Infrastructure and Development:
A Critical Appraisal of the Macro Level Literature

Stéphane Straub

The World Bank
East Asia and Pacific Sustainable Development Department
Operations and Policy Unit
April 2008
Abstract

This survey reviews the existing macro-level empirical literature on the link between infrastructure and development outcomes in a critical light. After providing a general framework that casts the relevant terms of the controversy on the real effect of infrastructure on growth in the context of an aggregate production function, it signals what are the relevant empirical questions to be addressed. This guides the systematic review of a number of empirical studies and the discussion of the main econometric challenges to the identification of the effect of infrastructure on output and productivity. Finally, building on related research, in particular in contract theory and political economy, the paper spells out several promising research avenues.

This paper—a product of the Operations and Policy Unit of the East Asia and Pacific Sustainable Development Department—is part of a larger effort in the department to examine the relationship between infrastructure and economic growth. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at stephane.straub@ed.ac.uk.
Infrastructure and Development: A Critical Appraisal of the Macro Level Literature

Stéphane Straub
University of Edinburgh

1 I thank Michael Warlters for initiating this research and sharing stimulating conversations and comments, and Antonio Estache, Marianne Fay, Douglas Holtz-Eakin, John Moore, Jonathan Thomas and Charles Vellutini, as well as participants in seminars at the World Bank and the University of Edinburgh for helpful comments.
1. Introduction

Infrastructure capital, understood as including transport related facilities (roads, railroads, ports and airports), water and waste water treatment facilities, telecommunications, and energy generation, transmission and distribution, is often mentioned as a prerequisite for the success of development policies and has therefore been for some time an important topic on the agenda of politicians and development practitioners that endow it with many virtues. Since the late 1980s, economists have produced hundreds of papers, mostly empirical, on the subject, mostly making use of macro level data, i.e. cross country or cross state data. To date, despite all this accumulated evidence, the link between infrastructure availability and economic productivity or growth is still subject to considerable uncertainty and debate.

This survey takes stock of the existing macro level literature in order to see how 15 years of sustained research have enhanced our understanding of this major development issue. To assess these contributions, it provides a general framework that spells out the relevant terms of the controversy on the real effect of infrastructure on growth in the context of an aggregate production function. It then signals in the context of this framework what are the relevant empirical questions to be addressed. This guides the systematic review of a number of empirical studies, which seeks to identify to what extent answers have been provided to these key questions and discusses the main econometric challenges to the identification of the effect of infrastructure on output and productivity. Finally, building on related existing research, in particular in contract theory and political economy, it spells out several promising research avenues.

Infrastructure: A Review of Issues

Infrastructure matters first because it provides key final consumption items to households, particularly water and to a lesser extent energy and telecommunications. Overall, a rule of thumb is that between one third and one half of infrastructure services
are used as final consumption by households (Prud’Homme, 2004; World Bank and IDB, 2005).

Moreover, basic services such as water and electricity often occupy a significant fraction of poor households’ budgets. Foster and Yepes (2005) show that households in developing countries spend a significant fraction of their income on water and electricity. For example, in a sample of Latin American countries, households in the poorest quintile often spend more than 5% of their income on water and more than 7% on electricity. In East Asia, figures from ADB, World Bank, JICA (2005) for 2003 show that the average share of total household expenditure spent on water services varies between 0.8% (China) and 3.2% (Cambodia), but can reach up to 16-33% for some of the poorest households in Indonesia for example. As for energy, average spending was 2.9% for Vietnam, 7.6% for China, 9% for Indonesia and 24% for Cambodia. From another angle, looking at the impact of privatization on welfare, McKenzie and Mookherjee (2003) show that service extension to previously unconnected customers resulted in large welfare gains for the poorest households.

The other half of infrastructure services corresponds to intermediate consumption, mostly by firms. For small producers and local firms of developing countries, access to distant markets and contacts with potential clients rely on the existence of a suitable and relatively cheap transport and telecommunication network. Specific channels through which productivity costs may emerge span an array of phenomena that go from the complete inability to access certain markets in some rural areas, to the impact of deficiency in infrastructure sectors on logistic costs and inventory levels (Guasch and Kogan, 2001). Recently, the development of mobile telephony has been shown to have an important effect on the ability to conduct business, for example in remote parts of Africa (Vodafone, 2005). Similarly, electricity is a vital input for many industrial and

---

2 For example, Japanese vehicle manufacturers in Thailand consider that Bangkok traffic congestion increases their costs by raising the amount of parts stock they need to hold (ADB, World Bank, JICA, 2005).
service activities. Deficient electricity networks, plagued by frequent power outages and unstable voltage, may induce high costs and even deter some types of investments.

Indeed, people living in developing countries are well aware of how infrastructure shortcomings affect many aspects of their daily life and work. Tap water that is available only a few hours a day, frequent power outages causing the breakdown of home appliances and machinery, communities that find themselves isolated each time it rains, frequently collapsed bridges, newly constructed roads already full of potholes, expensive mobile phone services as the only option when the hope of getting a fixed line installed is a distant and costly dream, are all common stories in these countries. At the entrepreneurial level, in surveys assessing the investment climate, businesses usually rank deficient infrastructure as an important barrier to their operation and growth. For example, the World Bank investment climate assessment (ICAs) indicate that a large proportion of respondents (between 20% in East Asia and the Pacific, and 55% in the Middle East and North Africa, as well as Latin America) view any of electricity, telecommunications or transport as a major or severe obstacle to doing business. Similarly, 33% of Japanese firms operating in Vietnam consider poor infrastructure as the major obstacle to their business (ADB, World Bank, JICA, 2005). Recently, Escribano and Guasch (2005) have developed an econometric methodology to assess the impact of investment climate assessments (ICAs) variables on firm-level productivity. Applying it to samples from Guatemala, Honduras and Nicaragua (Escribano and Guasch, 2006), they find infrastructure to have a major impact on productivity. For example, they find a one percent increase in the average duration of power outages to decrease productivity by between 0.02 and 0.1%, while access to internet increases productivity by between 11 and 15%.

In this context, there is therefore an intuitive presumption that infrastructure levels and quality matter for firms’ productivity and growth, and that a large part of the output and

---

3 See for example the World Bank World Development Indicators and Briceño-Garmendia, Estache and Shafik (2004) for statistics on access around the world.
productivity differences that we observe across countries could be due to different endowments of such capital. Formalizing some of the stylized facts mentioned above, a standard formulation of this line of thinking would probably start from the argument that key infrastructure services, such as transport, energy or communications, matter in a strongly complementary way to other productive input, and as such may constitute major bottlenecks if not or insufficiently available.5

Such a presumption underlies the inclusion of infrastructure indices in much publicized competitiveness indices such as the World Economic Forum’s Global Competitiveness Report and the IMD’s World Competitiveness Yearbook. However, to the extent that infrastructure indices are used as an input into multivariate competitiveness indices, problems of circular reasoning and potential simultaneity are obviously not absent from these indices, making their use in applied economic research problematic.

However straightforward the presumption might appear, it is subject to considerable dispute based on both theoretical and empirical arguments. Prominent scholars have contested the notion that shortage of capital could credibly account for the large productivity differences between developed and developing countries.6 Prescott (1998) points to differences in the incentive structure prevailing in these countries as a leading candidate explanation. Although he does not specifically discuss infrastructure capital, considering instead general physical as well as additional forms of intangible capital such as human capital, firm-specific learning-by-doing and organization capital, his approach might prove quite useful when applied to infrastructure, as it signals incentive arguments, applied specifically to infrastructure delivery and maintenance, as a potentially important topic that has largely been ignored in the literature. Related points are made by Hall and Jones (1999) and Parente and Prescott (2000). Formal models can be found in

5 An O-ring type of production function (Kremer, 1993), in which capital is made part of the multiplicative formulation in which quantity of any input cannot be substituted for quality of other key inputs like infrastructure, could be used to formalize such an argument.

6 The fact that output-per-worker in developing countries is lower than what is implied by the amount of capital available point is discussed for example in Prescott (1998) and Banerjee and Duflo (2005). This mirrors Lucas (1990)’s initial point that the marginal product of capital implied by differences in output-per-worker between developed and developing countries is higher than the average rate we observe. Note that all these discussions are framed within a Cobb-Douglas type of specification.
Kocherlakota (2001), who focuses on the interaction between limited enforcement and inequality to explain the failure to adopt high-TFP technologies, and Bental and Demougin (2006), who explain productivity differences by endogenous differences in incentive schemes along the development path.

Alternatively, Banerjee and Duflo (2005) stress that there are huge differences in rates of return within countries, so that many firms in developing countries are in fact not taking advantage of available technologies and investment opportunities. They point to “non-aggregative” reasons why this might be the case: government failures (inadequate regulations, excessive interventions), credit constraints and failing insurance markets, intergenerational constraints, among others. If this is the case, adding to the infrastructure capital stock may fail to significantly boost productivity as long as other types of bottlenecks remain.

Although not always grounded in the same theoretical arguments, polar views are also found in the large empirical literature that developed in the last 15 or 20 years to try to assess the real impact of infrastructure on output, growth and productivity. As we will see when reviewing these contributions in Section 3, authors have come up with everything from hugely positive output elasticities, as in the seminal work of Aschauer (1989) that launched a flurry of subsequent research, to zero or even negative elasticities. Of course, results are not always strictly comparable due to differences in the samples and time periods under study (covering alternatively time-series data for a single country, worldwide cross country samples or single country state-level panels) and econometric techniques used. We will discuss below to what extent these disparities may account for the variations in results found in the literature.

On top of an ever growing literature, several surveys have by now extensively reviewed the literature on the effects of infrastructure on output and growth, comparing and contrasting the different methodologies available and the different results obtained. From Gramlich (1994), who focused mostly on the US aspect of the debate to Sturm, Kuper

7 Note that the first point of the list overlaps with Prescott’s argument above.
and de Haan (1998) and Romp and de Haan (2005) among others, these reviews have compared different methodologies and results, try to evaluate to what extent the available results allowed the identification of potential shortages in infrastructure and focused on policy-oriented questions such as the best way to finance infrastructure. Gramlich (1994) had 51 references, Sturm, Kuper and de Haan (1998) 97, of which 33 from 1994 onwards, and Romp and de Haan (2005), building on the previous survey, had 93 references, of which 59 from 1998 onwards. Finally, Prud’homme (2005) is a more methodological survey that discusses several aspects, from the notion of infrastructure itself, to the challenges linked to the assessment of its contribution to economic development.

The structure of this paper is the following. Section 2 introduces the theoretical framework that will be used to structure the discussion of the empirical evidence. Section 3 then reviews a sample of macro level empirical studies in the light of this framework, and discusses the main econometric challenges to the identification of the effect of infrastructure on output and productivity. Finally, section 4 considers potentially fruitful developments integrating incentive arguments and shows how these relate to the key policy questions that are still on the agenda.

2. A Theoretical Framework

At this stage, it is useful to put the discussion on the effect of infrastructure and growth in a common framework. As in most of the literature, this debate can be framed by starting with a general form of the aggregate production function written as:

\[ Q = A \cdot F(K, L, I(K_i)), \]  \hspace{1cm} (1)

Where \( Q \) is real aggregate output, \( K \) is (non-infrastructure) aggregate capital stock, \( K_i \) the infrastructure capital stock, \( L \) aggregate hours worked by the labor force, and \( I(K_i) \) is an
intermediate inputs variable.\textsuperscript{8} \(A\) is here a standard productivity term, which allows for shifts in the production function. In this framework, changes in \(K_i\) lower the cost of related intermediate inputs, resulting in what Hulten, Bennathan and Srinavasan (2003) call a \textit{market-mediated effect} of infrastructure. In this framework, at the end of each period, agents get the return from their investments in physical and human capital and consume the realized output, maximizing some utility function.\textsuperscript{9}

There are several reasons to make infrastructure \(K_i\) enter the production function through the services \(I(K_i)\) provided by this type of capital, rather than simply as an additional factor of production as is often done in the literature (Romp and de Haan, 2005). First, introducing \(K_i\) directly assumes that infrastructure has pure public good attributes and produces services proportional to the stock of infrastructure in a non-rival and non-excludable way. However, this is only partially true as infrastructure is increasingly mediated through the market and has characteristics of standard private goods. In this case, its effect should indeed go, as in (1), through the production of specific services, like transport, communications, etc., that enter firms’ production functions.\textsuperscript{10} Second, despite the increasing market mediation of infrastructure, there is also strong evidence its costs and prices are largely not reflecting “fundamentals” of these activities, so it is implausible that this type of capital is remunerated according to its marginal productivity, even in a world of constant returns to scale.\textsuperscript{11} When the unit cost of infrastructure is not market determined, it is therefore questionable to include it as a factor in the production function, as firms would not be able to make informed decisions on the cost of the amount of infrastructure capital they use (Duggal, Saltzman and Klein, 1999). This has prompted several authors to instead consider that infrastructure is part of the total factor

\textsuperscript{8} Adding subscripts \(i\) and \(t\) would yield an inter-temporal production function at a lower level of aggregation, for example the regional or state level, but we abstract from this as it is not useful to our current discussion. A discussion of the assumptions behind the aggregation process that leads to (1) can be found in Banerjee and Duflo (2005).

\textsuperscript{9} To reconcile this with the fact that part of infrastructure services are directly consumed by households, consider that the composite capital stock of this economy (physical and infrastructure capital) is made of the unique final good, as is usual in such models. Thus in a sense, infrastructure capital is both consumed and used as intermediate input.

\textsuperscript{10} Fernald (1999) adapts this “service” approach to study the impact of the road infrastructure in the US on specific industrial sectors according to their vehicle-intensity.

\textsuperscript{11} See Pritchett (1996) on the issue of costs and prices.
productivity term $A$, for example because it influences productivity by lowering costs or through economies of scales resulting from market expansion. A generic formulation would be:

$$Q = A(\theta, K_I).F(K, L, I(K_I)), \quad (2)$$

where it is made explicit that outward shifts in the efficiency term $A$ may come from two sources: efficiency-enhancing externalities specifically linked to the accumulation of infrastructure capital, and any other type of efficiency-enhancing externalities $\theta$. To sum up, in what follows we will refer to the market-mediated effect, through the intermediate inputs, as the “direct” effects of infrastructure, while the efficiency-enhancing infrastructure externalities will be characterized as “indirect” effects.

Note that at this stage we make no specific assumption on the nature of returns to scale, leaving open the possibility of diminishing, constant or increasing returns. In other words, this framework is compatible with both a neoclassical exogenous growth model and an endogenous growth model. In particular, it can accommodate externalities generating some type of endogenous growth process in the logic of the AK model for example, a point to which we come back below.\(^{12}\)

Simple as it is, this framework allows us to highlight three sets of issues that merit consideration. The first of these deals with the magnitude and nature of the effects of infrastructure on output. This again may be divided into three sub-questions, starting with the obvious one, addressed in most of the empirical literature, namely the simple comparison of the elasticities of output with respect to the two types of capital $K$ and $K_I$.

\(^{12}\) The AK model is a simple endogenous growth model, in which on aggregate there are constant returns to reproducible inputs like capital despite the fact that each firm faces decreasing returns to private capital, so the long-run growth rate depends on the investment rate. Such a dynamic can arise through different mechanisms, for example learning-by-doing externalities when firms’ investments add to the general non-rival stock of knowledge (Romer, 1986) or through government investment in public goods (Barro, 1990).
The second important question is whether there are some sizable indirect infrastructure effects, which raises the problem of disentangling direct vs. indirect effects of infrastructure capital on output and deriving their relative signs and magnitude. Thirdly, a related issue concerns the relative importance of $\theta$ and $K_I$ in explaining the observed shifts in $A(.)$. Advocates of the “infrastructure matters a lot” view would argue that most of the gains in productivity over time ultimately stem from improved infrastructure services ($K_I$), while at the other extreme, it would be argued that other type of externalities ($\theta$), which we discuss in more details below, are responsible for the gains. Between these two extremes, some correlation may exist between both dimensions, so that the potential external benefits from infrastructure services only materialize in the presence of the right set of incentives, for example at the level of the regulatory framework, of the political game, etc. Before discussing the theoretical motivations for these questions, we summarize them formally:

**Question set 1:** Magnitude and nature of the effects of infrastructure on output.

A. Comparison of the output elasticities of $K$ and $K_I$.

B. Can we disentangle direct and indirect effects?

C. If indirect effects can be estimated, what are the respective contributions of generic ($\theta$) versus infrastructure ($K_I$) externalities to shifts in productivity ($A$)? Are there interactions between both?

The theoretical underpinning of growth externalities can be found in several literatures, including the endogenous growth theory and the new economic geography. As for specific infrastructure related externalities, a few examples include higher quality electricity supply making possible the use of more sophisticated machines, and better transport infrastructure that, by lowering transport costs, leads to economies of scale, a different pattern of agglomeration and better inventory management (Hulten et al., 2003; Baldwin et al., 2004). Other potential channels involve the pattern of specialization of

---

agents, as well as incentives to invest in innovation as the transport and communication infrastructure, and therefore access to market, change. Economies of scale due to network externalities are still another important explanation in the case of network industries.

As for more general types of externalities, the main candidate category is a broad one labeled “incentives”. Prescott (1998) mentions two specific historical examples. In the first one, drawn from Wolcott (1994), he argues that productivity differences between Indian and Japanese cotton mills, in the period 1920 to 1938, can be traced back to the ability of workers to resist organizational changes, itself due to differences in the composition of the population of workers, rather than to differences in technology. Similar reasons appear to explain the century-long pattern of productivity changes in strongly unionized coal mining in the US, with surges in productivity only occurring when changes in the environment (competition from low-price oil and from non-unionized mines) presented workers with an alternative between changing work practice to allow the use of more efficient technology or seeing the mines closed. Finally, Prescott argues that the large observed industry-level differences in productivity among developed countries such as the US, Japan, Germany, the UK and France, can hardly be attributed to differences in skills or the stock of useable knowledge, and must instead be related to constraints such as laws, regulations or union power.

Recent contributions that blend endogenous growth theory with contract theory are promising avenues to understand how the incentive structure prevailing in key areas such as R&D or government services affect the growth path of an economy (e.g. Martimort and Verdier, 2004; Sarte, 2002; Erlich and Lui, 1999). As for geographical linkages, Foster and Rosenzweig (2003) provide microeconometric evidence of changes in individual and geographical specialization patterns following the green revolution in India, with industrial investment clustering in low agricultural productivity regions where wages were lower. An intuitive question here is to determine to what extent the changes in regional specialization and the agglomeration of industrial and agricultural activities have been mediated or constrained by the availability of key infrastructure like roads and telecommunications, suggesting the relevance of potential correlations or
complementarities between the different types of externalities. Note however that in this case, the question of the attribution of productivity gains and output differences to infrastructure versus other causes obviously becomes a difficult empirical issue.

This points to another key question implicit in the theoretical framework presented above, namely whether an increase in the infrastructure capital stock will have a permanent or only a transitory effect on growth of per capita income in the specific geographical and temporal setting under study.

**Question Set 2:** Is the effect of additional infrastructure investment a permanent or a transitory one?

Ultimately, the question boils down to whether we believe that infrastructure (or its combination with other policies) generates enough externalities to induce constant returns on aggregate and leads to endogenous growth, in which case it will have a permanent effect on the growth rate, or that we are in a standard neoclassical case with decreasing returns where any investment in infrastructure will only have transitory effects, increasing the level of output but not the long-run growth rate? Note that this is not to say that a transitory positive shock, like the huge infrastructure investment in Vietnam since 1995 (near 10% of GDP on average) or in Thailand since 2001 (above 15% of GDP), would not be desirable. If it has the effect of shifting the economy to a higher level of output, even with the growth rate then going back to its previous level, this may still be a desirable policy, subject to cost-benefit considerations.

Another way to look at this debate is to think in terms of rates of return to infrastructure. If such rates of return are higher in situations of under provision, characteristic of many developing countries, but decline as the stock of infrastructure increases, infrastructure is likely to provide a one time boost to growth but not to have a permanent effect. Alternatively, the network effect story claims returns are higher at higher level of coverage (at least for some type of services/sectors), providing one possible mechanism
through which a one time investment in infrastructure, leading to cross a threshold in terms of service coverage, may have permanent effects on the growth rate.

Finally, the question of the optimality of infrastructure stocks introduces an additional twist in the debate. Indeed, knowing whether the stock of a given type of infrastructure is below or above its optimal point should have important policy consequences.

**Question Set 3:** Can we identify a country’s optimal infrastructure stock?

Given the (at least partial) public good nature of infrastructure, determining the optimal level of the infrastructure stock would require equating the social marginal benefits to marginal costs (Romp and de Haan, 2005). However, as discussed in Pritchett (1996), the determination of the cost of infrastructure is a major challenge. Moreover, there are additional issues with this macroeconomic approach. Indeed, one could imagine situations characterized both by an adequate aggregate stock of infrastructure and specific local bottlenecks. This would arise for example if roads development concentrates in specific regions for political reasons,¹⁴ or if the conditions that make a given stock of infrastructure optimal for private business development vary from regions to regions, as for example labor regulation across Indian states as shown in Besley and Burgess (2004).

As pointed out in Gramlich (1994), there have been several ways of estimating the optimal stock of infrastructure, including engineering assessments of needs, political measures based on voting behavior, measures of rate of return and econometric estimates, and none have provided definitive answers or methodology.

The discussion above raises the question of how the theoretical debate on the different questions we have identified is mirrored at the econometric level. The next section

---

¹⁴ We discuss the evidence for political biases in infrastructure investment in the next section. The repartition of the construction of dams across Indian States (Duflo and Pande, 2007) and of transport infrastructure investment decisions in France (Cadot, Röller and Stephan, 2006) are examples of this.
attempts to systematically organize what can be learned from the existing empirical literature according to this framework.

3. The Empirical Evidence

This section reviews a sample of 80 different specifications from 30 macro level studies, realized between 1989 and 2006, that include some measure of infrastructure as an independent variable and some measure of economic performance (output level or growth, productivity level or growth) as dependent variable. These studies were non-randomly selected as the result of skimming through and reading the literature, so the sample can be considered to be a subset of the most often quoted references. Sixteen of the 30 studies have been published in peer-reviewed outlets like the American Economic Review, the Journal of Monetary Economics or the Journal of Development Economics, others as book chapters or working papers.

A majority of these papers uses cross country data, the rest being either cross-state, cross-region or time series based. Their data sets cover periods that go as far back as 1949, up to 2003, and sample sizes (when not single country time series) vary from 8 to 121 countries/states. In terms of level of development, we find a relatively balanced composition between studies looking either at developed, developing or mixed settings.

In terms of technical characteristics, there is also a lot of variation. Looking at the whole set of specifications, we see that 66% use panel data, and fixed effects are included in 35% of them. The underlying theoretical framework is either a production function (58%), cross country regressions (37%), cost function (1%) or growth accounting (4%).

---

15 In a given paper, specifications testing the effect of different types of infrastructure services (telecom, transport, electricity, etc.) are considered separately, and so are estimations using different techniques (e.g. panel data, then collapsed cross-country) or those testing the effect on different dependent variables. The list of studies included, classified by type of data and techniques used is in the Appendix.

16 A similar “meta-analysis” exercise on a much larger sample of studies is currently being undertaken and will be the object of a forthcoming paper. The insights from the present sample should be considered as illustrative only.
As for the dependent variable being explained, it is either output (60%), output growth (30%), productivity (9%) or inequality (1%). The independent variable used as a proxy for infrastructure is either some measure of public capital (44%) or a physical indicator (56%).

3.1. General conclusions from empirical studies

What are the main questions addressed in the literature under review? The overwhelming majority of specifications (77 out of 80) limit themselves to estimating the output or growth elasticity of infrastructure capital without putting any more theoretical structure on the problem (our question 1.A). The disentangling of direct versus indirect effects is rarely tackled (5 cases), and so is the question of the nature of indirect effects (4 cases). Additionally, 36 specifications also attempt to distinguish permanent from transitory effects, although 22 of these are simply cross-country specifications that by construction imply the estimation of long term effects. Finally, only 5 specifications are concerned with estimating countries’ optimal stocks of infrastructure.

What can we infer from the 80 specifications under study? Overall, a little over half of them (45, equivalent to 56%) find a positive and significant effect of infrastructure, while 30 (38%) find no effect and 5 (6%) find a negative and significant effect. Below, we consider the results of these studies according to the taxonomy of questions outlined in Section 2, and consider specifically how the variations in sample, techniques and type of variables used affect the conclusions.

3.1.1. The Output Elasticity of Infrastructure (Question 1.A)

As mentioned above, most of the available empirical evidence (96%) simply reports some estimate of the output elasticity of infrastructure capital. This is done by inserting some measure of infrastructure in a specification already containing general capital as an
explanatory variable. Table 1 gives an overview of the distribution of results, classifying them as either negative and significant (-1), non significant (0) or positive and significant (+1), and depending on a number of characteristics of the studies, namely sample type, type of dependent and independent variables, and theoretical framework used.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1</td>
</tr>
<tr>
<td>Sample type</td>
<td></td>
</tr>
<tr>
<td>Developed (23)</td>
<td>8.70%</td>
</tr>
<tr>
<td>Developing (22)</td>
<td>9.09%</td>
</tr>
<tr>
<td>Mixed (32)</td>
<td>3.13%</td>
</tr>
<tr>
<td>Dependent variable</td>
<td></td>
</tr>
<tr>
<td>Output (48)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Output growth (24)</td>
<td>16.67%</td>
</tr>
<tr>
<td>Productivity (44)</td>
<td>25.00%</td>
</tr>
<tr>
<td>Other (1)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Independent variable</td>
<td></td>
</tr>
<tr>
<td>Public Capital (34)</td>
<td>14.71%</td>
</tr>
<tr>
<td>Aggregate (27)</td>
<td>18.52%</td>
</tr>
<tr>
<td>Transport (4)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Telecom (2)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Water (1)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Physical Indicator (43)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Electricity (11)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Roads (10)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Telecom (14)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Water (1)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Sanitation (1)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Synthetic (6)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Theoretical framework</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Prod function (46)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td></td>
<td>2.17%</td>
</tr>
<tr>
<td></td>
<td>36.96%</td>
</tr>
<tr>
<td></td>
<td>60.87%</td>
</tr>
</tbody>
</table>

A number of stylized facts emerge from this initial view of the data. Overall, positive effects of infrastructure are found more often in the sample of developed countries, and when the dependent variable is output level rather than output growth or productivity. As for the independent variable, more conclusive results are obtained by studies using physical indicators rather than measures of public capital. Within these categories, looking at the specific sectors for which more than a few studies are included, positive effects are found mostly for telecom, roads and electricity in that order. Finally, studies based on a production function framework reach more positive conclusion that those relying on cross-country regressions. These results are discussed further in what follows when specific related issues like permanent versus transitory effects, or the quality of different indicators of infrastructure are addressed.

As for the value of the estimates found in the literature, a first generation of studies on US state-level data, such as Aschauer (1989), Munnell (1990) and Ford and Poret (1991), found output elasticities of public capital varying between 0.31 and 0.54. Estimates of the marginal product of a unit of public capital from these elasticities are bound to be approximate, as the results are very sensitive to measurement errors in the ratio Y/G, but the rough implication is a marginal product around 100%, meaning that infrastructure would pay for itself in one year or less (Gramlich, 1994). For this reason, these numbers have often been dismissed as unrealistic. In particular, as Gramlich (1994) pointed out, they raise the question of why capital does not flow to infrastructure investment if rates of return there largely outperform those of other types of investments.
A first line of response to this critique relies on industry-level studies. For example, Fernald (1999) similarly estimated huge rates of return on investment in roads for US industries that use roads more intensively: in terms of a Cobb-Douglas specification like the ones used in state-level studies, he found an output elasticity of road investment around 0.35. After noting that this is consistent with the initial results from Aschauer, he argues that the massive interstate highway network built in the 1950s and 1960s generated a one-time boost in productivity (of approximately 1%) rather than a permanent one, also explaining the post-1973 slowdown in productivity. In short, initial large investments in infrastructure may produce very high rate of returns, but this is no guarantee that additional investments would also be characterized by the same returns. In this view, Aschauer’s results adequately captured the pre-1973 period. Additionally, this line of argument coincides well with the idea that once basic infrastructure is in place, adequate investment in maintenance might actually have a higher rate of return than new investment, as argued in Hulten (1996), who uses a cross-country sample similar to that of Easterly and Rebelo (1993) and finds that the impact of an effectiveness index of infrastructure is more than seven time larger than that of public capital itself (see also Rioja, 2003).

Addressing endogeneity sources of infrastructure

Holtz-Eakin (1994) argues that results are substantially modified when econometrically taking into account state- or region-level unobserved effects. Indeed, when introducing fixed state-level effects in his US state panel data, Holtz-Eakin (1994) and Garcia-Milà, McGuire and Porter (1996) find the effect of public capital to be insignificant. Accordingly, in the second generation of studies incorporating these concerns, the positive estimates found were significantly smaller than those of Aschauer, with elasticities around 0.1 to 0.2 (Romp and de Haan, 2005). Note, however, that these

---

17 The total post-1973 US productivity slowdown was about 1.3%. Yeaple and Golub (2004) using an industry level panel across countries similarly find significant positive effects of infrastructure on TFP growth and on industrial specialization.

18 While growth accounting using single-country time-series data implicitly controls for these unobserved region-specific effects, production function-based estimates do not so they need to explicitly incorporate fixed effects.
numbers are still quite high, as for the US case for example they imply rates of return of between 25 and 50%.

This conclusion is supported by a quick look at our sample of studies, as shown in Table 2 below. In panel settings, while three quarter of the specifications yield positive and significant results when fixed effects are not included, and none yield negative outcomes, the use of fixed effects leads to less positive estimation outcomes, and some significantly negative ones.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1</td>
</tr>
<tr>
<td>No (26)</td>
<td>0.00%</td>
</tr>
<tr>
<td>Yes (25)</td>
<td>8.00%</td>
</tr>
</tbody>
</table>

Beyond unobserved effect induced endogeneity, an additional problem that was pointed out in early studies was the potential reverse causality between output and infrastructure investment, with the potential upward bias in results it could generate. Endogeneity caused by reverse causality may not be entirely solved by the use of fixed effects, implying the necessity of some sort of instrumental variable approach.\(^{19}\) Several types of instruments have been used here, including the use of lagged values of the explanatory variables themselves, other lagged variables, and other outside variables. Most of the contributions have addressed this problem by using lagged values of the explanatory variables as instruments. While standard tests, such as Sargan tests, in general seem to support this strategy, it should be considered that lagged variables are only weak instruments. In particular, the effects of infrastructure may take time to materialize, for

\(^{19}\) In a nutshell, this involves the use of some outside variables that are correlated with the potentially endogenous explanatory variable (infrastructure), but not with the dependent variable to be explained (output or productivity growth for example). See Wooldridge (2002) for a discussion.
example if the construction of new transport links or electricity connections only leads to new business development with a significant lag, growth rate themselves have a distributed lag structure, and we are often dealing with relatively small samples, which casts doubt on the asymptotic properties of the IV-estimators (Holtz-Eakin, 1994). A related question may therefore concern the power of the tests involved. A closely related strategy is employed by Devarajan et al. (1996), who use a forward lag structure for the dependent variable. Alternatively, Easterly and Rebelo (1993) use as instruments continent dummies, as well as country level structural characteristics such as population size and share of agriculture in GDP. It is unclear, however, whether these last two are plausibly excluded from the growth regression as they claim.

Table 3 shows that the use of instruments makes a significant difference to the results derived in the context of cross-country regressions, where endogeneity problems are likely to the more acute, while very little difference is observed in the context of panel data studies.\(^\text{20}\)

<table>
<thead>
<tr>
<th>Panel</th>
<th>-1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (51)</td>
<td>No (34)</td>
<td>2.94%</td>
<td>35.29%</td>
</tr>
<tr>
<td>Yes (17)</td>
<td></td>
<td>5.88%</td>
<td>35.29%</td>
</tr>
<tr>
<td>No (26)</td>
<td>No (20)</td>
<td>15.00%</td>
<td>45.00%</td>
</tr>
<tr>
<td>Yes (6)</td>
<td></td>
<td>0.00%</td>
<td>33.33%</td>
</tr>
</tbody>
</table>

An alternative strategy to define instrumental variables is to make use of geographical or industry-level correlations. Examples are found in Holtz-Eakin (1994), who instruments US state level public capital by using other neighboring states' average levels of public

\(^{20}\) Garcia-Milà et al. (1996) are indeed unable to reject exogeneity in a panel of US states.
capital. In other contexts, Guasch, Laffont and Straub (2006) instrument the choice of contractual clauses such as the type of regulation using other countries’ contemporary average adoption rates of these clauses, and in the empirical industrial organization literature, Berry, Levinsohn and Pakes (1995), instrument product characteristics and prices using characteristics and prices of other substitute products. The common idea is that the correlation across regions or industries reflects some common global trends and is orthogonal to specific regional or industry level unobserved effects. However, while such instruments are well suited when the source of endogeneity is the presence of unobserved effects, because say more prosperous states/countries also have characteristics that make them more likely to spend more on infrastructure, their use must be subject to more caution in the case of reverse causation. Indeed, in this last case, such instruments would only be valid if the neighboring states’/countries’ variable used as instrument (e.g. infrastructure capital stock, or some infrastructure sector level indicator) is correlated with infrastructure in the state/country, but not with the output residual. If the instrument is linked to the infrastructure capital stock this would only be true if output has no spatial correlation conditional on observed inputs. In most case, this will hardly be verified as common business cycles are usually observed both at the state and the country level, with explanations including such diverse reasons as technological shocks, political context, etc.

This leads us to consider the two additional causes of endogeneity (see Wooldridge, 2002, for a discussion) that may be related to our theoretical framework. The first one is the potential correlation between infrastructure capital, used as an explanatory variable, and the error term because of omitted variables. To the extent that such estimations will necessarily omit some relevant aspects for which suitable proxies are typically difficult to find, in particular those making up what we characterized as generic types of incentives externalities \( \theta \), the risk is that estimates of the total effect of infrastructure will be artificially inflated. As noted above, while fixed-effects or first differencing may help address this problem under the assumption that the unobserved effects are time invariant,

\[ 21 \text{ I thank Douglas Holtz-Eakin for making this point.} \]
they would fail if these effects vary across time.\textsuperscript{22} In this case, an instrumental approach using the types of variables described in the previous paragraph would be appropriate.

Additionally, measurement errors may also create such bias. Although Garcia-Mila, McGuire and Porter (1996) test for measurement errors using the Griliches and Hausman test on US annual state-level public capital data and conclude to no significant measurement problems, data based on some form of public investment indicators do need to be treated with caution. Indeed, issues of data comparability and measurement errors are likely to be responsible for the huge discrepancy in results across studies.

The first problem is the choice of proxy. An example from the cross-country approach makes the point clear. Barro (1991) and Easterly and Rebelo (1993) find that an increase in public investment in transport and communications leads to higher growth, while Milbourne at al. (2003) find the effect not to be significantly different from zero and Devarajan at al. (1996) find a significant negative effect. In the discussion of their results, Devarajan at al. (1996) partly attribute the discrepancies to the fact that their data differ in many dimensions from those used in other studies. They mention first that Barro (1991) uses different criteria to decide which part of public expenditures are productive or not, second that Easterly and Rebelo (1993) use measures that aggregate government expenditures at all levels and expenditures by public enterprises, while they restrict to central government data, and third that Easterly and Rebelo do not control for level effects of expenditures, thus mixing composition and level effects. This raises the question of the choice of indicators.

\textit{Choice of indicators}

As mentioned above, two main types of infrastructure proxies have been used in the empirical literature: Public capital (based on some monetary measure of public

\textsuperscript{22} Note also that first differencing destroys the long term relationships in the data (e.g. for labour and private capital) so it is unclear whether it still allows their identification (Duggal et al, 1999; Sturm et al., 1998).
infrastructure capital investment) and physical indicators of service production or coverage.

Note first that in a world where in the last decades an increasing part of the infrastructure investment has been made by the private sector, public capital is unlikely to overlap completely with infrastructure investment. To take only a few examples, in the period 1996-2001, the share of total spending in infrastructure corresponding to the public and the private sector respectively amounted to 1.37 and 1.02% of GDP in Brazil, 0.27 and 0.98% of GDP in Mexico or to 2.93 and 4.35% of GDP in Bolivia (Calderón and Servén, 2004). To the extent that we are interested in the effect of infrastructure capital, however it is financed, on growth, rather than in the effect of fiscal policy, this is obviously problematic. If variations of the private share of infrastructure investment across sectors or countries are not random, an assumption that seems likely to be verified, relying on public capital as a proxy for total infrastructure investment therefore introduces a systematic measurement error.

The second problem has to do with the argument found in Pritchett (1996), who argues that whatever the measure of capital stock used, the numbers available overlook the fact that cumulated investment flows are not reasonable proxies for the true effective capital stocks, because the costs of these investments are likely to differ from their values. Justifications for this may include simple government inefficiency, potential corruption, or departure from efficiency for redistributive motives among others.

The weakness of public capital as a proxy for infrastructure services is reflected in the fact that estimations using aggregate public capital fail to find any effect about half of the time, as shown above in Table 1. Partly to circumvent the measurement problems discussed above, raw physical indicators, like kilometers of road, number of phone lines, or electricity generating capacity, have become a standard alternative. However, their

---

widespread use also raises important questions, both with respect to the accuracy and the quality of these measures. Consider for example an indicator supposed to capture the availability of transport infrastructure, “total road length”. The first issue is whether the raw indicator is significant, or should be scaled along or adjusted for different dimensions, such as population, land area, or other geographical characteristics. Moreover, anyone having traveled in a low or middle-income developing country is aware of how widely the quality of “paved road” can vary. Finally, it is well known that political considerations often lead to roads being paved where it serves the ruler or its friends rather than where it is more efficient, as already spelled out by Adam Smith in the 18th century.

Another widely used indicator is electricity generation capacity. Consider the case of Paraguay, a small landlocked country in the middle or South America, which happens to be host to Itaipú (the largest dam in the world, on the Paraná river along the border with Brazil) and Yacyreta (another large dam, lower down the Paraná river along the border with Argentina), leaving aside other smaller dams inside the country. Itaipú, owned together with Brazil, has 18 turbines, one of which alone provides 90% of all the electricity consumed in Paraguay. The rest is given to Brazil under an agreement that stipulates the payment of yearly royalties. As this description makes clear, Paraguay enjoys an electricity generating capacity that widely exceeds its need. However, a closer look reveals that the state of the energy infrastructure network in Paraguay is less than satisfactory. In Alto Paraná, the Paraguayan department where Itaipú is located, only 82% of rural households have electric connections and in a recent stay there, this author experienced six major domestic power outages in twenty days, not to mention the constant voltage jumps that plague the network. Again, the question arise of what other potential measures to use (an alternative might be the number of connected households or firms), and of suitable quality measures, which are notably absent from standard

24 Most studies simply ignore this issue and use the raw indicators.
databases.\textsuperscript{25} Note indeed that electricity services measured in this way is the indicator that more often fails to produce significant results (Table 1).

3.1.2. Indirect Effects (Questions 1.B and 1.C)

Few papers have addressed this question. In our sample, only Hulten et al. (2000, 2005), La Ferrara and Marcelino (2000) and Duggal et al. (1999) do so. Except the last one, these contributions use a growth accounting framework. Indeed, when discussing estimates of the effect of infrastructure, a first issue is simply to recognize that empirical analysis based on some version of a Cobb-Douglas production function approach (this is the case of 15 out of the 17 production function-based studies in our review sample) has in general little to say on the indirect effects, as this specification does not permit a distinction to be made between indirect and direct effects. Growth accounting techniques suffer from a similar problem, as they are unable to discriminate the direct effect of infrastructure for reasons discussed above, namely the difficulty to attribute a price to infrastructure capital. Indeed, as infrastructure is partially a public good, not remunerated at its marginal productivity, its share of the output can only be guessed, which makes the estimates subject to caution.\textsuperscript{26} Note that in most cases, growth-accounting studies find lower levels of infrastructure externalities for more developed countries or regions than for developing ones. For example, applying a similar framework to both cases, Hulten and Schwab (2000) show that US state level data displayed no significant infrastructure externalities on growth, while Hulten et al. (2005) found infrastructure to account for a significant part of TFP growth (about half of it in the case of highways and electricity) across Indian states in the period 1972-1992.

Two remarks follow. First, if the total effect of infrastructure on output is high overall, but externalities are low, it may be that most of the effect of infrastructure is a direct one,

\textsuperscript{25} See Briceño-Garmendia, Estache and Shafik (2004) for a discussion of the issue of quality indicators and an overview of existing data.

\textsuperscript{26} Hulten et al. (2000; 2005) attempt to disentangle these indirect effects for regional US manufacturing data and Indian manufacturing data respectively. They estimate the share of output of intermediate input by assuming it is constant over time.
for example through an AK type of dynamic. Alternatively, it could also be that the production function approach in fact captures as part of the direct effect a significant fraction of the externalities responsible for shifting the productivity term $A$ over time that are not due to infrastructure but rather to other types of factors falling in the “incentives” category. On the other hand, it is not clear to what extent the significant infrastructure externalities found for example in the Indian case are purged of other complementary externalities. Further work is still needed to tell these assumptions apart.

3.1.3. Permanent versus Transitory Effects (Question 2)

Although this appears to be a crucial issue of this line of research, again little convincing evidence emerges. We noted in Table 1 that specifications using output level as dependent variable are generally more supportive of a positive effect of infrastructure than those using either output growth or productivity. Although it does not control for other potentially non random specificities of these studies, this could be interpreted as an indication that transitory effects (shifting the aggregate level of output through a temporary investment shock rather than the long term growth rate) are more often observed.

This point can be further developed by considering a recent example from the empirical literature. Calderón and Servén (2004) have argued that raising Latin American quantities and qualities of infrastructure stocks to East Asian Tigers’ level would generate long-term per capita growth gains of around 3%. Note however that if the claimed causation running from the level of infrastructure stocks to growth rates were literally true, all growth in the US and Europe, which have high quantities and good quality of infrastructure and long-term growth rates of 2-3%, would be attributable to their levels of infrastructure stocks.\(^{27}\)

\(^{27}\) We thank Michael Warlters for pointing this out.
Perhaps another way to look at these estimates is to say that, because of distinct incentive structures between Latin America countries and East Asian ones, their economies are settled at different equilibria that display marked gaps in both infrastructure stocks and output growth. Again, the right objective would be to disentangle the part of these incentive differences that has to do with potential infrastructure externalities from the part that boils down to different types of problems (question 1.C above). One possibility would then be that the global estimate of Calderón and Servén (2004) is biased upwards because it includes part of the effect of what we have called generic incentives ($\theta$). Several contributions have shown for example that differences in the structure and quality of the regulatory framework, the quality of contracts, the political economy of the process, as well as the quality of the local bureaucracy, corruption, etc., have all had an impact on the business environment for infrastructure operators and the efficiency of their investments.²⁸ Moreover, as far as private sector involvement was concerned, Latin America went mostly for concessioning of retail and distribution facilities, while East Asia focused on BOTs for wholesale facilities (e.g. power plants), which raised fewer direct political concerns, and was more successful in managing the financing through its higher savings. So despite Latin America having more mature regulatory frameworks than East Asia, the characteristics of the process were such that Latin America may have experienced more severe incentive and information problems.

Alternatively, we could read the Calderón and Servén’s result as saying that with a huge investment in infrastructure, Latin America would generate additional per capita growth for a fairly long time. Obviously this is different from claiming that a higher stock of infrastructure capital implies a higher steady state growth rate, although transitory effects can look permanent when the transition period extends for long enough. In the limit, a permanent increase in the flow of infrastructure investment may imply an increase in the long-term growth rate. Again, this issue is clearly in need of more research.

²⁸ See for example Guasch, Laffont and Straub (2003, 2006), who show that regulatory quality had a key impact on the wave of concession renegotiations in the Latin America in the 1990s, an occurrence that has notably discouraged private investment in infrastructure there. Other contributions include Wallsten (2001), Dal Bo and Rossi (2003) and Cubbin and Stern (2005).
Finally, Canning (1999) and Canning and Pedroni (2004) have proposed an alternative methodology to address this issue, that partly overlaps with the issue of the determination of optimal stocks of infrastructure. We discuss these studies in the next subsection.

3.1.4. Optimal Stocks of Infrastructure (Question 3)

Despite the seemingly crucial relevance of this topic, it has attracted very little attention in the empirical literature, probably because of the technical challenges it raises. In our review sample, only three papers touch upon it. Canning and Pedroni (2004) consider it in the context of a panel of countries, while Aschauer (2000) looks at a panel of US States. Finally, Cadot et al. (2005) provide indirect evidence on the non-optimality of infrastructure investment decisions by showing that these are mainly politically driven.

Aschauer (2000) assumes a steady-state output elasticity of infrastructure of 0.30, in order to determine whether actual stocks are indeed optimal. As in previous work, however, this study fails to address issues of simultaneity between output and infrastructure, casting doubt on the relevance of the results. A more econometrically sophisticated study is Canning and Pedroni (2004). They start from a Barro-type of growth model, and use panel-based unit root and cointegration tests to determine the sign of infrastructure long run effects. A positive (resp. negative) sign is then taken to mean that the infrastructure stock is below (resp. above) its optimal level.

Two main shortcomings can be pointed out here. First, the meaning of an optimal aggregate stock of infrastructure may in fact be very limited, when instead the effects of specific types of infrastructure services on economic activities is typically a local issue. Ultimately, what is optimal for economic agents will depend on the national environment but also on a string of physical and institutional variables, part of which are of a local nature. Because it fails to inform the crucial question of the spatial distribution of services across regions, districts, etc., any macroeconomic answer is therefore bound to be of limited policy usefulness. Moving to a lower level of aggregation (the state level in
US or Indian data for example) should go some way toward solving this problem, but at the same time, because a large part of infrastructure is both of local and national use (for example roads that go through and connect different regions), and because there might be inter-regional spillover benefits from the infrastructure stock, it obviously raises other difficult questions with respect to allocation of costs and benefits of infrastructure to one local area rather than another.

Second, the results in Cadot et al. (2006) show that in a world of limited resources and non-perfectly benevolent governments, assessing the optimality of past and present infrastructure investment decisions (including the choice between new investments and maintenance of existing ones) would best be based on a positive theory of these decisions. Similar work in the context of developing countries would therefore be extremely useful. We discuss in the last section of this paper how the political economy literature could provide part of the needed ingredients for such an approach.

### 3.1.5. Other Issues

Having looked at the different problems likely to bias estimates of the effects of infrastructure and at how they have been addressed in the literature, we cannot reject the possibility that results may simply differ according to the geographical areas and time periods analyzed. While the level of aggregation should matter little as long as suitable controls are used (Holtz-Eakin, 1994), different pictures tend to emerge for groups of developing countries vs. developed one, or for a given country in the 1950s or 1960s vs. the 1990s. Several theoretical arguments are relevant here. First, as was already mentioned above, as initial investments to establish specific road or communication infrastructure for example may only have a one-time effect on productivity. In particular this should imply that developing countries with acute infrastructure shortages be characterized by high returns to these investments. Alternatively, network effects are often mentioned in the literature as a motive for why specific infrastructure investments may produce higher returns once some minimum critical threshold in term of coverage is
crossed. Note that the two arguments imply opposite effects and that their empirical disentangling may be difficult. Indeed, they can alternatively be used to justify that returns to infrastructure investment are either higher or lower in developing countries.

As for specific results on the relevance of the level of development, some contributions have provided dissenting views. Devarajan et al. (1996), using a sample of 43 developing countries, find the effect of public capital expenditures to be negative. Their interpretation is that developing country governments have been misallocating expenditures resulting in excessive capital spending. This, however, runs in the face of some anecdotal evidence, in particular from Latin America, showing that the capital part of public budgets has repeatedly been sacrificed to current expenditures (Easterly and Servén, 2003; IDB and WB, 2005). 29

Overall, the sample of specifications reviewed displays more systematically positive returns to infrastructure in developed countries. This, together with the fact that some studies find low returns to infrastructure in developing countries, however, does not necessary invalidate the productivity-boosting view and support the network externality one. Instead, it may indicate that the first effect would only materialize in the presence of a set of conditions enabling the development of productive activities, among which the right set of incentives and a minimum critical mass of suitable human capital. An interesting potential question appears therefore to concern interactions between infrastructure and proxies for these effects. We return to this issue in the next section.

Some evidence also suggests the existence of the aforementioned network externalities. Overall, 9 specifications, from 3 papers, claim to explicitly test for the existence of network effects. The first 2 papers (Röller and Waverman, 2001; Torero et al., 2005) simply do so by running estimations on subsamples of richer and poorer countries respectively, which raises among others the problem of sample selection. Röller and Waverman (2001) find that in the case of telecom investment, significant network

---

29 For East Asia, the presumption is that capital expenditure has been largely adequate, but maintenance has been inadequate.
externalities kick in at near universal service level, while Torero et al. (2005) find that the effect is stronger among middle income countries. Using an endogenous threshold panel model à la Hansen in which infrastructure itself is the threshold variable, Hurlin (2005) reaches a somewhat different conclusion. He argues that such effects are relevant especially at intermediate levels of infrastructure development, where the productivity of infrastructure investments is significantly higher than that of other types of investment, while it is not more productive at either low level of coverage or when the network is completed. It is unclear, however, what are the underlying theoretical mechanisms that could give rise to such a pattern.

Finally, a careful discussion of rates of returns to infrastructure requires us to think much more carefully about specific sector and project characteristics.30 Important differences may arise depending on whether the policy focus is on the rate of return of a marginal investment or on the average rate of return over a string of investments, for example over a period of time. While marginal rates are likely to depend on the state of completion of a network and be very case specific (the final investment allowing the termination of a transport corridor would be large, while intermediate ones might be very low), marginal rates would display such variations much less and may correlate better with the prevailing level of coverage, although no simple linear relationship is likely to hold. This clearly calls for a more microeconomic, or at least project-based, approach to the return of infrastructure investment.

Because different development stages may be characterized by important differences in characteristics of demand, incentive structures and in institutional frameworks, the discussion above clearly implies that we may expect different relative magnitudes of direct and indirect effects and, as far as externalities are concerned, different relative relevance of infrastructure vs. generic type of external effects. In the next section, we present the insights on these issues available in existing studies and in the applied theoretical literature, signaling what in our view are some promising areas for future research on the effects of infrastructure on growth and productivity.

30 I thank Marianne Fay for useful comments on this aspect.
4. Incentives

The effect of generic types of incentive externalities is relevant to our discussion on the growth or productivity payoff of infrastructure investment to the extent that some degree of complementarity exists between these and infrastructure externalities, or in other words, if potential infrastructure externalities may only materialize when other conditions are fulfilled. While in this case the attribution of the returns to one or the other source is bound to be arbitrary, the identification of such interactions would be extremely relevant from a policy point of view. In this section, we review three aspects that may exhibit such complementarities with infrastructure, namely regulatory frameworks and market structure, institutional quality and political economy.

Regulatory frameworks and market structure

Because they often have characteristics of natural monopolies, infrastructure sectors are generally subject to a regulatory framework. The theoretical literature has long stressed the role of imperfect information, adverse selection and moral hazard, in determining the second best nature of public regulation (see for example Laffont and Tirole, 1993 and 2000). In this context, the first relevant aspect to our discussion stems from likely regional variations in the extent of information asymmetries and in the commitment power of both infrastructure investors and operators and of host governments. Moreover, departing from the assumption that regulators are benevolent introduces another potential source of disparity in the externalities linked to the regulatory framework (Laffont, 2005). In particular, when governments have weak commitment power and large asymmetries of information exist, there are several ways in which the potential returns from infrastructure investment might be partly or entirely suppressed or appropriated, leading to a shortfall in such investment.
Indeed, when evaluating the returns to the infrastructure investments that do occur (Estache and Pinglo, 2004), it becomes important to consider that problems of contract enforcement, expropriations, and opportunistic renegotiations have in many cases plagued infrastructure projects and are likely be obstacles to the realization of such returns, especially when investment stems from private operators (Guasch, 2004). Guasch, Laffont and Straub (2003 and 2006) have shown that the choice of the incentive structure (price cap vs. rate of return for example), and the fact that a regulator is or is not in place at the signing of the contract, have a major impact on the likelihood of transport and water concession contracts renegotiation in Latin America.

The channels through which weak regulatory frameworks may affect the returns to infrastructure investments are various. By making opportunistic political interference more likely (Guasch, Laffont and Straub, 2006) they increase the likelihood of ex post expropriation of sunk investments, which often results in a degradation in maintenance, a lack of follow-up in planned investments, etc., and may jeopardize the realization of medium term returns. Even in the absence of renegotiations or expropriations, Cubbin and Stern (2005) and Estache and Rossi (2005) have shown that the electricity sector is more efficient in countries that enjoy a regulatory law and higher quality regulatory governance.

Moreover, market structure also appears to matter in conjunction with effective regulation. In those sectors and places where the introduction of competition has been successful, Wallsten (2001), among others, shows that telecommunications are significantly more efficient and reach higher level of coverage. Similar conclusions arise from a review of the UK’s privatization experience, as discussed in Parker (2004) and Newbery (2004). More generally, there is a presumption that the market structure that results from the ownership and regulatory choices imposed on a given sector will crucially affect potential access, prices, etc., all variables that are likely to be relevant to the social return of infrastructure, as shown for example in Estache, Laffont and Zhang (2005).
To our knowledge, the effects of the quality of the regulatory framework along such dimensions, as well as that of market structure, have never formally been considered in the context of studies looking at the effect of infrastructure investment on output or productivity, and may constitute a promising area for research. One important challenge here will be the fact that such aspects must be considered as endogenous. A combination of fixed effects estimations and/or the use of instrumental variables of the type developed in Holtz-Eakin (1994) and Guasch, Laffont and Straub (2006) (see section 3) should allow this problem to be addressed appropriately.

Institutional quality

Some of the aspects mentioned above have to do more generally with the quality of the institutional framework of the host economy (quality of contracts, enforcement, corruption, etc.). Here again, there is a widely held presumption, supported by some sparse empirical evidence, that better functioning institutions should in general contribute to the realization of infrastructure returns. Formally, Esfahani and Ramirez (2003), using a structural model of infrastructure and output growth, display estimates that support the key role of generic institutional capabilities, such as contract enforcement and bureaucratic efficiency, in enabling high infrastructure returns. The important role of such aspects in mediating disruptive events such as renegotiations is highlighted in Guasch, Laffont and Straub (2003 and 2006). Dal Bo and Rossi (2004) stress the fact that greater corruption, as measured by subjective perception indices in a cross-country of Latin American countries, is significantly associated with lower efficiency in electricity distribution.

Moreover, this last study points to an additional interesting aspect, which is the importance of the ownership structure, as a channel through which institutional weaknesses may affect the operation of infrastructure sectors. Indeed, Dal Bo and Rossi show that the effect of corruption is actually stronger for publicly owned firm. Martimort and Straub (2006) develop a model of privatization that highlights sectoral and economy-
wide conditions under which corruption may be higher under private ownership, and relate this to popular dissatisfaction with privatizations in Latin America.

**Political economy**

As mentioned previously, some authors have attributed differences in the estimates of infrastructure returns across geographical units to the fact that some may have exceeded their optimal level of infrastructure stock, while others may still be below it (Canning and Pedroni, 2004; Devarajan et al. 1996; see discussion in de Romp and Haan, 2005). Different conclusions have emerged depending on the econometric techniques used, from simple panel estimations to sophisticated unit root testing. While the question of the optimal stock of infrastructure is obviously an important one (see Gramlich, 1994), what has been missing, however, is a positive theory, and related empirical tests, of the decision to invest in infrastructure that could guide the optimality assessment. In this sense, the most promising area appears to be the political economy approach.

Indeed there is a presumption that decisions to invest in infrastructure, be it directly through public budget or by calling on the private sector through some form of PPI, respond to political motives rather than economic efficiency. Pork-barrel, electoral pandering, etc., provide theoretical channels through which investment may be suboptimal and returns may be affected. Examples of such analysis are in Robinson and Torvik (2005), Maskin and Tirole (2004; 2006) and Dewatripont and Seabright (2005) among others.

These theoretical insights are confirmed by the few empirical papers available. Cadot et al. (2006), specify a simultaneous-equation model, which explicitly considers the political-economy process that drives infrastructure investments. Their results, based on a panel of French regions over the period 1985-92, support the idea that such investments are mostly determined by electoral concerns and interest-groups activities, to the detriment of the maximization of economic returns. In their evaluation of the economic impact of dams in India, Duflo and Pande (2007) conclude that the microeconomic
impact depended on the local institutional framework, and in particular that the lack of redistribution from winners to losers was especially felt in areas where institutions, for historical reasons, favored the politically and economically well-connected agents.

Rauch (1995) showed, using data from US cities in the first two decades of the twentieth century that an educated bureaucracy and adequate formal functioning rules in terms of recruitment, tenure, etc., are instrumental in allowing the choice and successful realization of long-matured infrastructure projects. In his sense, professionalization of the bureaucracy, because to some extent it constrains politicians and limits opportunism, makes more likely a choice of infrastructure close to the optimal one, or at least that long-gestation-period projects are not replaced by short term ones for electoral considerations for example. Finally, Henisz and Zelner (2000) argue, on the basis of cross-country evidence from the electricity sector, that the investment incentives of private firms are affected by a combination of the level of political constraints on politicians and the degree of interest-group intervention in the organization and regulation of the industry.

Again, this potential policy endogeneity presents a challenge for econometric studies. Indeed, if infrastructure investment decisions are the result of an endogenous policy process, the expected signs may not be the obvious ones, a point forcefully made in Rodrik (2005) for the cross-country growth literature. A simple example would be the fact that if countries deciding to increase their infrastructure investment are precisely the ones where the existing stock is deficient, and the resulting improvement to the infrastructure stock is only partial still leaving this group of countries with lower average stocks, the expected sign would then be negative despite the fact that the effect of this investment might be positive for the specific group of countries. This therefore requires specification of the underlying model and may be addressed by one of the IV strategies discussed above (reverse causality or unobserved effects) or by the use of simultaneous-equation models, as in Cadot et al. (2006).
5. Conclusion

This survey has reviewed the existing macroeconomic level literature on the link between infrastructure and development in a critical light. It has shown that this literature suffers from several related problems. First, it often fails to lay down clearly the relevant theoretical questions to be addressed. Second, it also tends to ignore the fact that most relevant answers, from a policy point of view, cannot be meaningfully addressed with the type of data available. Within these limitations, the last section has highlighted some of the potential areas for further work, bridging the gap with related literatures such as the ones on regulation and political economy. It is clear, however, that major advances of the knowledge in this area will require both more theory and better data sets, that go beyond the macroeconomic level, to combine the existing insights with those from sector- and project-level microeconomic studies and allow policy makers to better assess the potential linkages between specific infrastructure investments and growth, and chose the right composition and sequencing of these investments. Assessing the corresponding priorities for research and data development is the object of ongoing follow-up work.
Appendix: Summary of empirical studies reviewed.

<table>
<thead>
<tr>
<th>Studies based on cross-country data</th>
<th>Number of specifications considered</th>
</tr>
</thead>
</table>

### Production function

- "Infrastructure’s Contribution to Aggregate Output", Canning, D., 1999  
- "Network Effects of the Productivity of Infrastructure in Developing Countries", Hurlin Christophe, 2005  
- "How much does infrastructure matter to growth in Sub-Saharan Africa?", Antonio Estache, Biagio Speciale and David Veredas, 2005  
- "Institutions, Infrastructure, and Economic Growth", Hadi Salehi Esfahani and Maria Teresa Ramirez, 2003  
- "Infrastructure Capital and Economic Growth: How Well You Use it May Be more Important than how much You Have", Charles Hulten, 1996  
- "The Impact of Telecoms on Economic Growth in Developing Countries", Waverman L. Meschi M. and Fuss M., 2005  
- "The Social Rate of Return on Infrastructure Investment" David Canning and Esra Bennathan, 2002  
- "The Effect of Infrastructure on Long Run Economic Growth", David Canning and Peter Pedroni, 2004  

### Cross-country regressions

- "Economic Growth in a Cross-Section of Countries", Barro R., 1991  
- "Infrastructure Investment and Growth, some Empirical Evidence", Blanca Sanchez-Robles, 1998  
- "I just ran 2 million regressions", Xavier Sala-i-Martin, 1997

### Studies based on cross-state or cross-regional data

38
**Production function**

"Public-Sector Capital and the Productivity Puzzle", Holtz-Eakin D., 1994 2
"Infrastructure and Productivity: A Nonlinear Approach", Duggal V., Saltzman C. and Klein L., 1999 1
"Contribution to Productivity or Pork Barrel? The two Faces of Infrastructure Investment", Cadot, O., L.-H. Roller and A. Stephan, 2005 1
"Is Public Expenditure Productive?", David Aschauer, 1989 2
"Infrastructure and Private-Sector Productivity", Robert Ford and Pierre Poret, 1991 1

**Growth accounting**

"TFP, Costs, and Public Infrastructure: An Equivocal Relationship", Eliana La Ferrara and Massimiliano Marcelino, 2000 3
"Does Infrastructure Investment Increase the Productivity of Manufacturing Industry in the US?" Charles Hulten and Robert Schwab, 2000 1
References


Sala-i-Martin (1997), “I just ran two millions regressions”, \textit{American Economic Review}.


World Bank and Inter-American Development Bank (2005), “Diagnosis and Challenges of Economic Infrastructure in Latin-America and the Caribbean”, mimeo.