in US non-metropolitan counties. Additional evidence on infrastructure externalities is surveyed by Holl (2004).

Effective infrastructure projects require knowledge on the spatial impacts on economic activities.

A concrete example to see these implications at work comes from the analysis of the impact of Portuguese public investment in transport infrastructure on regional disparities mentioned in Section 5.5. This analysis convincingly shows how the theoretical framework discussed in the present essay can be effectively used to evaluate and forecast the effects of specific transport policies on the spatial distribution of economic activities.

To sum up, not only effective infrastructural projects require knowledge of their impacts on the spatial distribution of economic activities but there is also evidence that NEG provides a useful means of developing that knowledge. Therefore, NEG concepts and modelling strategies should supplement the standard toolbox of impact studies on infrastructural projects.

References

Acs, Z., Audretsch, D., and Feldman, M. (1994). "R&D spillovers and recipient firm size". *Review of Economics and Statistics*, (76:2), pp. 336-340.

Ades, A. and Glaeser, E. (1995). "Trade and circuses: Explaining urban giants". *Quarterly Journal of Economics*, (110:1), pp. 195–227.

Anas, A., Arnott, R., and Small, K. (1998). "Urban spatial structure". *Journal of Economic Literature*, (36:3), pp. 1426-1464.

Aschauer, D. (1989). "Is public expenditure productive?" *Journal of Monetary Economics*, (23:2), pp. 177-200.

Audretsch, D. and Feldman, M. (2004). "Knowledge spillovers and the geography of innovation", in Henderson, V. and Thisse, J-F. (eds.), *Handbook of regional and urban economics*, Vol. 4, North-Holland, Amsterdam, The Netherlands.

Baldwin, R. (1999). "Agglomeration and endogenous capital". *European Economic Review*, (43:2), pp. 253-280.

Baldwin, R., Martin, P., and Ottaviano, G. (2001). "Global income divergence, trade and industrialization: The geography of growth take-off". *Journal of Economic Growth*, (6:1), pp. 5-37.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., and Robert-Nicoud, F. (2003). *Economic geography and public policy*, Princeton University Press, Princeton, USA.

Behrens, K., Gaigné, C., Ottaviano, G., and Thisse, J.-F. (2006). "Is remoteness a locational disadvantage?" *Journal of Economic Geography*, (6:3), pp. 347-368.

Behrens, K., Mion, G., and Ottaviano, G. (2007a). "Industry reallocations in a globalizing economy". CEPR, Discussion Paper No. 6049.

Behrens, K., Lamorgese, A., Ottaviano, G., and Tabuchi, T. (2007b). "Testing the 'home market effect' in a multi-country world". CORE-UCL, Louvain, mimeo.

Behrens, K., Lamorgese, A., Ottaviano, G., and Tabuchi, T. (2008). "Changes in transport and non-transport costs: Local vs. global impacts in a spatial network". *Regional Science and Urban Economics*, forthcoming.

Boarnet, M. (1998). "Spillovers and the locational effects of public infrastructure". *Journal of Regional Science*, (38:3), pp. 381– 400.

Borck, R., Pflüger, M., and Wrede, M. (2007). "A simple theory of industry location and residence choice". IZA Discussion Paper No. 2862.

Brenneman, A. and Kerf, M. (2002). "Infrastructure and poverty linkages: A literature review". The World Bank, mimeo.

Brülhart, M. (2001). "Evolving geographical specialisation of European manufacturing industries". *Review of World Economics*, (137:2), pp. 215-243.

Calderon, C.A. and Serven, L. (2003). "The output cost of Latin America's infrastructure gap", in Easterly W. and Serven, L. (eds.), *The limits of stabilization: Infrastructure, public deficits, and growth in Latin America*. Stanford University Press, Stanford, USA.

Calderon, C.A. and Serven, L. (2004). "The effects of infrastructure development on growth and income distribution". World Bank Policy Research Working Paper No. 3400.

Chamberlain, E.H. (1993). *The theory of monopolistic competition.* Cambridge, MA, USA.

Chandra, A. and Thompson, E. (2000). "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system". *Regional Science and Urban Economics*, (30:4), pp. 457-490.

Ciccone, A. and Hall, R. (1996). "Productivity and the density of economic activity". *American Economic Review*, (87:1), pp. 54-70.

Combes, P-P., Duranton, G., and Gobillon, L. (2004). "Spatial wage disparities: Sorting matters!". CEPR Discussion Paper No. 4240.

Combes, P-P. and Overman, H. (2004). "The spatial distribution of economic activities in the European Union", in Henderson, V. and Thisse, J.-F. (eds.), *Handbook of regional and urban economics*, Vol. 4, North-Holland, Amsterdam, The Netherlands.

Conley, T., Flyer, F., and Tsiang, G. (2003). "Spillovers from local market human capital and the spatial distribution of productivity in Malaysia". *Advances in Economic Analysis and Policy*, (3.1), p. 1229.

Coughlin, C., Terza, J., and Arromdee, V. (1991). "State characteristics and the location of foreign direct investment in the United States". *Review of Economics and Statistics*, (73:4), pp. 675-683.

Cronon, W. (1991). *Nature's megalopolis: Chicago and the Great West*, Norton, New York, USA.

Crozet, M. (2000). "Do migrants follow market potential? An estimation of a new economic geography model". Cahier de la MSE-Serie Blanche No. 2000-30.

Crozet, M. and Koenig-Soubeyran, P. (2002). "Trade liberalization and the internal geography of countries". CREST Discussion Paper No. 2002-37.

Devarajan, S., Swaroop, V., and Zhou, H. (1996). "The composition of public expenditure and economic growth". *Journal of Monetary Economics*, (37:2-3), pp. 313-344.

Duranton, G. and Puga, D. (2004). "Micro-foundations of urban agglomeration economies", in Henderson, V. and Thisse, J.-F. (eds.), *Handbook of regional and urban economics*, Vol. 4, North-Holland, Amsterdam, The Netherlands.

Easterly, W. (2001). "The lost decade: Developing countries' stagnation in spite of policy reform". The World Bank, mimeo.

Easterly, W. and Rebelo, S. (1993). "Fiscal policy and economic growth: An empirical investigation". *Journal of Monetary Economics*, (32:3), pp. 417-458.

Esfahani, H. and Ramirez, M.T. (2003). "Institutions, infrastructure and economic growth". *Journal of Development Economics*, (70:2), pp. 443-477.

Estache, A. (2003). "On Latin America's infrastructure privatization and its distributional effects". The World Bank, mimeo.

Faini, R. (1983), "Cumulative process of deindustrialization in an open region: The case of Southern Italy, 1951–73". *Journal of Development Economics*, (12:3), pp. 277–301.

Feldman, M. (1994), *The geography of innovation*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Fujita, M. and Krugman, P. (2004). "The new economic geography: Past, present and future". *Papers in Regional Science*, (83:1), pp. 139-164.

Fujita, M. and Ogawa, H. (1982). "Multiple equilibria and structural transition of non-monocentric urban configurations". *Regional Science and Urban Economics*, (12:2), pp. 161-196.

Fujita, M. and Thisse, J-F. (1996). "Economics of agglomeration". *Journal of the Japanese and International Economies*, (10:4), pp. 339-378.

Fujita, M. and Thisse, J.-F. (2002). *Economics of agglomeration: Cities, industrial location and regional growth*, Cambridge University Press, Cambridge (Mass.), USA.

Gannon, C. and Liu, Z. (1997). "Poverty and transport". The World Bank, mimeo.

Glaeser, E. and Maré, D. (2001). "Cities and skills". *Journal of Labor Economics*, (19:2), pp. 316-342.

Gramlich, E. (1994). "Infrastructure investment: A review essay". *Journal of Economic Literature*, (32:3), pp. 1176-1196.

Hanson, G. (1997). "Increasing returns, trade and the regional structure of wages". *Economic Journal*, (107:440), pp. 113-133.

Hanson, G. (1998). "Market potential, increasing returns, and geographic concentration". NBER Working Paper No. 6429.

Harris, C. (1954). "The market as a factor in the localization of industry in the United States". *Annals of the Association of American Geographers*, (44:4), pp. 315-348.

Head, K. and Mayer, T. (2002). "Market potential and the location of Japanese investment in the European Union". CEPR Discussion Paper No. 3455.

Head, K. and Mayer, T. (2004). "The empirics of agglomeration and trade", in Henderson, V. and Thisse, J.-F. (eds.), *Handbook of regional and urban economics*, Vol. 4, North-Holland, Amsterdam, The Netherlands.

Head, K. and Mayer, T. (2005). "Regional wage and employment responses to market potential in the EU". CEPR Discussion Paper No. 4908.

Head, K., Ries, J., and Swenson, D. (1999). "Attracting foreign manufacturing: Investment promotion and agglomeration". *Regional Science and Urban Economics*, (29:2), pp. 197–218.

Helsley, R. W. and Strange, W. C. (1990). "Agglomeration economies and matching in a system of cities". *Regional Science and Urban Economics*, (20:2), pp. 189-212.

Henderson, J. V. (1974). "The sizes and types of cities". *American Economic Review*, (64:4), pp. 640–656.

Henderson, V. (1978). *Economic theory and the cities*, Academic Press, London, UK.

Henderson, V. and Thisse, J.-F. (2004). *Handbook of regional and urban economics*, Vol. 4, North-Holland, Amsterdam, The Netherlands.

Holl, A. (2004). "Manufacturing location and impacts of road transport infrastructure: Empirical evidence from Spain". *Regional Science and Urban Economics*, (34:3), pp. 341-363.

Hulten, C. (1996). "Infrastructure capital and economic growth: How well you use it may be more important than how much you have". NBER Working Paper No. 5847.

Jaffe, A. and Trajtenberg, M. (2002). *Patents, citations and innovation: A window on the knowledge economy*, MIT Press, Cambridge (Mass.), USA.

Jaffe, A., Trajtenberg, M., and Henderson, R. (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations". *Quarterly Journal of Economics*, (108:3), pp. 577-598.

Kim, S. (1995). "Expansion of markets and the geographic distribution of economic activities: The trends in US regional manufacturing structure, 1860-1987". *Quarterly Journal of Economics*, (110:4), pp. 881-908.

Krugman, P. (1991). "Increasing returns and economic geography". *Journal of Political Economy*, (99:3), pp. 483-499.

Krugman, P. (1993), "The hub effect: Or, threeness in international trade", in Ethier W., Helpman, E., and Neary, P. (eds.), *Theory, policy and dynamics in international trade*, Cambridge University Press, Cambridge, UK.

Krugman, P. (1995). *Development, geography, and economic theory*, MIT Press, Cambridge (Mass.), USA.

Krugman, P. and Livas, R. (1996). "Trade policy and the third world metropolis". *Journal of Development Economics*, (49:1), pp. 137-150.

Krugman, P. and Venables, A. (1995). "Globalization and the inequality of nations". *Quarterly Journal of Economics*, (110:4), pp. 857-880.

Loayza, N., Fajnzylber, P., and Calderon, C. (2003). "Economic growth in Latin America and the Caribbean: Stylized facts, explanations and forecasts", The World Bank, mimeo.

Marshall, A. (1890). *Principles of economics*, Macmillan, London, UK.

Martin, P. (1999a). "Are European regional policies delivering?" *EIB Papers*, (4:2), pp. 10-23.

Martin, P. (1999b). "Public policies, regional inequalities and growth". *Journal of Public Economics*, (73:1), pp. 85-105.

Martin, P. and Rogers, C. (1995). "Industrial location and public infrastructure". *Journal of International Economics*, (39:3-4), pp. 335-351.

Monfort, P. and Nicolini, R. (2000). "Regional convergence and international integration". *Journal of Urban Economics*, (48:2), pp. 286-306.

Monfort, P. and Van Ypersele, T. (2003). "Integration, regional agglomeration and international trade". CEPR Discussion Paper No. 3752.

Moretti, E. (2002). "Workers, education, spillovers and productivity: Evidence from plant-level production function", UCLA, mimeo.

Moretti, E. (2004). "Human capital externalities in cities", in Henderson, V. and Thisse, J.-F., *Handbook of regional and urban economics*, Vol. 4, North-Holland, Amsterdam, The Netherlands.

National Research Council (NRC) (1987). *Infrastructure for the 21st century: Framework for a research agenda*, National Academy Press, Washington, D.C., USA.

Ottaviano, G., Tabuchi, T., and Thisse, J.-F. (2002). "Agglomeration and trade revisited". *International Economic Review*, (43:2), pp. 409-435.

Ottaviano, G. and Pinelli, D. (2005). "A 'new economic geography' perspective on globalization". *Italian Journal of Regional Science*, (4:1), pp. 71-106.

Ottaviano, G. and Thisse, J.-F. (2001). "On economic geography in economic theory: Increasing returns and pecuniary externalities". *Journal of Economic Geography*, (1:2), pp. 153-179.

Paluzie, E. (2001). "Trade policies and regional inequalities". *Papers in Regional Science*, (80:1), pp. 67-85.

Peri, G. (2002). "Young workers, learning, and agglomeration". *Journal of Urban Economics*, (52:3), pp. 582-607.

Puga, D. (1999). "The rise and fall of regional inequalities". *European Economic Review*, (43:2), pp. 303-334.

Puga, D. (2002). "European regional policy in light of recent location theories". *Journal of Economic Geography*, (2:4), pp. 372-406.

Puga, D. and Venables, A. (1997). "Preferential trading arrangements and industrial location". *Journal of International Economics*, (43:3), pp. 347-368.

Pyke, F., Becattini, B., and Sengenberger, W. (1990). *Industrial districts and inter-firm cooperation in Italy*, International Institute for Labour Studies, Geneva, Switzerland.

Rauch, J. (1993). "Productivity gains from geographic concentration of human capital: Evidence from the cities". *Journal of Urban Economics*, (34:3), pp. 380-400.

Redding, S. and Venables, A. (2000). "Economic geography and international inequality". CEPR Discussion Paper No. 2568.

Röller, L-H. and Waverman, L. (2001). "Telecommunications infrastructure and economic development: A Simultaneous Approach". *American Economic Review*, (91:4), pp. 909-923.

Rosenthal, S. and Strange, W. (2001). "The determinants of agglomeration". *Journal of Urban Economics*, (50:2), pp. 191-229.

Rosenthal, S. and Strange, W. (2004). "Evidence on the nature and sources of agglomeration economies", in Henderson, V. and Thisse, J.-F. (eds.), *Handbook of regional and urban economics*, Vol. 4, North-Holland, Amsterdam, The Netherlands.

Sanchez-Robles, B. (1998). "Infrastructure investment and growth: Some empirical evidence". *Contemporary Economic Policy*, (16:1), pp. 98-108.

Scitovsky, T. (1954). "Two concepts of external economies". *Journal of Political Economy*, (62:2), pp. 143-151.

Scotchmer, S. and Thisse, J.-F. (1992). "Space and competition: A puzzle". *Annals of Regional Science*, (26:2), pp. 269-286.

Smith, T. (1975). "A choice theory of spatial interaction". *Regional Science and Urban Economics*, (5:2), pp. 137-176.

Starrett, D. (1978). "Market allocations of location choice in a model with free mobility". *Journal of Economic Theory*, (17:1), pp. 21-37.

Sugolov, P., Dodonov, B., and von Hirschhausen, C. (2003). "Infrastructure policies and economic development in East European transition countries: First evidence". DIW Berlin WP-PSM-02.

Teixeira, A. (2006). "Transport policies in light of the new economic geography: The Portuguese experience". *Regional Science and Urban Economics*, (36:4), pp. 450–466.

Venables, A. (1996). "Equilibrium locations of vertically linked industries". *International Economic Review*, (37:2), pp. 341-359.

Venables, A. and Gasiorek, M. (1999). "Evaluating regional infrastructure: A computable equilibrium approach", in European Commission (ed.), *The socio-economic impact of projects financed by the Cohesion Fund – a modelling approach*, vol. 2, Office for official publications of the European communities, Luxembourg, Luxembourg.

Vives, X. (2001). "Globalización y localización", in García-Milà, T., *Nuevas fronteras de la política economica*, Centre de Recerca en Economia Internacional, Universitat Pompeu Fabra, Barcelona, Spain.

World Bank (1994). *World Development Report*, The World Bank, Washington, D.C., USA.

World Bank (2003). *Inequality in Latin America and the Caribbean*. World Bank Latin American and Caribbean Studies. The World Bank, Washington, D.C., USA.