Transportation Infrastructure, Productivity, and Externalities

Charles R. Hulten
University of Maryland
and National Bureau of Economic Research

August, 2004
Revised February, 2005

ABSTRACT

This paper summarizes the results of three studies linking investment in highway infrastructure to productivity growth in the manufacturing sector of the U.S., Spanish, and Indian economies. The goal of this research is (1) to trace the overall impact of highway investment on the growth of this strategic sector, (2) to examine the interregional effects of such investments, with particular attention to the issue of whether highway investment encourages regional convergence and relocation of economic activity, and (3) to assess the extent of the spillover externalities on manufacturing industry associated with such investments. This last issue is of particular importance for infrastructure policy, since spillover externalities tend to go uncounted in formal project investment analyses, leading to the possibility of under-investment. The comparative study of three countries at different stages of economic development using virtually the same model allows a fourth issue to be examined: the possibility that the effects of infrastructure investment differ according to the level of development and the extent to which existing infrastructure networks have already been built up. These issues are first framed in the larger context of the literature on infrastructure and productivity.

Paper prepared for the 132nd Round Table of the European Conference of Ministers of Transport, at the Joint OECD/EMCT Transport Research Center, Paris, France, December 2 and 3, 2004.

1. Transportation Infrastructure and Productivity: Historical Background

The idea that transportation infrastructure is a type of capital investment distinct from other forms of capital is an accepted part of the fields of economic development, location theory, urban and regional economics, and, of course, transport economics. In his classic treatise, Albert O. Hirschman (1958) classifies transport infrastructure systems as "social overhead capital (SOC)" to distinguish it from the type of capital that is used directly by industry to produce their goods and services (e.g., plant and equipment), which he calls "directly productive assets (DPA)." [1] Hirschman points to four characteristics that distinguish SOC from DPA: (1) SOC is basic to (and facilitates) a great variety of economic activities, (2) it is typically provided by the public sector or by regulated private agencies, (3) it cannot be imported, and (4) it is "lumpy" in the sense of technical indivisibilities. He also argues that the function of SOC investment is to "ignite" DPA, and that "Investment in SOC is advocated not because of its direct effect on final output, but because it permits, and in fact, invites, DPA to come in (page 84)."

A more modern treatment of these issues would frame in terms of the economic theory of partial public goods, or "clubs," modified to include network theory. Roads and highways, for example, are lumpy joint use networks with many different simultaneous users and uses. Unlike "private good" DPA investments, the conditions for optimal provision involve the summation of benefits across the different users ("members of the club"), adjusted for congestion effects. Moreover, the benefits associated with any one segment of the network (or 'mini-club') depend on the

---

[1] "SOC is usually defined as comprising those basic services without which primary, secondary, and tertiary productive activities cannot function. In its wider sense, it includes all public services from law and order through education, public health, to transportation, communications, power and water supply, as well as such agricultural overhead capital as irrigation and drainage systems. The hard core of the concept can probably be restricted to transportation and power. (Hirschman (1958), page 83)."

size and configuration of the entire network, and not just with that segment. Spillover externalities between network segments are therefore potentially important. So are Hirschman's "igniting" effects, since the addition or expansion of key networks effects can have a magnified effect throughout the network. he example in the U.S. of the intercontinental railroads which opened up the western regions of the country in the middle of the 19th century is a case in point.

Sorting out the complex structure of benefits involved in network "club" is notoriously difficult. A partial solution to this problem is to limit attention to the production side of the economy, where the more-or-less immediate impact of SOC infrastructure on the growth of output is more easily measured. This is the was used by David Aschauer (1989a,1989b), who estimated aggregate production functions for the U.S. that included a public capital variable. Aschauer's work received a great deal of attention because of the startling result that the supply-side effects of public capital alone translate into a gross return of 100 percent per annum or more, and a payback period of one year or less (a point noted in Gramlich (1994)). These seemingly implausible results were soon supported by other supply-side studies that also reported a large public capital effect. One implication of this literature was that the U.S. was drastic under-investing in its public capital and infrastructure systems, and that one culprit was the failure of conventional micro-economic project evaluation procedures to count the large spillover externalities associated with these systems, though no explicit estimates of the magnitude of the presumptive externalities were offered. However, the "igniting effects" associated with the building of the U.S. intercontinental railroads was often mentioned as an example of how this might occur.

It is not surprising, in view of the magnitude of the infrastructure effect, that these results triggered a large amount of subsequent research. Gramlich notes "at least forty other econometric studies using different data and techniques" appeared in the five years between the first paper published by Aschauer and the publication his survey article. He also observes that, over this period, "the bubble has happened and may even be beginning to burst (page 1177)," since the emerging research suggested that the econometrics of the macro-production function approach were fragile and not robust to changes in estimation technique, scope, and data. For example, estimation using "first differences" in the data often produced far lower, and sometimes statistically insignificant results, compared to "level" estimates; much the same was true when "panel data" with a regional dimension were employed. Moreover, the direct estimate of the link between infrastructure and output was subject to the problem of "reverse causality." While it is likely that transport and other infrastructure investment make possible ("cause") a larger volume of national output, it is equally clear that a growing volume of output leads to more investment in such systems. To assume that the statistical correlation between infrastructure and output is caused entirely by infrastructure is not only to commit the econometric sin of simultaneous equations bias, but also to invite an overstatement of the return to public infrastructure.

More recent econometric studies that have employed more flexible functional forms have obtained rates of return that are within a more normal range of experience. Nadiri and Mamuneas (1994), for example, use a flexible cost function approach to estimate an average annual "social" rate of return to infrastructure of around 7.2% in U.S. manufacturing industry for the period 1955-86, compared to a 8.7% annual rate of return to private capital. In similar work, they report plausible rates of return to highway infrastructure investment that diminish as the Interstate Highway System program matured. However, the problem of model specification continues to present a significant challenge to the application of production function techniques. One challenge is to specify models in such a way that the externality effects are isolated. An even more formidable challenge is to allow for the fact that most forms of core infrastructure, including transport systems, come in the form of spatially distributed networks of individual investments, and that the productive capacity of a network depend on its configuration as well as its total size.[2] Unfortunately, statistics on the network capital

---

[2] Unlike most DPA capital, the marginal product of network capital depends on where in the system the incremental investment is made, and not just on the how much capital is already in place (Hulten (1994)). Thus, the practice of adding an infrastructure stock variable to a production function when it is measured in the same way as private DPA capital results in model mispecification. The practical consequence of this mispecification is that the parameter (elasticity) associated with the network capital will vary according to the evolution of the network, and, as noted later in this paper, may be very large at some times and zero at others.

essentially measure only the latter.

These considerations provide an intellectual backdrop for the body of empirical research described in the following sections. This research is based on a model developed by Hulten and Schwab (1984,1991,2000) that attempts to avoid some of the econometric problems described above and, in particular, to isolate the externality effects. This approach focuses on manufacturing industry rather than on the entire economy, and has both a regional and a time dimension. Both features help reduce the problem of reverse causality, for reasons noted below. However, the principal difference between the Hulten-Schwab approach and the majority of the production function literature on infrastructure is the focus on total factor productivity rather than on real output as the left-hand variable of interest. This shift in focus reduces the dependence on econometric specification and the associated problems, further reduces the problem of reverse causality, and most importantly, provides a means for isolating infrastructure externalities as they affect manufacturing industry. Since this last issue is of considerable importance to infrastructure policy, it is developed in some detail in the following section, before turning to the description of the data and results.

2. Infrastructure “Channels” in the Theory of Production

The Hulten-Schwab approach to isolating externalities follows the general approach of Meade (1952) in assuming that the effects of infrastructure capital operate through two different channels in manufacturing industries.[3] In the first channel, the benefits to manufacturing industries from infrastructure investments are received *indirectly* in the form of inputs purchased from those sectors

---

[3] Meade distinguishes between unpaid factors of production and pure spillover externalities in the context of a production function, but does not deal with infrastructure capital per se. However, his model has a natural application to the production-side aspects of the infrastructure problem, since infrastructure can be expected to exhibit both of Meade’s effects. The idea that infrastructure operates through different channels is certainly not uncommon in the transportation literature, with examples of different channel taxonomies appearing in recent papers presented at these Round Tables by Berechman (2001) and Prud’homme (2001). The contribution of the current model is to show how to use these channels to estimate externalities in infrastructure-using industries.

involved in the production of infrastructure services (for manufacturing, this is mainly transportation, and various utilities). Roads and highways, for example, are combined with vehicles, workers, fuel, warehouses, etc., by the transport industry to produce transportation services sold to other sectors. Similarly, electric utilities combine the services of infrastructure networks with inputs of DPA capital, labor, and fuels and sell the output directly to other industries. From the perspective of the manufacturing sector these infrastructure services appear as an intermediate input purchased from the upstream producing industry. In this process, the unpaid infrastructure inputs are converted to a *paid* factor of production in the downstream industry, and any improvement in the quantity or quality of the infrastructure network upstream appears as a *reduction* in the cost of the intermediate purchases of transportation services and electricity downstream, or as an improvement in the quality or scope of these services. A similar process occurs with the primary factor inputs, labor and capital. Improved transportation may, for example, lower the labor costs by expanding the pool of available workers or by reducing the cost of housing workers near the work place. In any event, upstream infrastructure externalities are internalized in the market for purchased services and factor inputs by the time they arrive at the downstream user.

If this were the only channel through which infrastructure affects manufacturing output, there would be little or no role for externalities in that industry. However, infrastructure may also affect manufacturing industries *indirectly* through a second channel: network externalities. The expansion of capacity at one point in an existing infrastructure system can have effects throughout the network through the addition or extension of critical links, or the elimination of bottlenecks. These indirect can lead to an overall increase in productive efficiency as, for example, when lower transport costs lead to an expansion in the size of product and input markets, in turn leading to efficiency gains through economics of scale and scope, increased competition, and to greater input specialization. These system-effects may also permit the use of newer more efficient technologies (e.g., just-in-time inventory management) or those allow more efficient use of existing technology (e.g., fewer vehicles and drivers per unit output as congestion is reduced). These second channel effects are external to the firms located at any point on the network, and, unlike the first channel effects, they operate largely outside the market place and are not mediated by prices.

The two channels can be given a more precise analytical form in terms of the production functions underlying the descriptive analysis. Imagine an economy with only two products, transportation services, T, and a manufactured good, Q. The production of transport services is $T = T(D_T,B)$, where D_T is DPA input used by the sector and B is the transportation network; the technology for the manufactured good is $Q = F(D_Q,T)$, with D_Q as that sector's DPA input and T the transport service purchased from the other sector. The technology for manufacturing does not make use of the transport network as a direct input. The entire impact of transport infrastructure, of manufactured goods, in this case, operates through the markets for the primary good and the intermediate transport good purchased by that sector, and the change in manufacturing output associated with an increase in infrastructure can be expressed as $\Delta Q/\Delta B = [(\partial Q/\partial D_Q)(\partial D_Q/\partial T) + (\partial Q/\partial T)]\Delta T/\Delta B)$. This is the first channel effect defined above, and there is no second channel effect, $(\partial Q/\partial B)$, in this case.

Second channel effects in the production of manufactured goods can be modeled by introducing a "shift" term, A(B), into the manufacturing production function, which can now be expressed as $Q = A(B)F(D_Q,T)$. The second channel effect operates in this expanded model through the term $\partial Q/\partial B)$ that is, through the increase in output due to infrastructure holding purchased inputs constant. The change in infrastructure now affects output indirectly by allowing the purchased inputs, D_Q and T, to be used more efficiently. Moreover, transport infrastructure now operates influences output through both of the channels described above, and the total effect can be written as $\Delta Q/\Delta B = [(\partial Q/\partial T)(\Delta T/\Delta B)] + (\partial Q/\partial B)$. The term in square brackets is the first channel effect, and the last term on the right-hand side of the expression is the second channel operating as an externality outside of the market place.

The implications of this framework depend on whether the focus of the analysis is on individual nodes within the network or on the network. First, standard location theory and economic geography (e.g., Krugman (1998)) predicts that a positive output effect across all nodes, $\Delta(\sum_i Q_i)/\Delta(\sum_i B_i) > 0$, does not necessarily is imply a positive effect at each node, $\Delta Q_i/\Delta B_i$. A reduction in transportation costs due to an expansion in the road network exposes inefficient (or otherwise high-cost) producers to competition from producers at lower-cost nodes of the network, and can result in a relocation and concentration of production to those nodes. The relocation effect may be sufficiently strong that $\Delta Q_i/\Delta B_i < 0$ at some nodes. Such effects can have redistributional implications that are at least as important for assessing the net benefits of an infrastructure building program as the overall expansionary effects.

The level of aggregation is also important because the individual nodal effects, $\Delta Q_i/\Delta B_i$, will usually depend on more than just the infrastructure at the node in question. For example, the impact of expanding a road linking nodes A and B will depend, in part, on the adequacy of the road link between B and C. The derivative $\Delta Q_i/\Delta B_i$ must therefore be regarded as a function of all the links in the network $\{B_{j,k}\}$. The micro-project analysis of a specific investment project in the vicinity of node i will tend to miss the more distant effects of the investment on production at other nodes, and thereby tend to understate the true benefits associated with the investment (and possibly miss the "igniting" effect noted by Hirschman).[4] These uncounted effects are precisely what Aschauer and others counted on to justify the large magnitude of their macro-economic estimates. However, macro analysis solves one problem at the expense of another: while the macro approach incorporates the output effects across all nodes of the network by focusing on total output, $\sum_i Q_i$, it does so by aggregating the change in infrastructure across nodes into a single number $\sum_i B_i$. Moreover, ratio of

---

[4] The situation when the full range of marginal benefits, $\Delta Q_i/\Delta B_{j,k}$, is taken into account can be very complex. Incremental investments may have a very large payoff during a period in which a network is undergoing an "upsizing" at all nodes. On the other hand, cost considerations often dictate building capacity in advance of need during periods of upsizing, and investment in additional capacity has no immediate effect on output and therefore appears to have a zero marginal product. A similar outcome can arise in mature, built-up, networks before capacity constraints (e.g., congestion, bottlenecks) set in: incremental investments in this situation tend to be made in links that substitute for other links, rather than complement them as in the early stages of building, and will also tend to have a low marginal product.

these aggregate variable (or, more accurately, the corresponding elasticity $\gamma = [\Delta(\sum_i Q_i)/\Delta(\sum_i B_i)][\sum_i B_i/\sum_i Q_i]$) is treated as a constant parameter over time in the Aschauer-style empirical analysis, even though the parameter (elasticity) associated with the network capital will vary according to the evolution of the network and may be very large at some times and zero at others. This problem is a characteristic problem of macro analysis, which necessarily abstracts from the underlying heterogeneity of the micro world, and will be of importance in interpreting the results presented in the following sections.

3. Empirical Results: India

The model described in the sections which follow is an elaboration of the basic theory set out above. The basic idea is to estimate the $F(D_Q,T)$ component of the manufacturing production function separately using non-econometric "index number" techniques, and thereby isolate the shift term A(B) function that embodies the second channel externalities. A Hicks' neutral production function with both primary and intermediate inputs is a common specification that allows for this dichotomy:

$$(1) \qquad Q_{i,t} = A_{i,t}(B) \quad F^{i}(K_{i,t}, L_{i,t}, M(B)_{i,t}) \ ,$$

where Q denotes gross output, M is intermediate inputs, L is labor input, K is privately owned (non-infrastructure) capital, and B is the infrastructure stock. The variables and subfunctions can have both a time "t" and region "i" dimension. The shift term $A_{i,t}(B)$is specified as

$$(2) \qquad A_{i,t}(B) = A_{i,0} e^{\lambda_i t} B_{i,t}^{\gamma_i}.$$

This specification treats productive efficiency as a multiplicative function of the initial level of efficiency in each region, $A_{i,0}$, the exogenously determined average annual rate of technical progress in each region, λ_i, and the second channel infrastructure externality effect whose elasticity, γ_i, is assumed to be constant over time but can vary among regions.

The multiplicative form of the specification (1) and (2) is a standard feature of the literature on the Solow residual, which allows the shift term to be estimated using index number techniques (as opposed to applying econometric techniques directly to the production function). Solow's "total factor productivity" is conventionally defined as the ratio of real output to total input, which, in terms of the production function above, is equivalent to $TP_{i,t} = Q_{i,t}/F(K_{i,t},L_{i,t},M_{i,t})$. Total factor productivity (which is really "total productivity" in this context because it is based on gross output and intermediate input)[5] is therefore directly related to the parameters of interest in $A_{i,t}(B)$. Expressing productivity in logarithmic form results in

$$\ln TP_{i,t} = \ln A_{i,0} + \lambda_i t + \gamma_i \ln B_{i,t}. \tag{3}$$

The utility of this specification is that it isolates the parameter of interest, γ_i, in a form that can be estimated using regression techniques given estimates of total productivity and of the stock of transport infrastructure, both of which can be measured. It is the basis for the empirical results shown below.

The total productivity variable required for (3) is estimated in two steps. The growth rate is first estimated using the Solow residual method, which involves subtracting the growth rates of the inputs (L, K, M), each weighted by its shares in total income, from the growth rate of real gross output (Q).[6] The resulting residual estimate of the *growth rate* of total productivity must then be converted

---

[5] Total factor productivity is expressed as a ratio of real value added to an index of the primary factors, labor and capital. It is used primarily for measuring productivity for the aggregate economy. Total productivity includes intermediate goods in both the numerator and denominator of the productivity ratio, and is used primarily at the industry level of detail. The survey of the productivity literature by Hulten (2001) provides a more complete description of this topic and, more generally, of the Solow productivity framework.

[6] This procedure assumes that input prices are proportional to marginal products, the output elasticities of K, L, and M, are equal to the corresponding cost shares, and that the residual measures the shift in the production function. This pricing assumption is the characteristic limitation of the

to levels. This yields an estimate of variable $TP_{i,t}$ in the estimation equation (3) for each year and each region, and permits an analysis of the regional evolution of productivity as a byproduct.

This approach is applied to the manufacturing sector of the Indian economy in Hulten, Bennathan, and Srinivasan (2003), using data from India's Annual Survey of Industries for the years 1972-1993. This source includes annual estimates (in current prices) of gross output, intermediate inputs, labor input, and the book value of capital stocks for manufacturing firms registered under the Factory Act, which are the larger enterprises in the manufacturing sector. The estimates of output and input were then converted to constant (real) prices using a new output price deflator developed in the study, since previous approaches were deemed inadequate. The resulting growth rate estimates for the manufacturing sector as a whole are shown in the first columns of Table 1, and the productivity level estimates in Table 2.

Table 1 indicates that real gross output grew at a sustained rate of over seven percent a year for the two decades of our sample. It also shows that the growth of inputs explained most of the growth in output, with productivity only a small contributor. Since this productivity estimate includes the externalities associated with transport infrastructure, γ_i, it would appear to leave little scope for this effect. However, while the 0.5 percent annual growth rate of productivity may seem small, it appears small only because it measures the impact of innovation and infrastructure investment on a very broad base of inputs. When converted to a value-added basis, as shown in the second column of Table 1, the size of the productivity residual is a more conventional and very respectable magnitude of two percent.[7]

Solow residual method (Hulten (2001)). Its advantage is that it avoids the need to specify and estimate the "input" segment of the production function, $F(K_{i,t},L_{i,t},M_{i,t})$, and thus avoids some of the econometric problems that have troubled the literature on productivity of transport infrastructure.

[7] The two concepts of productivity are algebraically related: the growth rate of total factor productivity is equal to the growth of total productivity divided by the sum of capital's and labor's share of income (in effect dividing the latter by around 0.20). A two percent growth rate for total factor productivity is quite respectable when compared to similar estimates by Young (1995) for some of the highly successful East Asian economies: the manufacturing sector of South Korea grew at an

The Table 1 estimates of the annual growth rate of productivity (in both its measures) for Registered Manufacturing across *all* the states of India combined, and it thus excludes the regional dimension. Since the ultimate goal is to isolate the infrastructure spillover component of manufacturing productivity, regional differences in productivity and infrastructure across geographical regions are a potentially important source of variation. A sources-of-growth table similar to Table 1 was therefore calculated for each of the 16 states in the regional sample, and the results are summarized in Table 2 for Indian states grouped into terciles according to the initial level of productivity. The bottom five states ranked according to this criterion experienced a more rapid rate of both gross output and productivity growth, with the result that those states that started with the lowest levels of productivity narrowed (but did not eliminate) the gap with the leaders by 1992. This pattern of regional convergence was achieved with strong output growth in all regions, without major net relocations of manufacturing activity between the upper and lower groups of regions. This results stand in stark contrast to the experience in the United States.

4. Cross-National Comparisons

The results for India are based on the model first developed and applied to U.S. manufacturing. The sources-of-growth results obtained by Hulten and Schwab (2000) for the U.S. are reported in Table 1, based on data from the Census of Manufactures and the Annual Surveys of Manufactures for the years 1970-1986. They show a different pattern of growth compared to India: the rate of growth of output is significantly lower in the U.S. and it is driven primarily by productivity growth, while labor input growth was slightly negative.

The regional distribution of U.S. manufacturing growth is summarized in Table 3. The results for the nine Census regions in the original study are aggregated to the level of “Sun Belt” and “Snow

---

average annual rate of 3.0 percent over the period 1966-1990, while Taiwanese manufacturing grew at

Belt" to highlight the key fact exposed by the analysis: the shift in the manufacturing base from the older regions of the Northeast and Midwest of the U.S. (the "Snow Belt") to the South and West of the country (the "Sun Belt") that occurred after World War II. This shift attracted much comment in the 1970s and 1980s, for reasons that are apparent in Table 3, where Sun Belt growth rates of output and input are seen to be much larger. Labor input, in particular, shows a negative growth in the Snow Belt and no net growth overall, suggesting the presence of a significant relocation of manufacturing activity.

Some of the factors involved in this relocation are shown in this table. The rate of return to private capital was higher in the Sunbelt throughout the period and wages were lower, making these regions attractive from the standpoint of business. This was also a period of macroeconomic stagflation, weakening the power of labor to resist the relocation of businesses to low wage – high return areas. Public capital, the measure of "infrastructure" used in this study, also grew significantly more rapidly in the Sunbelt. The public capital variable includes transport infrastructure as one of its most important components. In an earlier study, Hulten and Schwab (1991) estimated that the growth rate of highway expenditures was also much more rapid in the Sun Belt, and it is worth noting that the U.S. Interstate Highway System program, begun in the 1950s, was largely completed during this period and surely played a role in accommodating the regional shift in the manufacturing base (as predicted by location theory).

Significantly, productivity differences among the regions was *not* a factor determining regional growth differentials.[8] The level of manufacturing total productivity was basically the same across regions, at both the starting and end points of the period studied. Essentially, manufacturing firms functioned at the same average level of productivity efficiency throughout the U.S. This finding is consistent with a rapid rate of technological diffusion among the regions, but contrasts with the case

---

1.7 percent over this period.

[8] The 1991 Hulten-Schwab study also found that the growth rates of productivity in the Sun Belt and Snow Belt regions are essentially the same. That study focused only on rates of growth and used real value added as the measure of real product, and therefore did not implement the full

of India, where productivity differentials existed and were compressed during the sample period.

The index number methodology used for the U.S. and India was applied to Spain by Mas et. al. (1998). The Spanish study focuses on the aggregate economy rather than on manufacturing industry alone, so the second-channel externalities cannot be isolated as with the U.S. and India. However, the Spanish study does use productivity rather that output as the variable of interest, as do the U.S. and Indian studies, so some insights can be obtained by a comparison of all three. The pattern of growth exhibited by the Spanish economy in Table 1 is, in fact, roughly similar to that of U.S. manufacturing, in that productivity change is the most important source of output growth and the role of labor is insignificant. However, the level of productivity differed by region in Spain, and showed a tendency to converge over the sample period.

5. Estimation of the Infrastructure-Productivity Link: India

The results reported in Tables 1 and 2 provide the data with which to study the relation between the level of productivity and the stock of infrastructure, as set out in equation (3) above. Following Hall (1988), this equation was expanded to allow for the possibility of increasing returns to scale and to allow for the possibility of a bias in the Solow residual method due to non-competitive pricing (a problem thought to be particularly characteristic of markets in developing economies). The externality parameter, γ_i, was assumed to be equal across regions in the variant of the econometric analysis reported in Table 4, and was estimated using a fixed effects model.[9]

The results shown in the first column of table 4 indicate a substantial and statistically

---

methodology of the 2000 study (on which the estimates reported in this paper are based).

[9] The estimates of the transport infrastructure variable, B, were proxied by paved roads obtained from annual issues of the Ministry of Transport's "Basic Road Statistics of India." This measure consists of the lengths of the following categories of paved roads: national highways (arterial roads for interstate movement), state highways (arterial roads for inter-district movement, linking up with national highways and adjacent state highways), and district roads ("Other Public Works Roads"). Unfortunately, adequate data on road capacity (lanes) were not available, nor were data on

significant externality effect. The implication of this estimate of γ for the gross return (marginal product) of transport capital is shown in Table 5: this rate of return increases from 2% in 1974 to 5% in 1993. While this is not a large number when compared to the overall return to private capital, 29%, it is nevertheless impressive since it represents only the second-channel externality effect, over and above the direct return to transport infrastructure.

Table 6 provides another look at the importance of the infrastructure effect. This table allocates the overall total productivity residual (the 0.04 estimate in Table 1) into the four components shown in the rows of Table 6, with the result that the transport externality effect is found to account for almost a quarter of total productivity growth. When expressed on a value-added basis as in the second column of Table 1, the transport externality is found to account for 0.25 percentage points per year, which is a very large effect in growth accounting terms.

The other regression estimates shown in Table 4 are statistically significant and of a conventional magnitude. The estimate of the scale effect implies mildly increasing returns to scale (3.8 percent), and the markup parameter suggests an 8.2 percent markup of price over marginal cost.

These results are but a subset of the results presented in Hulten, Bennathan, and Srinivasan (2003), and an even smaller subset of the results underlying the complete paper. For example, the project also considered the role of electricity generating infrastructure, and found a large externality effect there, as well. The gross return to electricity was found to be 5 percent in 1993, and the combined highway-electricity effect was 9 percent. When translated into the decomposition framework of Table 6, the combined effect explains approximately half of the annual growth in total productivity. The magnitude of this effect may seem implausibly large (and it may well be), but it is fairly well established “on the ground” that inadequate road transport and electricity generating capacity have exacted a significant penalty on Indian economic growth.

Finally, this analysis has assumed that spillover effects occur within the boundaries of each state, with no allowance for spillovers to neighboring states. When adjacent highway and electricity

---

capacity utilization. State road lengths were normalized by state area.

networks are included in the estimation, the implied spillover elasticities are much larger than in previous cases -- a combined 12.7 percent -- although the levels of significance are marginal, due possibly to the presence of multicollinearity. These estimates should therefore be interpreted with care.

6. Estimation of the Infrastructure-Productivity Link: The U.S. and Spain

A similar but less elaborate analysis was carried out for the U.S. manufacturing sector in Hulten and Schwab (2000).. The key result appears in column (2) of Table 4, where the estimated elasticity associated with infrastructure, γ, is not statistically different from zero. This is hardly surprising in view of Table 3, where the interregional differentials in the level of total productivity are effectively zero. Since total productivity is the dependent variable in the regressions reported in Table 4, there is little opportunity for infrastructure to matter in the cross-sectional dimension of the study.

The infrastructure variable used in the U.S. study includes all public capital, and is therefore broader that the transport variable used in the Indian study. That said, the comparison of the two countries invites the surmise that the effect of infrastructure investment, and the extent of uncounted externalities, depends on the extent of pre-existing networks. Given the interconnected nature of networks discussed in the opening sections of this paper, it is plausible to expect that such investment has a different effect in built-up, infrastructure rich, environments than in situations where there are significant infrastructure deficits. A comparison of the India and U.S. studies lends support to this hypothesis. The perceived inadequacies and needs underlying India's major effort to increase its highway network with the National Highway Development Project (the "Golden Quadrilateral") adds verisimilitude.

The regression results for Spain do little to confirm or reject this last hypothesis, given the economy-wide focus of the analysis and the resulting non-comparability of the regression estimates.

However, they confirm the general importance of the infrastructure productivity link at the aggregate level of economic activity in an economy that is in the later stages of economic development.

7. Concluding Remarks

The three studies reviewed in this paper shed light on the issues raised at the outset of the paper and have implications for infrastructure policy. The evidence suggests that investment in infrastructure networks does have an effect on the pattern of economic growth, and that the impact may depend on the stage of economic development. This evidence is by no means conclusive, but it does provide some support for the theoretical hypothesis that the effects of transportation network investments are highly non-linear: in built-up networks like the U.S., the primary effect of lowering transport costs is to relocate economic activity to lower-cost regions without a significant change in productivity or perhaps even system-wide output, whereas the addition of capacity to under-developed or capacity-constrained networks will tend to cause an improvement in productivity efficiency and lead to expansion in net output. The available evidence suggests that uncounted "second channel" externalities are important in the latter case and may be associated with systemic network under-investment. However, the three studies suggest that infrastructure investment is associated with convergence in regional growth in both built-up and infrastructure-poor networks, though to claim that the infrastructure causes convergence would be to over-interpret the evidence.

These conclusions are relevant for European transport policy, as the E.U. expands to incorporate lower wage regions in Central and Eastern Europe. Both the theory and evidence reviewed in this paper point to two important effects associated with improvements in the transportation systems connecting the new member states with existing E.U. members, and with improvements within the new members. On the one hand, a certain relocation of the existing manufacturing base toward the lower-wage regions can be expected. This is already occurring, according to Walter (2004), who writes:

> “Still, the prospect of a larger Europe, encompassing tens of millions of new, low-income workers, has many Western Europeans afraid that workers will migrate west in search of economic opportunities. In reality, the integration process is taking a different direction. Rather than workers moving west, it is investment capital that is moving east. This movement of capital, much more that the migration of people, is already shaping Europe’s economic future.”

Substitute “north” for “west,” “south” for “east,” and “America” for “Europe,” and this statement is reminiscent of the debate in the U.S. over the growth of the Sun Belt relative to the Snow Belt. However, both theory and evidence also hold out the prospect of an expansion effect that benefits the E.U. as a whole. The relative strength of the two offsetting effects will be an important determinant in sorting out the net economic gains and losses by region. This is a matter for further study, where the focus should be on relative rates of productivity growth and levels among regions relative in comparison to wages and transport costs.

Table 1

Sources of Output Growth in Three Economies
(Average Annual Rates of Growth or Ratios)

	Indian Manufacturing 1973-92 Gross Output Basis	Indian Manufacturing 1973-92 Value Added Basis	U.S. Manufacturing 1970-86 Gross Output Basis	Spanish Economy 1964-93 Value Added Basis
Real Product	7.3%[1]	7.1%[2]	2.5%[1]	3.7%[2]
Materials	7.4%	——	2.0%	——
Labor	2.1%	2.1%	-0.1%	-0.3%[3]
Capital	6.8%	6.8%	2.5%	1.2%[3]
Total Input	6.9%	5.0%	1.1%	0.9%
Productivity	0.4%[4]	2.1%[5]	1.4%[4]	2.8%[5]

Sources: India: Hulten, Bennathan, and Srinivasan (2003); U.S.: Hulten and Schwab (2000); Spain: Mas et. al. (1998). Detail may not add due to rounding error.

[1] Real gross output
[2] Real value added
[3] Weighted by income shares
[4] Total productivity
[5] Total factor productivity

Table 2

Levels and Growth Rates by State
(Ranked by Terciles)

Rank by 1973[3] Level of Total Productivity	AAGR[1] Gross Out. Q	AAGR[1] Tot. Prod. TP[2]	TP[2] Level 1973	TP[2] Level 1992
Top 5	7.76%	0.42%	1.020	1.105
Middle 5	7.85%	0.45%	0.988	1.076
Bottom 5	9.30%	0.58%	0.927	1.035

Source: Hulten, Bennathan, and Srinivasan (2003)

[1] AAGR is average annual growth rate
[2] TP is total productivity
[3] Ranking by tercile excludes Kerala

Table 3

Regional Sources of Growth[1/]
U.S. Manufacturing Industry
1970-1986

	Sun Belt	Snow Belt	Total
Average Annual Growth Rate of:			
Gross Output	3.75%	1.53%	2.49%
Intermediate Input	3.20%	1.02%	1.99%
Labor Input	1.26%	-1.06%	-0.08%
Capital Input	3.54%	1.57%	2.46%
Total Productivity	1.30%	1.38%	1.34%
Total Productivity Level			
1970	0.9945	1.0027	1.0000
1986	1.2251	1.2505	1.2386
Rate of Return to Capital			
1970	16.4%	15.3%	15.9%
1986	10.7%	9.7%	10.3%
Index of Wage Level			
1970	0.937	1.012	0.970
1986	3.061	3.321	3.177
AAGR of Public Capital	2.09%	1.30%	1.70%
AAGR of Highways	1.43%	0.69%	1.43%

Source: Hulten and Schwab (2000); “highways” estimates are taken from Hulten and Schwab (1991)

Table 4

Parameter Estimates of Basic Model[1]
Comparison of Three Studies
(Elasticities)

	India[2]	U.S.[2]	Spain[3]
Infrastructure Variable[4]	0.044 (2.71)	-0.043 (0.58)	0.101 (2.08)
Time	0.004 (4.81)	0.014 (8.66)	0.024[5]
Scale Variable	0.038 (4.12)	-0.053 (1.24)	0.043 (0.66)
Markup Variable	0.082 (7.31)	0.226 (4.14)	N/A
R-squared	0.809	0.794	0.978

[1] t-statistics in parentheses; state fixed effects not shown
[2] Dependent variable is log total productivity
[3] Dependent variable is log total factor productivity
[4] Infrastructure variable is national and state roads and highways for India, and broader measures of public capital for U.S. and Spain
[5] Arithmetic average over regions of Spain

Table 5

Comparison of Gross Marginal Products
All-India Average for Manufacturing Industry
(Average Gross Return per Rupee of Capital)

	1974	1993
Highways	0.02	0.05
Private Capital	0.29	0.29

Source: Hulten, Bennathan, and Srinivasan (2003)

Table 6

Decomposition of the Growth Rate of Total Productivity
All-India Average for Manufacturing Industry, 1973-1992
(Average Annual Growth Rates)

Core Productivity	0.30%
Highways	0.09%
Subtotal:	0.39%
Scale Effect	0.24%
Markup Effect & Residual Error	-0.23%
Subtotal:	0.01%
Total Productivity:	0.40%

Source: Hulten, Bennathan, and Srinivasan (2003)

REFERENCES

Aschauer, David A. (1989a), "Is Public Expenditure Productive?," *Journal of Monetary Economics*, 23.

Aschauer, David A. (1989b), "Public Investment and Productivity Growth in the Group of Seven," *Economic Perspectives*, Federal Reserve Bank of Chicago, 13.

Berechman, Joseph (2001), "Transport Investment and Economic Development: Is There a Link?" *Transport and Economic Development*, Report of the Hundred and Nineteenth Round Table on Transport Economic, Economic Research Centre, European Conference of Ministers of Transport.

Gramlich, Edward M. (1994), "Infrastructure Investment: A Review Essay," *Journal of Economic Literature*, 33, 3.

Hall, Robert E. (1988), "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96, 5.

Hirschman, Albert O. (1958), *The Strategy of Economic Development*, New Haven: Yale University Press.

Hulten, Charles R. (1994), "Optimal Growth with Infrastructure Capital: Theory and Implications for Empirical Modeling," University of Maryland, unpublished.

Hulten, Charles, and Robert M. Schwab (1984), "Regional Productivity Growth in U.S. Manufacturing: 1951- 1978," *American Economic Review*, 74, 1.

Hulten, Charles, and Robert M. Schwab (1991), "Public Capital Formation and the Growth of Regional Manufacturing Industries," *National Tax Journal*, 44, 4, 1.

Hulten, Charles, and Robert M. Schwab (2000), "Does Infrastructure Investment Increase the Productivity of Manufacturing Industry in the US?" In: Lawrence J. Lau (ed.), *Econometrics and the Cost of Capital,* MIT Press.

Hulten, Charles R., Esra Bennathan, and Sylaja Srinivasan (2003), "Infrastructure, Externalities, and Economic Development: A Study of Indian Manufacturing Industry," World Bank.

Krugman, Paul R. (1998)*, Development, Geography and Economic Theory*. Cambridge, Mass: The MIT Press.

Lee, Kyu Sik, and Alex Anas (1992), "Costs of Deficient Infrastructure: The Case of Nigerian Manufacturing," *Urban Studies*, 29, 7.

Mas, Matilde, Joaquin Maudos, Francisco Pérez, and Ezequiel Uriel (1998), "Public Capital, Productive Efficiency and Convergence in the Spanish Regions (1964-93)," *The Review of Income and Wealth*, 44, 3.

Meade, James E. (1952), "External Economies and Diseconomies in a Competitive Situation," *Economic Journal*, March.

Nadiri, M. Ishaq, and Theofanis P. Mamuneas (1994), "The Effects of Public Infrastructure and R&D Capital on the Cost Structure and Performance of U.S. Manufacturing Industries," *Review of Economics and Statistics,* 76, 1.

Prud'homme, Rémy (2001), In *Transport and Economic Development*, Report of the Hundred and Nineteenth Round Table on Transport Economic, Economic Research Centre, European Conference of Ministers of Transport.

Solow, Robert M. (1957), "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 39.

Walter, Norbert (2004), "The Unavoidable Pains of a Delayed Reunion," *International Herald Tribune*, May 13.

World Bank (1994), *The World Development Report, 1994: Infrastructure for Development*, The World Bank, Washington, D.C.

Young, Alwyn (1992), "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore," in Blanchard, O.J., and S. Fisher (eds.), *NBER Macroeconomics Annual 1992,* MIT Press, Cambridge MA.