Infrastructure and Productivity: What are the underlying mechanisms?
William P. Anderson and T.R. Lakshmanan
Center for Transportation Studies, Boston University


In recent years there has evolved a rich literature on measuring the effect of public infrastructure investments on the productivity in the private economy\(^1\). While this literature is beset by theoretical and methodological controversies, some studies provide evidence of a significant productivity effect. Most studies, however, are of a macroscopic nature and therefore do relatively little to identify the underlying mechanisms. The goal of this paper is to suggest the most likely mechanisms, to review the empirical evidence to date and to identify the most fruitful avenues for further research.

The approach of most empirical studies is to introduce some measure of public capital into either an aggregate production function or an aggregate restricted cost function.\(^2\) In either case, the role of public capital is subject to interpretation. Take the production function. One can clearly envision why private capital and labor have positive effects on output, but the reasons for the presence or absence of a positive effect for public capital is less clear. Public capital is an even more heterogeneous category than private capital; it is not used exclusively by any firm, but rather used collectively by diverse firms and households; and for any firm one might consider there are large portions of the public capital stock that they do not use at all, at least not directly.

This ambiguity is reduced somewhat by studies that isolate particular components of the stock of public capital. In particular, the role of transportation infrastructure permits a narrower range of interpretations. Even in this case, however, the empirical literature does not answer a number of questions that need to be addressed. How does the presence of more and better roads affect the relationships between spatially dispersed production centers within a national or continental economy? How do individual firms

---

\(^1\) A review and discussion of this literature is provided in a companion paper, Lakshmanan and Anderson 2004.

\(^2\) The two approaches imply different assumptions. The production function approach is consistent with perfectly inelastic factor supplies while the cost function approach is consistent with perfectly elastic supplies (Haughwout, 2002).
respond to better infrastructure by altering their logistics, production methods and locations? Does better infrastructure lead to more efficient spatial configurations of firms and households? Can transportation infrastructure ever have a perverse effect – effecting a reduction in productivity?

For the purpose of this paper we limit our discussion to factors related to transportation infrastructure. This is not meant to downplay the importance of other categories, such as water and sewage, public utilities or educational and health care facilities. Rather, we limit ourselves to transportation infrastructure in order to keep the discussion manageable and closer to our fields of expertise.

The remainder of the paper is organized as follows. The next section discusses ways in which transportation infrastructure affects productivity over a broad geographical scale. These include gains from interregional and international trade as well as coordination problems in economic development. The next section switches from the macro to the micro, considering how the supply of transportation infrastructure allows individual firms to function in a more productive manner. We focus here on logistical and locational effects. The following section attempts to draw out the linkages between transportation infrastructure and agglomeration economies. The New Economic Geography provides a theoretical lens for looking at agglomeration in a general equilibrium framework, but it is limited to addressing certain types of agglomeration mechanisms. Infrastructure may play different roles in different types of agglomeration. We also consider the possibility that investments in transportation infrastructure may reduce productivity by working against agglomeration. In the final section of the paper we propose some fruitful directions for continued research.

**Economy Wide Mechanisms**

These are mechanisms that increase productivity in many regions or countries simultaneously because of increased integration, expansion of markets and mutually reinforcing growth. We may separate these into conventional gains from trade, the “big push” to overcome coordination failures and technological shifts that are made possible by complementary transportation infrastructure.
Gains from Trade

Adam Smith observed in 1776 that the division of labor is limited by the extent of the market. In the absence of institutional barriers, transportation infrastructure extends the markets of individual producers. Thus, one of the most important ways that transportation infrastructure contributes to growth in productivity is by increasing specialization and trade. This is true for international trade, whether at the global scale or within trade blocs such as the European Union. It is equally true for interregional trade within economies with a high degree of regional differentiation, such as the United States. The US Interstate Highway System, the Trans European Network program and the emergence of super efficient ocean ports all contribute to “Smithian” growth – growth arising due to specialization and trade.

The productivity benefits of specialization and trade can arise via two mechanisms:

- Each region has a different endowment of natural resources, labor (availability and skills), capital goods, and institutions that make it highly efficient in some categories of production and less efficient in others. Specialization and trade allows regional resources to be concentrated in those forms of production for which they are best suited. This is basically David Ricardo’s theory of comparative advantage, whose theoretical extensions are embodied in the Heckscher – Ohlin – Samuelson framework (Findlay, 1995.)

- Even if regions have similar endowments of resources, there is still a benefit to interregional trade via scale economies that are realized as producers target broader markets. Furthermore, production and trade across a national (as opposed to regional) market makes it possible to provide consumers with a broader variety of goods. This is the explanation of gains from trade provided in the new economic geography (Fujita, Krugman, and Venables, 1999; Venables and Gusiorek, 1999.)

Transportation infrastructure is important here because specialization and trade will only occur if the efficiency gains from trade are greater than the transportation costs required. For example, Limao and Venables (2001) found that inadequate transportation infrastructure is a major impediment to the export performance of developing countries,
especially in Africa. This type of productivity effect is difficult to capture in cost-benefit analyses, however, and may therefore be missed in policy analyses.

The contribution of transportation infrastructure to productivity need not be a slow process. Kelly (1997) has shown that transportation networks may have critical density levels below which markets are largely isolated and above which they become integrated, expanding the market for the most efficient producers in all regions. Kelly relates this theoretical argument to the rapid commercialization of the Chinese economy that occurred after completion of interregional waterways in the 12th century.

Coordination Failure and the Big Push

Rosenstein-Rodan (1943) and Hirschman (1958) described a problem of development in traditional economies whereby investment in increasing returns to scale technologies (i.e. industrialization) in a single production sector is not profitable, but simultaneous investment in such technologies by several sectors is profitable. This can occur for various reasons. Most notably, when there are linkages between sectors by means of intermediate goods, expanding one sector expands the market of other sectors. Thus if all sectors industrialize at once there are mutually supportive intermediate demands that allow them all to achieve scale economies. This is the justification for a “big push” industrialization policy as a means to address the coordination failure that prevents such simultaneous industrialization from occurring without intervention.

Another reason for coordination failure is the fact that no single sector can support the transportation or other infrastructure necessary for its industrialization. If all sectors industrialize at once, they can jointly support this infrastructure (Murphy et al, 1989). In light of this, investment in public infrastructure may be viewed as a policy to overcome coordination failure for two reasons. First, such investment may be sufficient to make it profitable for all sectors to industrialize independently. Second, even if it is not sufficient to create independent profitability, it may provide a signal to firms in all sectors to anticipate widespread industrialization.
Technological Shifts

There have been times in history when expanded freight services have made radical changes in the structure of production possible. For example, the development first of canals and later of railroads made it possible for huge areas of the central lowlands of the US to be developed for specialized agriculture serving a national market. To a degree this fits into the standard comparative advantage argument described above, except that rather than a shift from autarky to specialization it involved the creation of new economic regions whose growth was driven from an early stage primarily by export commodities. Furthermore, it involved a fundamental transformation of production technologies, achieving much higher productivities through specialization and large-scale production. It can be argued that a host of improvements in agricultural technology were induced, at least in part, by the expanded market opportunities made possible by freight improvements.

Another example is the industrial revolution in textile production that occurred on a global scale in the 19th century. In this case, improved freight made it possible to develop a production system that required the movement of materials inputs of cotton from production region (US South, Egypt, India) to production centers in England and New England. Thus, unlike the example above where freight made it possible for a specialized production region to reach broader markets, in this case freight made it possible for widely separated but complementary regions to be integrated into a specialized production system. Again, this story has elements of comparative advantage, but it involves a fundamental shift in technology made possible by improved freight.

Microeconomic Effects

These are mechanisms that act largely through the business decisions of individual firms. The logistics revolution of the past two decades has received considerable attention as a productivity benefit of infrastructure investment. But such investments may also affect location decisions in ways that may enhance productivity. While most research on the infrastructure productivity link focuses on freight, personal transportation is also important because transportation costs affect the frequency of face-to-face communication.
Logistics

The savings in transportation cost due to improved infrastructure can in some cases be transferred into savings in non-transportation cost elements such as storage, interest and insurance costs. Suppose a manufacturing firm has to receive a fixed quantity \( m \) of some material input in each one-year production cycle. One of the decisions the firm must make is whether to receive the input in one large order or in a number of smaller shipments. Define total logistics cost \( TLC \) as the sum of procurement, carrying, and transportation costs associated with that input. Procurement costs \( P \) will be lowest if one large shipment is received because it will only have to be ordered and processed once. Transportation costs \( T \) will also be lower because it is generally cheaper on a per unit basis to ship large batches of goods. However carrying costs \( C \), which include interest, insurance and storage costs, will be lower if the input is received in small shipments so that the amount held in inventory is minimized. (This is one of the principal benefits of Just In Time systems.)

Thus the optimal input shipment size – and the associated optimal level of inventory – depends on the trade-off between procurement and transportation costs on the one hand and carrying costs on the other. A similar argument can be made with respect to the firm's outputs. By delivering goods more frequently and in smaller batches the firm saves on inventory carrying costs by purchasing more (or better) transportation services.

Figure 1. Total Logistics Costs

![Figure 1. Total Logistics Costs](image-url)
This is illustrated in Figure 1 where the sum of transportation and procurement costs $P+T$ and carrying costs $C$ are graphed against the average size of shipments $B$. The optimal $B$ is found where $TLC = T+P+C$ is at a minimum. When a reduction in transportation cost from $T$ to $T'$ occurs – shown here as a downward shift in the $P+T$ schedule – the effect is a decrease in the optimal $B$. Thus, cheaper freight services leads to reduced storage, insurance and interest costs associated with carrying inventory. (For a theoretical treatment of the effect of transportation cost on optimal shipment size see McCann, 1998.) Of course, maintaining lower inventory increases the risk that the production process will be interrupted because a shipment of some critical input is late. Thus, if freight services are unreliable, greater than optimal inventory must be maintained as a buffer against risk. Improvements in freight services in recent years – brought about in part by application of information and communications technology – have improved the reliability of delivery within narrow time windows, thus allowing producers to reduce inventory and thereby reduce carrying costs. This illustrates that improvement in the quality, as well as reductions of the cost, of freight services yield benefits.

Until recently, empirical evidence of the magnitude of the logistics effect was limited to case studies and small sample surveys. For example, Table 1 summarizes results from a study by Hickling et al (1995) relating reductions in freight travel times to changes in total logistics costs. Only a portion of savings are derived from reduced transportation costs, with reductions in inventory carrying costs comprising another significant component of savings. This in part explains why medical and surgical instrument producers, which carry very high value inventories, can reap greater benefits from transportation improvements than the other industries.
Table 1. Measures of Travel Time Reduction Impacts on Costs (Hickling, Lewis, Brod, 1995)

<table>
<thead>
<tr>
<th>Industry</th>
<th>Logistics cost / travel time elasticity</th>
<th>Logistics cost savings as % of sales 20% reduction in travel time</th>
<th>Logistics cost savings as % of sales 45% reduction in travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail Food</td>
<td>.055</td>
<td>.04%</td>
<td>N/A</td>
</tr>
<tr>
<td>Automotive Parts</td>
<td>.234</td>
<td>.20%</td>
<td>.45%</td>
</tr>
<tr>
<td>Telecommunication Equipment</td>
<td>.103</td>
<td>.02%</td>
<td>.05%</td>
</tr>
<tr>
<td>Medical / Surgical Instruments</td>
<td>.548</td>
<td>.88%</td>
<td>1.98%</td>
</tr>
</tbody>
</table>

A very recent study by Shirley and Winston (2004) attempts to draw an empirical link between the provision of road infrastructure and inventory levels over a much more comprehensive data set than earlier studies. Theory suggests that lower transportation costs and higher reliability allow firms to maintain lower inventories. If costs and reliability are related to the capacity and condition of highway infrastructure, then inventories should be inversely related to highway infrastructure expenditures. Estimating an econometric specification derived from the theory of logistics costs on firm level data from the Census Bureau's Longitudinal Research Database, the authors find empirical support for this hypothesis. Their results indicate, however, that the marginal inventory reductions for each dollar of infrastructure expenditure are declining over time.

Location Choice

New infrastructure often leads to locational shifts. For example, a circumferential highway may encourage firms to locate further from the urban core. In many cases locational shifts don’t confer any benefits because they simply represent relocation of activities within the economy. If, however, the location shift leads to improved productivity, it constitutes an economic benefit.

One such case is where reduced freight costs allow a multifacility firm to concentrate its production on a smaller number of locations in order to take advantage of scale economies. One of the most persistent themes in the theory of industrial location is
the trade off between scale economies and transportation costs. Imagine a manufacturing firm that wants to sell its product in a large number of urban markets dispersed throughout a regional or national economy. If its production technology has scale economies, it will minimize production cost by operating in a single location to serve all markets. However it will minimize delivery costs by producing in smaller facilities located near important markets. It must therefore choose some optimal configuration of facilities and locations based on the counteracting effects of scale economies and transportation costs. If freight costs decline – whether from improved infrastructure, technical progress, or enhanced productivity in service provision – the firm will have an incentive to close some facilities and expand others, or build one or more large-scale facilities in new locations.\(^3\) In this case there is a transfer of jobs and income from locales where closures occur to locales where expansions occur, but there is also an aggregate efficiency benefit due to scale economies.

A useful way to look at the consolidation effect is as a substitution between production inputs and transportation services (HLB, 2001). Achieving scale economies means that on a per unit of output basis, the firm will be able to purchase fewer inputs of labor, capital, and other inputs. Since the scale economies are achieved by consolidating facilities, however, they come at the expense of more ton-miles of transportation services.

\(^3\) Even in the absence of scale economies, there may be benefits from spatial concentration of production. For example, if there is uncertainty in the level of demand in individual markets, concentration of production may have the effect of pooling risk – aggregate demand is more predictable than local demand (Comacho and Persky, 1990.)
This substitution is depicted in Figure 2 by the isoquant II, which defines different combinations of production inputs and ton miles that can be combined to produce a given amount of output. If transportation costs are reduced, the slope of the iso-outlay line AB is reduced to the line CD. The points of tangency define the optimal combination before (T) and after (T’) a reduction in transportation costs.

**Face-to-face communication**

In information intensive industries, personal transportation may be more important than freight transportation. The benefit of face-to-face communication is generally increasing with the complexity of the information being exchanged. Industries that are research intensive or involved especially complex supply chains or financial instruments therefore require more face-to-face than mature industries. A high level of face-to-face interaction is consistent with high levels of technological advance and creativity, which in turn promote productivity growth.

Kobayashi *et al* (1998) argue that face-to-face meetings are underconsumed in the economy, leading to significant welfare losses. The reason is that for any meeting to occur, the expected benefit for each participant must be above some threshold level. This threshold is a function of the value of time, which is constantly increasing. In cases where one person would choose to have a meeting while another would not, there is a welfare
loss to the former person. Kobayashi et al have shown that this welfare loss is increasing in personal transportation costs.

In order to consume sufficient face-to-face meetings, firms that engage in the transfer of complex information may choose to cluster in large cities or cities that specialize in some sort of information intensive activities. Alternatively, they may locate so as to have the opportunity to engage in meetings with people in other cities at a sufficiently low cost in terms of time and money. High speed rail in Japan and Europe and air travel in the US contribute to productivity by allowing face-to-face meetings to occur over wider spatial fields.

**Agglomeration Economies**

To traditional economic geographers, agglomeration economies are the most important spatial process promoting productivity growth. Early studies showed a link between city size and productivity (Mera, 1975). More recently, Ciccone and Hall (1996) provide further evidence that concentration of economic activity promotes higher productivity. Thus, a reasonable proposition is that infrastructure promotes productivity to the extent that it supports agglomeration. This raises two questions. First, what are the mechanisms by which infrastructure supports agglomeration? Second, does transportation infrastructure always support agglomeration or can it sometimes support dispersion?

*3 Types of Agglomeration*

McCann and Shefer (2004) contend that agglomeration economies can be defined as falling into three categories:

1. *pure agglomeration*: Firms benefit directly from the existence of an agglomeration – their relationships with other firms are essentially transient and the only price they pay for membership is the higher land rent.

2. *industrial complex*: Firms are interlinked by exchanges of goods and services and therefore have longer term interests in each others’ locations.
Therefore there are high entry and exit costs. Such complexes are local, but not necessarily urban.

3. **social networks:** Firms are interrelated by means of a culture of mutual trust. The absence of opportunism reduces transaction costs and promotes risk-taking and joint ventures. Physical agglomeration is not absolutely necessary but spatial proximity helps to promote such an environment.

Transportation infrastructure plays different roles in different types of agglomeration. In the case of pure agglomeration, transportation infrastructure is important to allow intrametropolitan commuting and allow diverse markets to be served from a central location. For the industrial complex, intraregional freight movement is critical and infrastructure requirements may be more specialized, depending upon the industrial sectors involved. McCann and Shefer (2004) argue that the social network model depends on social rather than physical infrastructure, although infrastructure for personal transportation clearly plays a role here.

*The New Economic Geography*

The New Economic Geography refers to a suite of models that adopt the monopolistic competition framework of Dixit and Stiglitz (1977) to cast spatial economic problems in a general equilibrium setting (Fujita *et al.*, 1999). These models are able to treat agglomeration explicitly, but are limited to the pure agglomeration version.

The basic two region model defines core and periphery regions, where the latter produces undifferentiated, constant returns to scale products. Differentiated products with scale economies trading in monopolistically competitive markets can be produced in either of the two regions, but the core region has the benefit of a larger market for these goods. In general, investments that reduce transportation costs tend to promote concentration of differentiated production in the core – in other words agglomeration. Thus such investments enhance productivity (via scale economies) but also increase regional disparity. This may be good or bad, depending on policy priorities.

The relationship between transportation costs and agglomeration in this general framework are subject to some ambiguity, however. Venables and Gasiorek (1999)
developed a computable general equilibrium model of the NEG type and conducted simulations to address questions related to the effect of transportation infrastructure. A simulation was devised to illustrate the case where one region has higher wages before the reduction in transportation cost occurs. The results indicate that if transportation costs are so high as to result in virtual autarky in the initial case, marginal reductions in transportation cost will cause wages to diverge. At that point, firms are attracted to the more lucrative market, but eventually wage increases in the center region cause shifts in the other direction. Thus past some critical level of transport costs, reductions will result in dispersion of production and wage convergence. For this reason, an inverted U-shaped relationship between transportation cost and regional income equality is indicated. (Kilkenny, 1998, found a very similar relationship.) So the effects of a transport investment on both agglomeration and interregional equity depend on the initial context and may be very difficult to predict.

*Infrastructure, Agglomeration and Dispersion*

A pervasive problem in discussions of infrastructure and productivity is that we tend to treat infrastructure as a homogeneous stock without paying enough attention to its spatial structure and how new additions affect existing networks. While there are theoretical arguments for why infrastructure promotes agglomeration it is also perfectly plausible that specific additions to the infrastructure may promote dispersion.

Haughwout (2000, 2001) argues that agglomeration is the outcome of positive externalities and is therefore underprovided in the economy. A firm that must decide whether to move from a central to a peripheral location in order to economize on labor or land costs does not take into account the negative impact that its departure may have on other firms. In principal, infrastructure can be a powerful tool to promote agglomeration. In practice, however, (at least in the US) new infrastructure is equally likely to promote dispersion. Highway infrastructure investment dollars in the US are distributed via a system whereby a fixed pool of funds from a federal gas tax is distributed to the states for proposed highway improvement projects. This process is naturally a political one and investments are likely to go to the areas and special interests that are most influential. For this reason, infrastructure investments in the suburbs take precedence over investments in
central cities. Furthermore, among the main beneficiaries of new infrastructure are holders of undeveloped land whose value will be enhanced by increased accessibility. Since most such land is in the periphery, there is political pressure for peripheral development.

This argument rests on the supposition that agglomeration is beneficial for the productivity of all firms. It may be the case, however, that different firms benefit from different locations. The product life-cycle model (Vernon, 1966) argues that products and the industries that produce them pass through stages in their histories over which their main business requirements evolve. At an early stage of development, acquiring appropriate labor skills, developing product innovations, and market penetration are the key concerns, while at a later stage, implementing process innovations, and economizing on the cost of inputs such as labor and land are the main competitive strategies. The spatial analogue to this argument is that early stage firms do best in urban core locations while late stage firms do best in the periphery. An alternative argument is therefore that a transportation infrastructure system that provides sufficient capacity and connectivity benefits firms by expanding the range of locations from which they can choose (Lakshmanan and Anderson, 2002). The net impact therefore rests on whether the benefits of choice outweigh any losses in positive externalities.

**Conclusions and Research Directions**

The foregoing discussion identifies a rather long list of distinct mechanisms by which investments in transportation infrastructure may increase productivity in the private economy:

1. increased specialization and trade
2. the “big push”
3. change of production technology
4. reductions in total logistics costs
5. facilities consolidation
6. enhanced face-to-face interaction
7. pure agglomeration effects
8. industrial complex development
9. development of social networks

One could probably identify others. We have also identified at least one way in which investments in transportation infrastructure may detract from productivity: by promoting spatial dispersion. Naturally, it would be interesting to know what the relative contributions of these mechanisms are. Macroscopic empirical studies provide rather little guidance in this regard.

In order to better understand these mechanisms, empirical studies are needed that are based on firm or establishment level data and more targeted to specific hypotheses. Shirley and Winston’s (2004) study on infrastructure and inventory levels is a good example: it is based on a panel of establishment level observations and it is focused on a particular type of benefit from infrastructure.

Another useful avenue of research comprises *ex post* assessment of major infrastructure projects and programs. By means of a more historical perspective we can ask a number of critical questions such as: To what extent did improved infrastructure lead to increased specialization and expansion of markets? How do freight service firms take advantage of improved infrastructure to offer cheaper and better services? What logistical, technological and locational innovations followed in the years after the project’s completion? How did such economic adjustments translate into higher productivities and incomes? Naturally, such assessments must be more than just case studies; they must apply appropriate analytical methods to identify those economic changes that can be attributed to the infrastructure improvement from those that would have occurred without it.
References


McCann, Philip (1998), *The Economics of Industrial Location: A Logistics-Costs Approach*, Heidelberg: Springer.


on Trunk Road Assessment, London: Department of Environment, Transport and the Regions.