INTERNATIONAL MONETARY FUND

Benchmarking Infrastructure Using Public Investment Efficiency Frontiers

WORKING PAPER

Javier Kapsoli, Tewodaj Mogues, and Geneviève Verdier

WP/23/101

2023
MAY

IMF Working Paper
Fiscal Affairs Department

Benchmarking Infrastructure Using Public Investment Efficiency Frontiers
Prepared by Javier Kapsoli, Tewodaj Mogues, and Geneviève Verdier

ABSTRACT: With limited financing options, increasing investment efficiency will be a critical avenue to building infrastructure for many countries, particularly in the context of post-pandemic recovery and rising debt emanating from higher energy costs and other pressures. Estimating investment efficiency, however, presents many methodological pitfalls. Using various methods——stochastic frontier analysis, data envelopment analysis (DEA), and bootstrapped DEA—this paper estimates efficiency scores for a wide range of countries employing metrics of infrastructure quantity and utilization. We find that efficiency scores are relatively robust across methodologies and data used. A considerable efficiency gap exists: Removing all inefficiencies could increase infrastructure output by 55 percent overall, when averaging across 12 estimation approaches—in particular, by 45 percent for advanced economies, 54 percent for emerging countries, and 65 percent for low income countries. Infrastructure output would increase by a still-sizeable 30 percent if instead of eliminating all efficiency, countries achieved the efficiency level of their income group's 90th percentile.

Benchmarking Infrastructure Using Public Investment Efficiency Frontiers

Prepared by Javier Kapsoli, Tewodaj Mogues, and Geneviève Verdier[1]

Contents

I. Introduction

Across the world, accumulated investment spending has been eroding. Advanced economies' public capital stock (as a share of GDP) has been on a steady long-term decline over decades (Figure 1).[2] In low-income and emerging economies, while the public capital stock steadily increased in the 1970s and most of the 1980s, this was followed by a sharp dip and then a precipitous decline in the 1990s and 2000s, although some mild recovery took place after the 2008 financial crisis. Much of the changes in capital stock can be explained by the behavior of public investment, which, despite strong infrastructural needs to adapt to and mitigate climate change, has been on a decline in AEs stagnated in LIDCs, and been relatively volatile in EMEs. Broadly speaking, public investment ratios of each income group follow a time trend similar to that of the capital stock but preceding it by a few years.

This trend predates the Covid-19 pandemic, which most likely exacerbated the decline in public investment in emerging market and low-income countries, as governments focused on an urgently needed sanitary response and other spending measures needed to protect vulnerable households and firms. Meanwhile, in advanced economies such as European Union countries, public investment is expected to rise in the years following pandemic restrictions, as recovery packages include sizeable resources for infrastructure (Brasili et al., 2022). Public investment could have a prominent role in underpinning the recovery: It may have particularly high multipliers by serving as a strong catalyst of private investment, especially in the current economic environment strongly characterized by uncertainty (IMF 2020), and green investments may have particularly high multipliers (Batini et al., 2021, Hepburn et al., 2020).

Figure 1. Trends in Public Investment and Capital Stock

a. Public Investment   b. Public Capital Stock
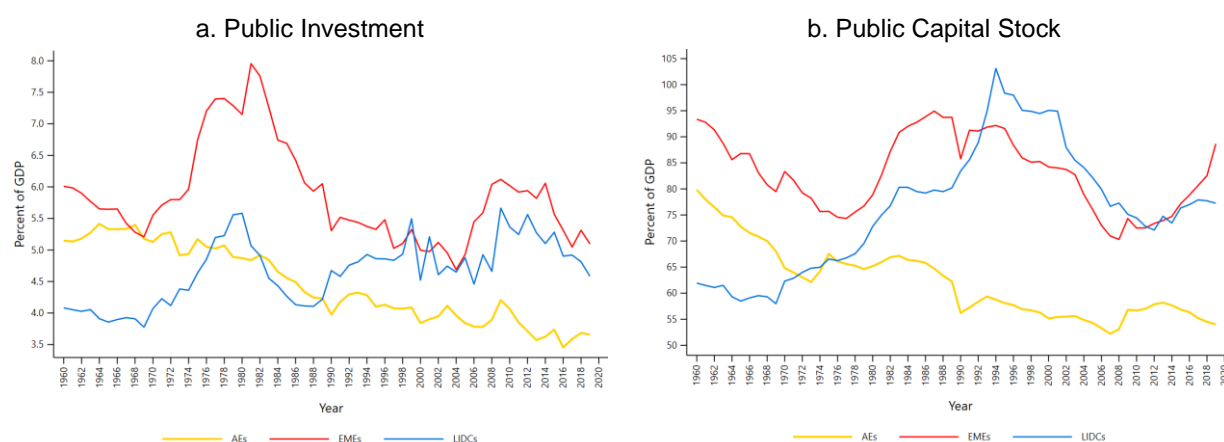
Source: Authors' calculations based on IMF's Investment and Capital Stock Dataset (Xiao et al., 2021).

---

[2] We use the IMF Investment and Capital Stock Dataset available at https://infrastructuregovern.imf.org. For a methodological explanation of these estimates, see Xiao, et al., (2021).

With the end of the historically low interest rate era—as policymakers tighten monetary policy in reaction to rising inflation in the wake of global shocks (IMF 2022c)—emerging market economies (EMEs) and low-income economies will face severe challenges to ramp up much-needed public investments. With high debt levels exacerbated by the pandemic (IMF 2022a), access to capital markets to obtain financing for investment projects will remain highly constrained in many EMEs and LICs.

For these countries, readying their infrastructure to bolster the economic recovery in the wake of the pandemic will require, among other things, strengthening the efficiency with which capital expenditures generate infrastructure outputs. Investment inefficiencies can arise from many factors, including poor public investment management systems (Baum, et al., 2020), political interference in project selection (Marcelo et al., 2016), corruption (Tanzi and Davoodi, 2002), and weak medium-term fiscal planning.

A necessary step to address this critical issue is gauging the size of the problem. Benchmarking could help, allowing the identification of countries that perform best in developing infrastructure using public investment as an input. Benchmarking is based on the idea that best performers shape a "frontier", which will be used to assess the efficiency of all other countries. While there is abundant literature on benchmarking health and education spending, such exercises are less abundant for public investment.[3] Albino-War et al. (2013) applied benchmarking methods to measure the efficiency of public investment by using the World Economic Forum's quality of infrastructure survey as output. IMF (2015) and Baum, et al. (2020) proposed a more comprehensive approach on the measurement of the efficiency of public investment by using indicators of social infrastructure alongside the traditional physical infrastructure outputs.

Estimation of public investment efficiency has come to play a role in IMF surveillance reports (e.g., IMF 2022b) and in providing advice to member countries in IMF capacity development. For example, the IMF's Public Investment Management Assessment (PIMA) uses efficiency scores as a starting point to discuss how governance reforms can increase efficiency and growth (IMF, 2015).[4] There are, however, drawbacks to the use of estimated efficiency scores on their own.[5] In particular, they can potentially vary depending on data availability and methodology used. In addition, they can often leave out important but exogenous determinants of efficiency—e.g., geography or climate. The purpose of this paper is to provide a careful and systematic exploration of analytical considerations when selecting and using inputs and outputs to public investment efficiency analysis. This paper adds to the literature by expanding the methodology proposed in IMF (2015), updating the efficiency estimates, and providing additional estimates using different methodologies and models. The empirical analysis mainly serves to explore the implications of applying a selection of the techniques discussed, to the extent that data availability allows. In so doing, we focus on "core" physical infrastructure; this

---

[3] IMF (2020). Fiscal Monitor, October 2020.

[4] See Herrera and Pang (2005) and Grigoli and Kapsoli (2018) among others.

[5] Due to this, PIMAs usually discuss these drawbacks and complement countries' efficiency scores with more granular information and additional indicators.

choice does not lessen the importance of social infrastructure but acknowledges that finding proper metrics for it requires a number of assumptions. The paper, however, also compares these baseline results with those arising from a wider definition of infrastructure, as described later.

The results suggest that efficiency scores as currently estimated do consistently capture the same phenomenon. While aggregated average efficiency scores vary by income group, estimation method, and measures of investment spending and outcomes—efficiency gaps are large and range between about one-quarter to nearly 80 percent—we find that individual efficiency rankings are broadly robust to estimation methods. More disaggregated, granular information, however, does provide a more complete picture. Our results using sectoral public investment spending in Latin America suggest that the use of aggregated data may produce somewhat overestimated efficiency scores. In addition to improvements in data quality, future work should focus on the drivers of inefficiencies.

The remainder of this paper is organized as follows. The next section discusses the benchmarking methodology, including several analytical approaches to gauge the efficiency of public investment. The paper provides in Section II a fine-grained discussion of key analytical considerations in the selection and use of inputs and outputs in investment efficiency analysis. In Section III, we present the empirical results of the paper, with an emphasis on exploring the application of selected techniques discussed in previous sections. The final section concludes and proposes tasks for future research.

II.   Measuring the Efficiency of Public Investment

A. Basic Concepts

Benchmarking is the systematic comparison of the performance of one unit against others. The methodology relies on the principles of production economics, and compares decision-making units (DMUs), under the assumption that they are implementing the same transformation process using inputs to produce goods and services (outputs). DMUs can be firms, industries, companies, or, as in this paper, countries.

A critical component of the benchmarking methodology is the implementation of performance evaluations, of which one element is the concept of efficiency. Efficiency is measured by identifying the best-performing units in a sample of DMUs and using them to build an efficiency frontier, which defines the best outcomes (relative to inputs) that are possible for a given technology. The performance of all other DMUs is then assessed vis-à-vis the frontier. Units are deemed inefficient if they are located below the frontier.

The modern literature on estimating efficiency scores started with Farrell's (1957) seminal paper, which establishes two types of efficiency, technical and allocative. Figure 2 illustrates both concepts assuming a production function with two inputs x_1 and x_2 that produce a fixed amount of output, Y_0. The AA' curve represents the different combinations of inputs that an efficient unit would use to produce Y_0, and Q^{IE} is an input bundle of an inefficient unit that produces Y_0. If this unit were, instead, able to reduce its use of inputs by the same proportions to arrive at point Q^{TE}, it would become technically efficient. Thus, the ratio OQ^{TE}/OQ^{IE} is a measure of the technical inefficiency of Q^{IE}, meaning that the distance $Q^{TE}Q^{IE}$ could be saved if inputs were used efficiently. The latter is a view of efficiency entirely based on the technical capacity to obtain the same level of output while consuming the minimum amount of inputs.

Figure 2. Technical and Allocative Efficiency
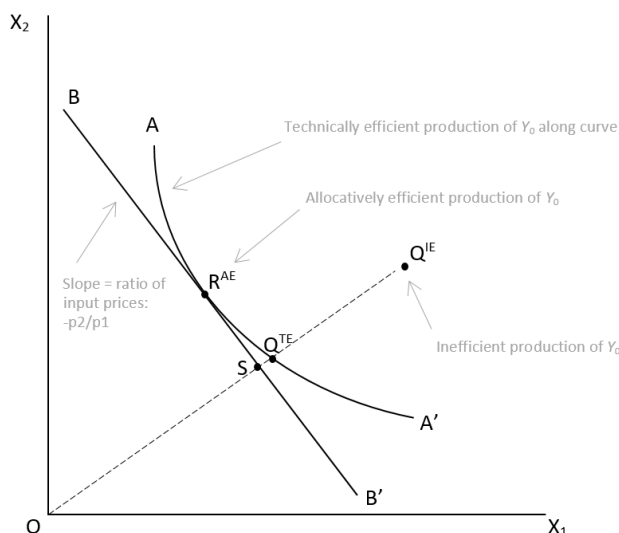
We can also see efficiency from a cost-minimizing perspective. Let p_1 and p_2 be the prices of inputs x_1 and x_2 and therefore $-p_2/p_1$ the slope of the budget constraint. BB' constitutes the budget line corresponding to the lowest possible total budget that enables the application of an efficient input mix to produce Y_0. Then, R^{AE} is not only a technically efficient, but also an allocatively efficient input bundle. For the unit that employs inputs associated with point Q^{AE}, the ratio OS/OQ^{TE} is a measure of its allocative or price efficiency. Allocative efficiency thus measures the amount of resources that could be saved if, given input prices, the consumption of inputs were used to minimize the total cost. In this vein, the cost efficiency of producing at point Q^{IE} is measured by OS/OQ^{IE}—cost efficiency is equivalent to multiplying allocative and technical efficiency. Examples of cost efficiency analyses in the infrastructure sector include e.g., Wheat, et al. (2019) for railways, and Agrell, et al. (2005) for electricity. The lack of comparable multi-country data on prices of inputs used in infrastructure creation renders difficult the estimation of allocative or cost efficiency measures. In this paper we focus on the estimation of technical efficiency.

Technical efficiency can be estimated using input-oriented or output-oriented models. In an input-oriented model, the efficiency scores measure by how much inputs could be scaled back while leaving the level of outputs unchanged. Conversely, efficiency scores from an output-oriented model measure how much outputs could be boosted while leaving input consumption unchanged. The choice of orientation does not change the frontier, nor which DMUs are deemed efficient, but it affects the scores of inefficient units. In certain investment efficiency analyses, the use of input-oriented measures is commonly justified by the fact that the decision makers do not have much influence over the amount infrastructure to be created and must therefore achieve efficiency by cutting costs. For example, Fritzsche (2019) uses the input efficiency approach arguing that counties in Germany (the DMUs) have a fixed set of obligations on the length and type of roads to build. On the other hand, in the context of other types of infrastructure, the inputs may be relatively fixed. In this vein, Lozano and Gutiérrez (2011) find the output-oriented approach more suitable to studying airport efficiency, where the inputs are characteristics of airport capacity that cannot be varied in the short run.

B. Parametric and Nonparametric Methods to Gauge Investment Efficiency

Two families of methodologies exist to estimate technical efficiency—parametric and non-parametric—each with its advantages and drawbacks. Parametric methods use econometric models, which require assumptions on the distribution of the stochastic errors and the functional form underpinning such models. This approach has the important merit that it assumes a stochastic relationship between inputs and outputs, implying that the estimated efficiency scores discriminate between real inefficiencies and statistical noise stemming from measurement error, omitted variables, and other issues in the data. In their study of the efficiency of airports, Martín, et al. (2009) remark on the advantage of stochastic frontier analysis in that the estimation is more robust to noise and various uncontrollable factors, rendering it more suited for drawing sensible policy recommendations.

Conversely, non-parametric methods are based on mathematical programming and therefore do not require any assumptions on stochastic distributions or functional forms. This has especially been an attractive property when assessing the efficiency of broad categories, such as total transport infrastructure (Kyriacou, et al., 2019), for which there is limited guidance on the functional form of the input-output relationship. Unlike statistical methods, nonparametric results are not as easily compromised by a small sample size—an important consideration when, for example, there are only a limited number of DMUs (see for example Wanke (2013) who studies the efficiency of 27 ports in Brazil using nonparametric methods).[6]

The main shortcoming of non-parametric models is the fact that they are deterministic, and as such their estimated scores reflect all factors affecting the efficiency of DMUs, including relatively uncontrollable factors

---

[6] The sample size in a prominent nonparametric method—data envelopment analysis—should be, at the minimum, the larger of ($I \cdot O$) and [$3 \cdot (I + O)$], where I and O signify the number of inputs and outputs used, respectively (Cooper et al. 2002).

such as weather, topography, etc. Because of this, there is a risk that a DMU could be deemed technically inefficient due to factors not subject to change by the decision-making unit (Section II.C discusses some approaches to account for such non-discretionary variables). Additionally, and for the specific case when DMUs are countries, measurement and reporting errors and coverage issues often cause data outliers. In the nonparametric approach, these outliers (extremely high outputs or extremely low inputs), can strongly shape the frontier, leading to an underestimation of other DMUs' efficiency. This problem generates a consistent but biased estimator of the true frontier.[7]

B.1. Data Envelopment Analysis

Data envelopment analysis (DEA) is the most widely used non-parametric method in the investment efficiency literature. DEA solves a mathematical programming application by executing the two main tasks involved in a benchmarking exercise: calculating a piecewise-linear frontier based on the best performing units and evaluating the performance of all DMUs relative to the frontier.

Assuming I DMUs, N inputs, and M outputs, the mathematical model underpinning a DEA solution—in the case of output-oriented efficiency analysis—is:

$$\max_{\lambda,\phi} \phi_j \tag{1a}$$

subject to:

$$\phi_j \, y_{jm} \leq \sum_{i=1}^{I} \lambda_i y_{im} \tag{1b}$$

$$x_{jn} \geq \sum_{i=1}^{I} \lambda_i x_{in} \tag{1c}$$

$$\lambda_i \geq 0 \tag{1d}$$

$$\sum_{i=1}^{I} \lambda_i = 1 \tag{1e}$$

where y_{jm} and x_{jn} are, respectively, output m and input n for DMU j, and λ_i are non-negative weights (constraint in equation 1d). The solution of this program is a peer group of DMUs satisfying the conditions that they should produce a factor ϕ_j (which is greater or equal to 1), pertaining to DMU j, of output relative to the analyzed

---

[7] See Kneip, Park, and Simar (1998) for the consistency proof of DEA estimators.

DMU's output (constraint in equation 1b) but with at most the same level of input consumption (constraint in equation 1c). ϕ_j would then be the maximized efficiency coefficient, and $1/\phi_j$ is the technical efficiency score that ranges from 1 (fully efficient) to 0 (fully inefficient).

Without imposing the convexity constraint (equation 1e), the program would assume constant returns to scale (CRS), implying that all DMUs are operating at an optimal scale. This is a strong restriction for real world applications, particularly if the DMUs are countries. As explained by Coelli et al. (2005), because of imperfect competition, financial constraints, diverse institutional frameworks, and other factors, DMUs usually do not operate at their ideal scale. To manage this problem, we estimate DEA models based on a variable returns to scale (VRS) assumption, which allows each DMU to be compared to others of similar size (see Murillo-Zamorano, 2004).

B.2. Stochastic Frontier Analysis

Stochastic frontier analysis (SFA) constitutes the flagship approach of the parametric family of frontier estimation methods. As in any parametric model, the SFA is based on assumptions about the model's functional form and its underpinning distributions. The canonical SFA model is:

$$ln(y_i) = f(x_i; \beta) + \varepsilon_i \tag{2a}$$

with

$$\varepsilon_i = v_i - u_i \tag{2b}$$

where y_i is the output of DMU I; $f(\cdot)$ is the frontier output of unit i; x_i and β are each an n-by-1 vector of inputs and parameters, respectively; v_i is a zero-mean random error; and $u_i \geq 0$ is a measure of inefficiency. Note that equation (2a) defines the frontier restriction. If $u_i > 0$, then unit i is technically inefficient by the magnitude u_i, and if u_i equals zero, unit i is technically efficient, i.e., on the frontier. Then, by definition, the expected value of the composed error term ε_i is negative. This fact can be used to test if a particular empirical model is suitable for the application of the SFA methodology: The OLS residuals are tested under the null hypothesis of normality versus the alternative of non-normality due to skewness (see D'Agostino, et al., (1990) for details on the test).

Estimating model (2) requires the imposition of a functional form on $f(x_i; \beta)$ and assumptions on the stochastic distribution of u_i. The model's parameters are estimated via the maximum likelihood method. In applications of this method, v_i is nearly universally assumed to have an i.i.d. Gaussian distribution, i.e.:

$$v_i \sim iid\ N(0, \sigma_v^2) \tag{2c}$$

In contrast, alternative considerations are possible regarding the stochastic distribution of the inefficiency parameter u_i. A common prior on its statistical properties is the half-normal distribution:

$$u_i \sim iid \, N^+(0, \sigma_u^2)$$ (2d)

The half-normal model assumes a zero mean for the inefficiency term. If the estimated variance $\hat{\sigma}_u^2$ is low, most of the inefficiency terms will be proximate to zero, meaning that most DMUs would be close to efficient.

B.3. Bootstrapping the DEA model

The standard DEA frontier can be defined as an estimation of the true frontier based on a single sample drawn from an unknown population. Under this interpretation, we can use bootstrapping to overcome the above-mentioned problems of the DEA method (Simar and Wilson, 1998, 2000). As the DEA frontier is based on best-performing units, it captures only the lower bound of the true frontier. This, by definition, generates an upward bias in the estimated efficiency scores. Bootstrapping corrects the bias in the efficiency scores. It is a statistical method based on the generation of an artificial dataset obtained by sampling with replacement from given data, with these samples used to calculate statistics called "replicates." The procedure is repeated many times, each time generating new replicates until we have a sample of them. Based on this sample we can infer conclusions on the distribution of the original data under the assumption that it mimics the distribution of the bootstrapped sample.

If we assume that the distribution of the difference between the DEA and the bootstrapped DEA efficiency scores mimics the distribution of the difference between the estimated and the true efficiency scores, we can estimate the bias, correct the efficiency scores, and establish their confidence intervals (for further details on the use of bootstrapping in the DEA context see Bogetoft and Otto, 2011). The method has been used in investment efficiency analysis, including in Cavalieri et al. (2017) analyzing the efficiency of health infrastructure, and Lorenzo and Sánchez (2007) focused on street lighting infrastructure.

III. Outputs and Inputs for Investment Efficiency Analysis

In the literature on infrastructure efficiency, analyses in which the decision-making units are countries are relatively uncommon. Instead, the vast number of studies are undertaken below the country level, with assessments of how efficiently units such as subnational governments, private companies, or parastatals within a given country create infrastructure and provide associated services such as electricity distribution in Sweden (Agrell, et al., 2005), port services in Mexico (Estache and González, 2002), or energy in China (Lv, Yu and Bian 2017). Several studies also blend cross-country and country-level analyses by considering decision-making units such as cities or firms that span multiple countries (e.g., Pina and Torres, 2006; Growitsch, et al.,

2009). Thus, discussions in this section on the choice of inputs and outputs in the infrastructure efficiency literature will draw from diverse empirical scopes, although the paper's empirical analysis uses countries as DMUs.

A. Outputs: Quantity, Quality and Utilization

'Infrastructure' refers to the physical structures that serve as the underlying foundation for the functioning of an economy. In its narrow definition, this comprises physical facilities such as transportation systems (rail, roads, airports, etc.), communications systems (telephone lines, broadband), power systems (electrical grids, dams, etc.), and water provision and treatment (e.g., irrigation and sewage). A wider scope of the concept of infrastructure can include buildings that support production and human capital development, such as office structures, factories, schools, and hospitals. The term has been extended into multiple spheres, through concepts such as institutional infrastructure, financial infrastructure, and so forth. In a review of infrastructure investment, Gramlich (1994) considers that from an economic perspective it makes sense to focus on a tighter definition including public-sector tangible capital-intensive facilities such as highways, water and sewage systems, hospitals, and school buildings. An early study introducing a cross-country panel database of global physical infrastructure includes the number of telephones and telephone lines, kilowatts of electricity generating capacity, and kilometers of roads, paved roads, and railway lines (Canning 1998).

Infrastructure output can manifest, and thus be measured, in a number of ways. Generally, outputs in efficiency analysis can be classified into three types: measures of quantity, of quality, and of utilization. Quantity is reflected, for example, in road length (Wheat, et al., 2019) or road area (Fritzsche, 2019), gigawatt-hours of electricity supply (Growitsch, et al., 2009), or length of railway routes (Kyriacou et al. 2019). It is also common to include multiple outputs in the same efficiency analysis in order to reflect both quantity and quality of infrastructure, as done, for example, in IMF (2015) and Baum, et al. (2020). This is especially relevant in transport, where quality may be proxied by road condition indexes based on information about cracks, unevenness, etc. (Fritzsche, 2019), the share of paved roads (Kyriacou et al. 2019), the number of accidents caused by bad road conditions (Kalb, 2014), the proportion of roads for which maintenance should be considered (Wheat, et al., 2019), the average speed of buses for bus transport efficiency (Pina and Torres, 2006), and the travel mean speed between large cities (Moszoro and Soto, 2022).

Finally, the third type of output in infrastructure—measures of utilization—reflects the extent to which consumers, passengers, etc., access and make use of the infrastructure. The literature has captured this dimension, for example, as the count of workload units (number of passengers and of tons of cargo transported in an airport) (Martín, et al., 2009), the sum of distances travelled by each train passenger and by each ton of freight (Coelli and Perelman 1999), irrigated agricultural land area in a study of water facilities (Ali and Klein, 2014), the number of electricity-receiving customers (Growitsch, et al., 2009), or the annual bus-seat kilometers and bus boardings (Pina and Torres, 2006). In certain contexts, it may be difficult to disentangle supply and

utilization, and here the output considered may reflect both dimensions, such as the number of telephone connections in telecommunications (Bartels and Islam, 2002) and the volume of merchandise handled in ports (Estache and González, 2002).

B. Single (Composite) Output versus Multi-output Analysis

When the inputs to investment are available only in aggregated form, for example as the total public capital stock, and yet outputs can be measured as distinct types of infrastructure, such as road, electricity, and water facilities, consideration needs to be given on how to link outputs to inputs. To mitigate the extent of a mismatch, it would be ideal to avoid assessing the efficiency with which governments translate total public investment into, say, a better road network, ignoring the other outputs that result from public investment.[8]

There are generally two ways to proceed in such a context. One is to accommodate multiple output variables into the same efficiency analysis. For example, Coelli and Perelman (1999) derive a railway country-level efficiency score by accounting for two outputs: the sum of distances travelled of each passenger, and the sum of distances travelled of each ton of freight. Another approach is to aggregate multiple outputs into one. Baum, et al. (2020) create a composite output by first standardizing, then averaging over variables for roads, electricity, water, and proxies for health and education infrastructure.

In the case of multi-output analysis, DEA places greater weight on those outputs on which a DMU performs better, usually resulting in higher efficiency scores than when these output variables are aggregated. This is seen in Rayp and Sijpe (2007), who consider a total of five outputs and conduct their efficiency analysis with different degrees of aggregation: One analysis conducts a multi-output efficiency analysis with five outputs, a second aggregates four of the variables into two pairs (resulting in a total of three outputs), and a third aggregates all five variables into one output. The mean efficiency score is higher when more disaggregated outputs are used. The same pattern is seen in Fritzsche (2019), who considers two different outputs of road quantity and quality in DEA analysis, and then combines these outputs in aggregated form. In the results, the number of fully efficient DMUs is higher when outputs are incorporated as separate variables.

C. Inputs: Stocks versus Flows, Direct versus Indirect Inputs

Investment efficiency analyses most commonly express inputs in monetary terms, e.g., as countries' total investment spending on roads, rail, waterways, ports and airports in a cross-country study on transport (Kyriacou et al., 2019). A financial measure enables the integration of different types of input in one variable, and monetary data tend to be more readily available than variables on all physical inputs required for analysis.

---

[8] Having said this, there are instances where broader investment measures are used in the literature as inputs for sectoral analysis, such as in a study of telecommunications efficiency by Bartels and Islam (2002).

Obtaining appropriate physical input variables tends to be more feasible when only one sector or subsector is examined. Such is the case in Coelli and Perelman (2000), who examine the efficiency of countries' railways, with the inputs being levels of staff for train and station services, tonnage capacity of freight wagons, the number of seats for coach, and the total length of lines. Utilizing both monetary and physical inputs may be warranted when seeking to include both quantity and quality metrics. E.g., Growitsch, et al. (2009) study utility companies' efficiency in electricity distribution among multiple countries, with total expenditure (operating and capital spending) as a measure of quantity and the duration of outages as a (negative) quality-related input.

In cases where infrastructure outputs considered are flows of services, it may be appropriate to capture inputs as flow measures, as done in Pina and Torres (2006), who consider how three inputs—operating costs, annual bus fuel consumption, and investment spending—translate into annual measures of bus transport service provision. However, when it is of interest to derive the efficiency of creating a stock of infrastructure, then inputs should be measured accordingly. Fritzsche's (2019) road production efficiency analysis is illustrative, deriving the public capital stock by summing annual discounted road spending across years in East German counties. In a cross-country analysis, Baum, et al. (2020) use the general government real capital stock in PPP prices as one of two inputs.

The inputs should also be as consistent as possible with the outputs considered in the analysis. For example, if the intention is to assess the efficiency of investments in digital infrastructure in advanced economies, it would be important to also account for private investment or private capital stock in the inputs. Even in the case of infrastructure mostly undertaken by the public sector, it would be important to decide whether inputs should include government investment spending, or also spending from public-private partnerships and/or state-owned enterprises (SOEs). In practice, data limitations may render inclusion of the latter prohibitive or difficult, particularly in cross-country analyses. This is because limitations in the availability of consistent data on investments by SOEs, and the constraints on data make it difficult to parse out the public component in PPP investments.

D. Factors Affecting Efficiency

In the standard investment efficiency analysis, the input variables are those factors understood to (i) directly bear on the output of concern, such as electricity production or kilometers of road; (ii) are under the control of decisionmakers—that is, they can be increased or decreased; and (iii) are direct "ingredients" to producing the output, that is, would be featured in a classic production function for the output. Typical cases of input variables may include public and/or private investment in the infrastructure of interest, capital stock, or inputs measured as non-financial units, such as labor hours, units of equipment, etc.

On the other hand, variables that fulfill criterion (i) but not necessarily (ii) or (iii), are thought of as environmental variables that affect infrastructure creation but may not be under the managerial control of decision makers and

would not be included in a typical production function. For example, Growitsch, et al. (2009) consider two environmental factors in the delivery of electricity by utility companies: the number of customers per network kilometer (as higher customer density facilitates reaching many customers for a given expenditure level), and country dummies, to capture differences such as regulatory environment in the energy sector. In stochastic frontier analysis, such non-controllable inputs can be captured as a determinant of the mean of the inefficiency term.[9] That is, they are not treated as regular inputs, but rather as factors that determine the efficiency frontier. In the DEA approach, non-discretionary variables that influence infrastructure outcomes can feature in the way that DMUs are benchmarked to determine their efficiency: DMUs are compared with those reference units that have equal or lower values of the non-controllable variables. For example, Agrell, et al. (2005) benchmark electricity distribution concessions against reference DMUs that are in less favorable climate zones and net length (km) (both treated as non-controllable factors in the DEA).

An alternative approach to accounting for environmental variables is to conduct a second stage analysis, in which the investment efficiency scores are regressed on these non-discretionary factors. In a study of investment in street lighting, Lorenzo and Sanchez (2007) derive the efficiency scores for lighting using personnel, streetlamps, and electricity use as inputs, and the area lit, the time lighting is used, and the time lamps remained unrepaired as outputs. Subsequently, they explore the effect of non-controllable factors—like geographical area, population density, hours of daylight, and prevalence of vandalism—on lighting efficiency. In a similar spirit albeit not in regression form,[10] Pina and Torres (2006) assess in a second stage the extent to which contextual variables—including urban and job density, local GDP per capita, and road investment—differ between cities that are efficient and those that are inefficient in delivering urban transport.

More generally, there is a large literature concerned with the drivers of investment efficiency—a question that this paper does not seek to investigate. For example, Baum, et al. (2020) find that that the average country could close more than half of the investment efficiency gap if it adopted the infrastructure governance and public investment management practices of the best performers. The inquiry of the determinants of efficiency may consider not only environmental and institutional factors, but also how efficiency changes with the core inputs themselves. For example, infrastructure efficiency may be nonlinearly related to the stock of public capital. The lumpiness of infrastructure implies that at very low levels of accumulated investment spending, efficiency may be limited. Economies of scale arising from network externalities can also contribute to low efficiency at low capital levels. Once a solid foundation across different infrastructure types is built, small additions to the capital stock may contribute highly to additional infrastructure generation. At very high levels of capital stock, however, diminishing returns may apply, reducing efficiency gains. While these are speculative considerations, future research could empirically examine this question.

---

[9] This is possible because the authors use a truncated-normal SFA model, in which the model value need not be zero but can be positive.

[10] This study uses the Mann-Whitney U test, akin to a nonparametric variant of a t-test.

IV.   Empirical Results on Investment Efficiency

A. Public Infrastructure Output Indexes

In this study, we are primarily concerned with the narrow definition of infrastructure discussed in Section I.A, but also include analysis based on wider definitions for comparison. In particular, our baseline model includes three dominant types of infrastructure following Calderón, et al. (2015): roads, energy, and telecommunications. We develop models with additional variables to examine robustness, namely by including railroads in the set of infrastructure considered (also following Calderón, et al., (2015)), by considering only energy and transport infrastructure (rail and roads), and by examining a model that goes beyond the narrow definition and capturing also infrastructure in the water, health and education sectors, as was done for the physical infrastructure index in Baum, et al. (2020). To summarize, the four models are as follows: Model 1 (baseline): roads, energy, telecommunications; Model 2: roads, energy, telecommunications, railroads; Model 3: roads, energy, railroads; and Model 4: roads, energy, water, health, education.

Variables used for each of the infrastructure types, and their source, are: the length of the road network (in kilometers); electric power consumption (kWh) per 100 people; the number of fixed telephone lines per 100 people; kilometers of rail line route per 100 people; percent of the population using safely managed drinking water facilities; and, as in Baum, et al. (2020) and IMF (2015), health and education infrastructure proxied by the number of secondary school teachers per 100 people and the number of hospital beds per 1,000 people, respectively. The roads variable is obtained from the International Road Federation and the CIA Factbook, the electricity data from the International Energy Agency and the U.S. Energy Information Administration, and all other variables are drawn from the World Bank's World Development Indicators (WDI) database.

There are a few limitations associated with several of these indicators, and their use reflects the absence of data—or the existence for only a small sample of countries—for what may have been preferred variables. For example, as in past IMF studies (e.g., IMF, 2015), we use the education proxy in light of the absence of cross-country data on physical educational capacity (e.g., the number of seats per classroom times the number of classrooms). Another limitation relates to the extent to which the measures used are created by public investment. Cross-country data on each type of infrastructure disaggregated by privately versus publicly created components is virtually non-existent. Given that our input measure is the *public* capital stock, we sought to focus on the types of infrastructure in which the public sector still dominates, generally speaking. It is for this reason that we have not, for example, included mobile telephony or internet systems in our

telecommunications infrastructure measure.[11] Finally, in Section II.A. we discussed output measures that capture the quantity, quality, and utilization dimensions of infrastructure. The variables in our analysis are a combination of quantity indicators (our road, rail, health and education variables) and indicators that in part reflect utilization (electricity consumption, use of safe water, active telephone lines). We include the same quality indicators in Model 4 as in Baum, et al. (2020).

As each variable is measured in a different scale, they need to be standardized before proceeding to any form of aggregation. We consider physical infrastructure as a multidimensional phenomenon, and thus follow the approach of Calderón, et al. (2015) in using principal components analysis to build an index of infrastructure provision based on the individual sectoral indicators. After aggregation, each model implies a different output variable.

These variables are presented (by income group) in Figure 3. They are scaled relative to the average of advanced economies, so that income groups' distributions can be compared against the AE average index of 100. We see that irrespective of model applied, for the most part the stock of infrastructure is larger for EMs than for LIDCs, and is larger still for AEs. The income groups are quite far apart from each other in infrastructure output: In all but one model, the interquartile range of the income groups never even overlap. LIDCs' infrastructure is relatively concentrated within a narrow range, while AEs' values display a wider distribution (especially in the third model that focuses on transport and energy).

Figure 3. Public Infrastructure Output Indexes, by Income Group
(Average of advanced economies = 100)
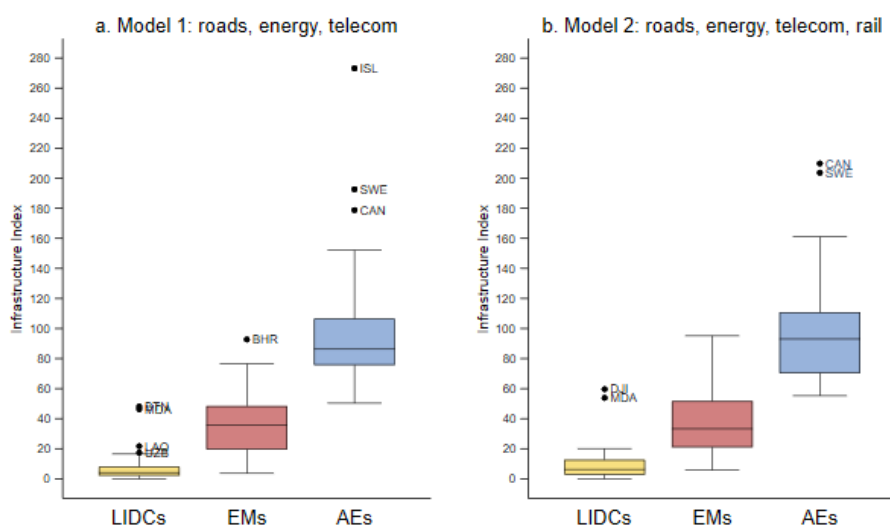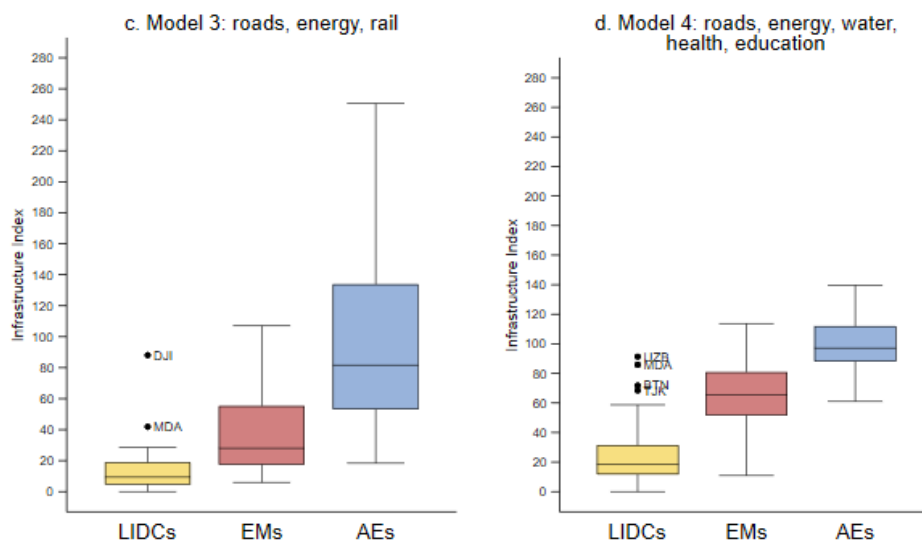
Source: Authors' estimations.

B. Estimated Efficiency Scores

The main input variable is the public capital stock per capita in monetary real dollars, following the approach of IMF (2015) and Baum, et al. (2020). It should be noted that the selection of inputs in monetary form using cross-country data poses challenges. Specifically, the fact that factor prices are higher in richer countries could make AEs seem more inefficient than they are compared to middle- and low-income countries (Gupta and Verhoeven, 2001, Herrera and Pang, 2005).[12] There are various approaches to address this problem. A common solution is to simply estimate frontiers by income group (Gupta and Verhoeven, 2001; Grigoli, 2014; Kapsoli and Teodoru, 2017). Another is to derive the input in monetary units as a share of GDP, which may be more suited in cases where the input is a flow variable (Kyriacou et al. 2019). Yet another approach is to include an input variable that can proxy economic development, such as per capita GDP (see also Baum, et al. 2020). We take this latter route.

Both parametric methods (SFA) as well as nonparametric methods (DEA and bootstrap DEA) were discussed in Section I.B. Before estimating SFA models, it is necessary to verify that this technique is suitable for the specifications proposed in this paper. For this we use the skewness test described in Section 1.B.2. In the baseline model, as well as Model 4 with an infrastructure index akin to Baum, et al. (2020). Figure 4 shows the residuals histograms, and Table 1 the test statistics and associated p-values. As Table 1 shows, the null hypothesis of no skewness is confidently rejected for all four models, providing support in favor of a left-skewed error distribution implied in a SFA model. Finally, because of data issues, not all models include the same

---

[12] The fact that price levels in wealthier countries are higher than in poorer countries is known as the Harrod-Balassa-Samuelson effect. See Obstfeld and Rogoff (1996), Chapter 4.

number of countries. The baseline model has a maximum coverage of 168 countries, Models 2 and 3 have 112 countries—as the railroads variable has more limited coverage—and Model 4 has 153 countries.

Figure 4. Histograms of Residuals from OLS Regressions of Infrastructure on Inputs
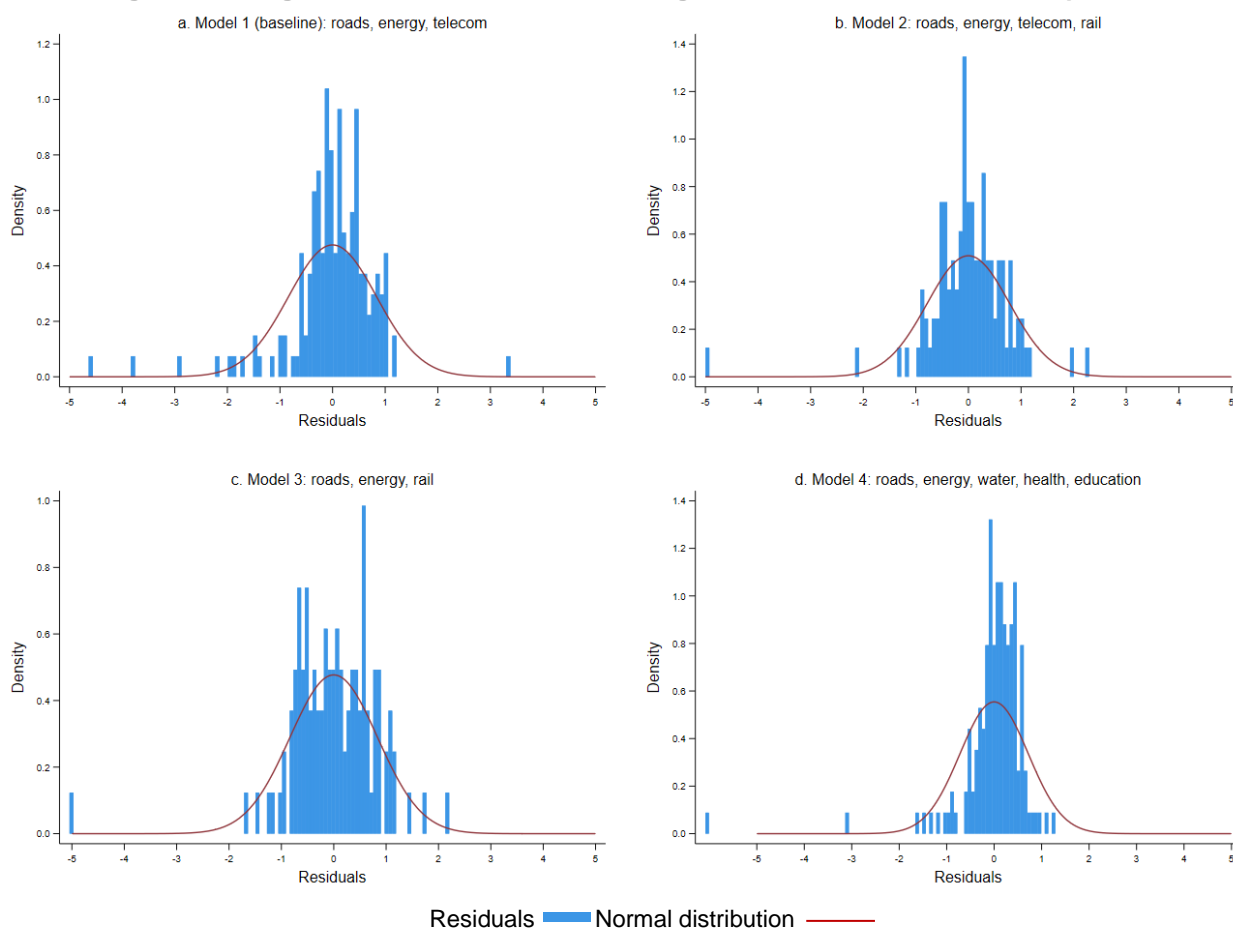
Source: Authors' estimations.
Notes: The histograms depict the distribution of the residuals from an OLS regression of the output on the inputs, using the variables of the efficiency analyses. This is compared against a normal distribution with the same standard error as the residuals.

Table 1. Normality Tests

Model	Skewness	p-value	Test
1. Roads, energy, telecommunications	-1.670***	0.000	84.285
2. Roads, energy, railroads, telecommunications	-2.120***	0.000	81.402
3. Roads, energy, railroads	-1.753***	0.000	66.975
4. Roads, energy, water, health, and education	-4.601***	0.000	187.044

Source: Authors' estimations. *, **, and ***: Statistically significant at the 10, 5, and 1 percent level, respectively.

Figure 5 and Table 2 summarize the main results of the paper. The efficiency scores vary somewhat by model and estimation methodology.[13] The median efficiency score rises with income in nearly all models and estimation methods. However, the median SFA score of EMEs is nearly equal to that of AEs in the baseline model, and higher in the fourth model that is akin to the main model in Baum, et al. (2020).

Figure 5. Public Investment Efficiency Scores

Model 1 (baseline): roads, energy, telecommunications
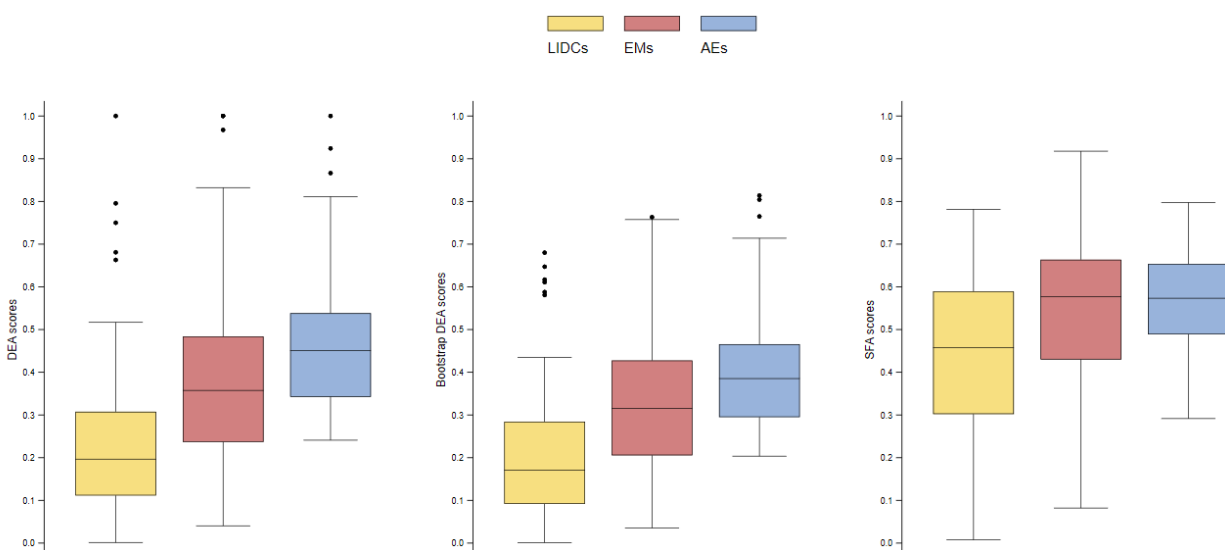
Model 2: roads, energy, telecommunications, railroads
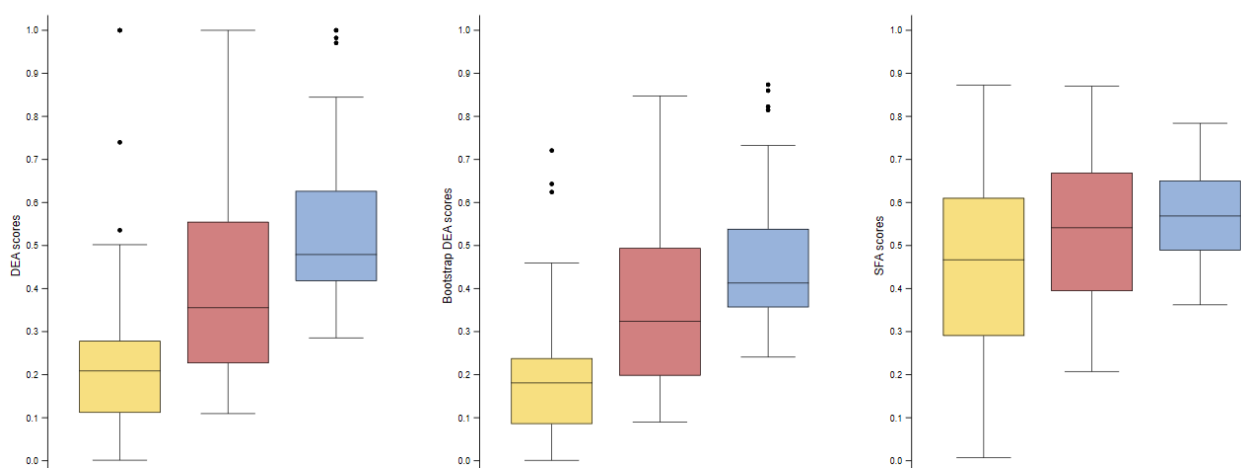
Model 3: roads, energy, railroads
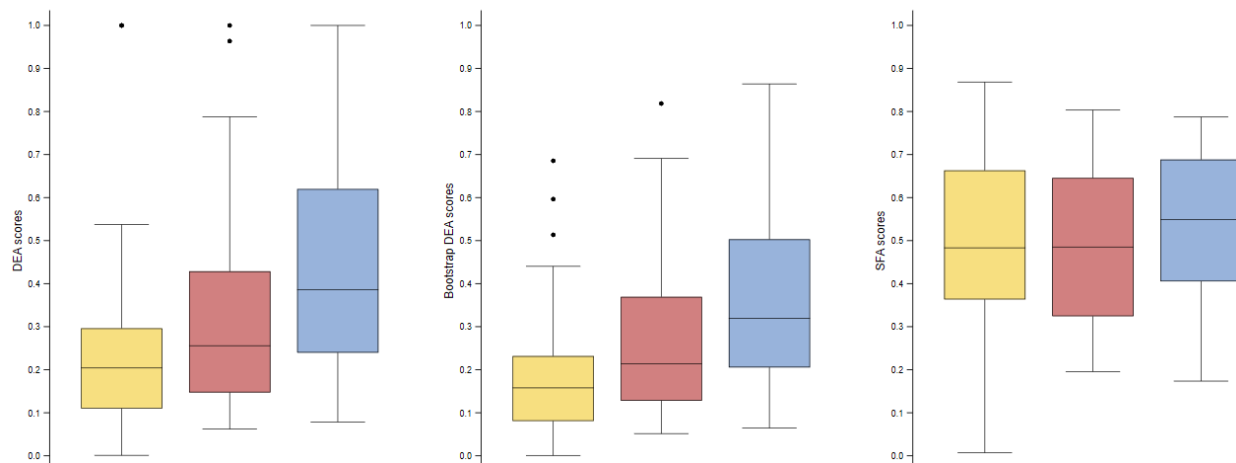
Model 4: roads, energy, water, health, education
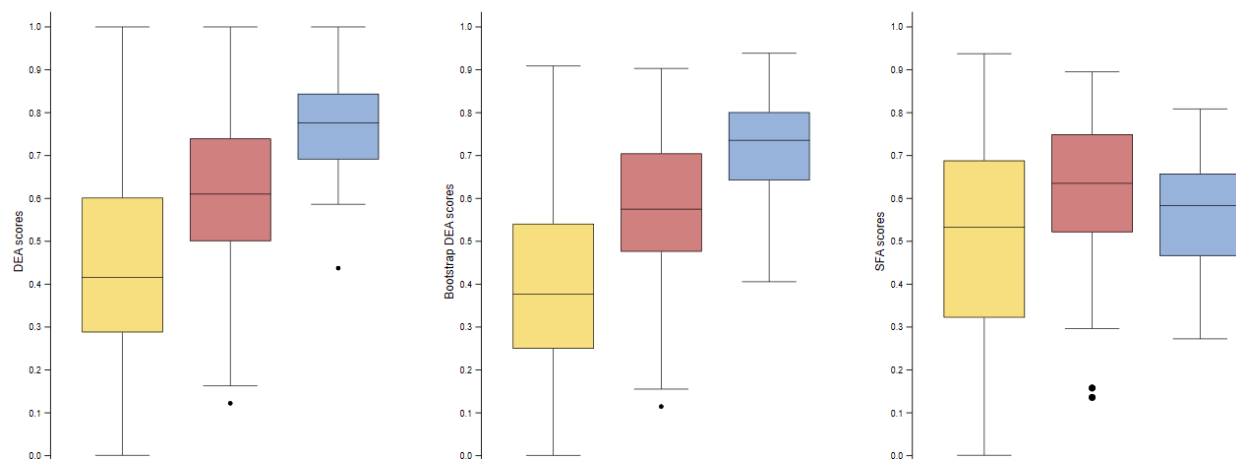
Source: Authors' estimations.

Notes: The upper and lower adjacent values (top and bottom horizontal lines in a box of a boxplot) are defined as follows: Let F be the efficiency score variable for a given model, estimation method, and income group. If F_i is the ith efficiency score when scores are ordered from lowest to highest, and $F_{[25]}$ and $F_{[75]}$ are the 25th and 75th percentiles, respectively, then the upper adjacent value is the Φ_i^U such that $\Phi_i^U \leq U$ and $\Phi_{i+1}^U > U$, where $U = F_{[75]} + 1.5(F_{[75]} - F_{[25]})$. Similarly, the lower adjacent value is the Φ_i^L such that $\Phi_{i-1}^L < L$ and $\Phi_i^L \geq L$, where $L = F_{[25]} - 1.5(F_{[75]} - F_{[25]})$.

The results also suggest that the average global public investment efficiency gap (calculated as 1 minus the average efficiency score) is significant and ranges between 23 and 79 percent, depending on the income group, infrastructure model, and estimation method (Table 2). In our baseline model, the average efficiency gap across methods for AEs, EMs, and LIDCs is 50, 57, and 70 percent, respectively. Averaging all 12 indicators (4 models and 3 methodologies) we find that removing all inefficiencies could increase infrastructure output by more than half overall, and by 45 percent for advanced economies, 54 percent for emerging countries, and 65

percent for low-income countries. It is also worth considering the gains from a more modest, or realistic, target of closing the efficiency gap with each income group's 90th percentile country. Even in this case the additional infrastructure that would be achieved is substantial: Countries could increase infrastructure by 30 percent overall. In AEs, EMs, and LIDCs, this gain would amount to 29, 27, and 35 percent, respectively.

Table 2. Estimated Average Efficiency Gaps, by Income Group

Model	Estimation method	Full efficiency gap				Efficiency gap to the 90th percentile of income group			
		AEs	EMs	LIDCs	Total	AEs	EMs	LIDCs	Total
Model 1	DEA	0.506	0.604	0.750	0.628	0.317	0.292	0.267	0.340
	Bootstrap DEA	0.577	0.657	0.785	0.679	0.290	0.268	0.366	0.296
	SFA	0.426	0.462	0.561	0.485	0.167	0.199	0.303	0.226
	Average	0.503	0.574	0.699	0.597	0.258	0.253	0.312	0.287
Model 2	DEA	0.436	0.584	0.715	0.576	0.412	0.415	0.455	0.420
	Bootstrap DEA	0.516	0.638	0.770	0.637	0.334	0.307	0.394	0.306
	SFA	0.419	0.461	0.533	0.467	0.176	0.209	0.291	0.215
	Average	0.457	0.561	0.672	0.560	0.308	0.310	0.380	0.314
Model 3	DEA	0.531	0.675	0.733	0.651	0.508	0.380	0.271	0.415
	Bootstrap DEA	0.608	0.725	0.792	0.710	0.409	0.318	0.305	0.323
	SFA	0.456	0.504	0.514	0.493	0.230	0.241	0.256	0.235
	Average	0.532	0.635	0.680	0.618	0.382	0.313	0.277	0.324
Model 4	DEA	0.226	0.368	0.533	0.389	0.226	0.245	0.533	0.301
	Bootstrap DEA	0.273	0.409	0.588	0.436	0.194	0.235	0.410	0.268
	SFA	0.432	0.368	0.493	0.418	0.183	0.190	0.349	0.228
	Average	0.310	0.382	0.538	0.414	0.201	0.223	0.431	0.266
Average across models	*DEA*	0.425	0.558	0.683	0.561	0.366	0.333	0.381	0.369
	Bootstrap DEA	0.493	0.607	0.734	0.616	0.307	0.282	0.369	0.298
	SFA	0.433	0.449	0.525	0.466	0.189	0.210	0.300	0.226
	Overall average	**0.450**	**0.538**	**0.647**	**0.547**	**0.287**	**0.275**	**0.350**	**0.298**

Source: Authors' estimations.

While efficiency scores vary by output model and efficiency estimation method in aggregates such as means and medians, we examine to what extent this also applies when considering the country estimates. Specifically, we assess whether the results across approaches closely correlated at the country level, by producing the Spearman rank-correlations between the different efficiency methodologies for any given model. As Table 3 shows, all 12 correlation coefficients (three estimation methodologies for each of four models) are highly statistically significant and range from 0.73 to 0.99. Individual country efficiency scores are closely correlated across methodologies used to estimate the scores which suggests that while numerical values for efficiency levels vary across estimation methods, country rankings remain stable.

Table 3. Correlations between Methodologies

Model	Methodology	DEA	Bootstrap DEA
Model 1: roads, energy, telecom	Bootstrap DEA	0.979*** (0.000)	
	SFA	0.913*** (0.000)	0.886*** (0.000)
Model 2: roads, energy, telecom, rail	Bootstrap DEA	0.994*** (0.000)	
	SFA	0.917*** (0.000)	0.912*** (0.000)
Model 3: roads, energy, rail	Bootstrap DEA	0.995*** (0.000)	
	SFA	0.939*** (0.000)	0.931*** (0.000)
Model 4: roads, energy, water, health, education	Bootstrap DEA	0.993*** (0.000)	
	SFA	0.738*** (0.000)	0.731*** (0.000)

Source: Authors' estimations.
Note: p-values in parentheses. *, **, and ***: Statistically significant at the 10, 5, and 1 percent level, respectively

The next robustness check examines the relationship among the different model specifications. In this case we choose one method at a time and estimate the correlation among different model specifications and test its significance. Again, all correlation coefficients are strongly statistically significant and range from 0.52 to 0.95 (Table 4). Together, these results suggest that countries generally preserve their efficiency ranking regardless of the outcome measure used.

Table 4. Correlations for Different Models and Estimation Methods

| Methodology	Model	Model 1	Model 2	Model 3
DEA	Model 2	0.881***		
		(0.000)		
	Model 3	0.777***	0.948***	
		(0.000)	(0.000)	
	Model 4	0.681***	0.751***	0.686***
		(0.000)	(0.000)	(0.000)
Bootstrap DEA	Model 2	0.872***		
		(0.000)		
	Model 3	0.763***	0.949***	
		(0.000)	(0.000)	
	Model 4	0.695***	0.734***	0.670***
		(0.000)	(0.000)	(0.000)
SFA	Model 2	0.882***		
		(0.000)		
	Model 3	0.753***	0.943***	
		(0.000)	(0.000)	
	Model 4	0.614***	0.594***	0.522***
		(0.000)	(0.000)	(0.000)

Source: Authors' estimations.
Note: p-values in parentheses. *, **, and ***: Statistically significant at the 10, 5, and 1 percent level, respectively

C. Efficiency Scores with Disaggregated Inputs

The assessment of public investment efficiency poses aggregation issues not only on the output side, as discussed Section 2.B, but also on the input side. Consistent public investment statistics—for a reasonably large number of countries—are only available at an aggregate level. This makes it challenging to find output measures that reflect the various targets outlined in an investment budget (roads, hospitals, schools, etc.). Efficiency estimations presented in the main results of this paper use the public capital stock— public investment cumulated over time—as input, which requires the use of an aggregated form based on different outputs.

This section further explores the impact on efficiency scores of using public investment disaggregated by economic sector. To this end, we take advantage of the INFRALATAM database on public investment (including general government and SOEs), a joint effort by the Economic Commission for Latin America and the Caribbean (ECLALC), the Inter-American Development Bank (IDB) and the Development Bank of Latin America (CAF). The database, published October 2022, includes 22 Latin-American countries from 2008 through 2021 and covers various infrastructure sectors. Figure 6 shows that the roads sector constitutes by far the largest portion of infrastructure investment, followed by electricity. It is also apparent that there has been a

steady decline in public investment in Latin American countries since 2016, both overall and across most sectors, with this decline accelerating in the first pandemic year (2020).

Figure 6. Public Investment by Sector in Latin America
(Percent of GDP)
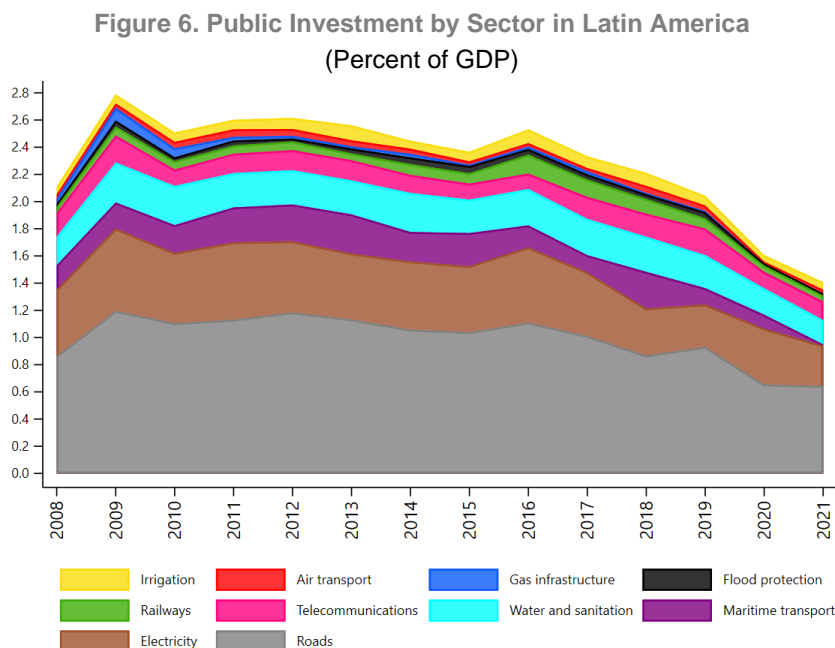
Source: Authors' estimations based on INFRALATAM dataset.

Using this database on Latin American countries, we can compute efficiency scores at the sectoral level for those sectors included in in the first three of our four models (since the INFRALATAM data does not have investment data for the health and education sectors) and compare the results with those produced using the original methodology described in Section 3.B. To be clear, this section does not attempt to perform a comprehensive study of the efficiency of public investment in LAC countries. It only seeks to show the impact on the efficiency scores resulting from the use of disaggregated inputs by replicating the models with aggregated and disaggregated inputs.[14]

We estimate efficiency scores for road, rail, telecommunications, and electricity for Latin American countries. In each case the output is the corresponding sub-component of the index described in Section 3.A, while inputs are the sectoral public capital stock and per capita GDP. In other words, for each country and method we derive four efficiency scores: one for each of the abovementioned infrastructure types. The final efficiency score for each country is constructed as the average of each sectoral score that Models 1, 2, and 3 include. We only present results of the two non-parametric methods, as the small number of observations precludes the use of parametric methodologies such as SFA.

---

[14] It should be stated as a caveat that while our main results include capital stock only from general government, INFRALATAM additionally includes SOE-created capital stock. Thus, the larger amount of input in the latter data implies that ##.

The left panel of Table 5 conducts this comparison based on DEA, and the right-hand panel based on the bootstrapped estimation. Presented are the efficiency scores averaged over Latin American countries for each model. In the case of the DEA estimation, while scores using the paper's main approach of the aggregated infrastructure output index and aggregated public capital stock are higher than sector-specific efficiency scores in two of the three models, these differences are not statistically significant. However, aggregation-based scores appear to be systematically higher than sector-specific scores when bootstrapping is applied. This suggests that our estimated efficiency gaps in Table 5 may be larger if we had the data to assess efficiency scores with more disaggregated data, and thus the gaps should potentially be treated as a lower-bound.

Table 5. Average Efficiency Scores for Latin American Countries

	DEA			Boostrap DEA		
	Aggregation-based scores	Sector-specific scores	t-test (p-value)	Aggregation-based scores	Sector-specific scores	t-test (p-value)
Model 1	0.626	0.551	1.208 (0.246)	0.571	0.436	3.558*** (0.002)
Model 2	0.564	0.621	-0.454 (0.666)	0.644	0.493	2.183* (0.072)
Model 3	0.631	0.519	0.801 (0.454)	0.611	0.378	3.281** (0.017)

Source: Authors' estimations

V. Conclusion

Public investment in infrastructure can be an integral part of countries' longer-term economic recovery from the pandemic and its economic ripple-effects in the years ahead. However, many countries, especially low-income economies, will face barriers to expanding infrastructure through substantially increasing investments when access to capital markets remains constrained, and thus increasing investment efficiency will need to be a critical avenue to building infrastructure. This paper draws out ways to assess investment efficiency. In particular, it takes a focused lens to the various analytical considerations when selecting inputs and outputs and integrating them in investment efficiency analysis using various methods and carries out empirical analysis to further explore these analytical options.

We use stochastic frontier analysis, data envelopment analysis, and bootstrapped DEA to estimate efficiency scores for a wide range of countries using four different scopes of infrastructure, employing mainly metrics of infrastructure quantity and utilization. Our primary analysis generates a composite measure of infrastructure output, given that investment inputs are globally only available in aggregated form. We also provide different robustness checks to ensure that the results are meaningful. However, the existence of a disaggregated dataset for Latin America enables us to compare the findings for this region based on aggregated inputs and outputs with those separately estimated by sector of infrastructure and type of investment, and the differences

in results suggest that the efficiency gaps from the main results using aggregated inputs and outputs should be seen as a lower-bound.

We found a sizeable efficiency gap, averaging all 12 indicators (4 models and 3 methodologies) we found that removing all inefficiencies could increase infrastructure output by 55 percent overall, and by more in low-income developing countries (65 percent) than in emerging market economies (54 percent), and in the latter by more than in advanced economies (45 percent). Even considering a more realistic goal of closing not the full efficiency gap but the gap to the 90th percentile within each income group, a still sizeable 30 percent increase in infrastructure output could be gained.

While this paper considers different substantive and analytical implications of the ways that inputs and outputs are used in efficiency analysis, only some of these options could be put to the test in the empirical analysis. Future research could expand the empirical applications to more of the options that were discussed conceptually in this paper, if data were to become available to do so. Future work could also contribute to the literature on the drivers of investment inefficiencies and possible ways to overcome these weaknesses. Finally, as public infrastructure is usually one of the most important determinants of growth, the estimates in this paper could be used also to add to the body of work identifying the relationship between the efficiency of public investment and economic growth.

References

Agrell, Per J., Peter Bogetoft, and Jørgen Tind (2005) "DEA and dynamic yardstick competition in Scandinavian electricity distribution," *Journal of Productivity Analysis*, 23: 173–201.

Albino-War, Maria, Svetlana Cerovic, Francesco Grigoli, Juan Carlos Flores, Javier Kapsoli, Haonan Qu, Yahia Said, Bahrom Shukurov, Martin Sommer, and SeokHyun Yoon, 2014, "Making the most of public investment in MENA and CCA oil-exporting countries," IMF Staff Discussion Note 14/10.

Ali, Md Kamar, and K. K. Klein (2014) "Water use efficiency and productivity of the irrigation districts in southern Alberta" *Water Resource Management* 28: 2751–2766.

Bartels, Robert and Towhidul Islam (2002) "Supply restricted telecommunications markets: The effect of technical efficiency on waiting times" *Journal of Productivity Analysis*, 18: 161-169.

Batini, Nicoletta, Mario di Serio, Matteo Fragetta, Giovanni Melina, and Anthony Waldron. 2021. "Building back better: How big are green spending multipliers?" IMF Working Paper No. WP/21/87.

Baum, Anja, Tewodaj Mogues, and Geneviève Verdier (2020) "Getting the Most from Public Investment," Ch. 3 (pp. 30 – 49), In: Schwartz, Gerd, Manal Fouad, Torben Hansen, and Geneviève Verdier, *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*. Washington DC: International Monetary Fund.

Bogetoft, Peter, and Lars Otto (2011) "Benchmarking with DEA, SFA, and R," Springer.

Brasili, Andrea, Atanas Kolev, Debora Revoltella, and Jochen Schanz (2022) "Public Investment in the Pandemic—Europe at a Glance," In: Floriana Cerniglia, Francesco Saraceno, and Andrew Watt (eds.), *The Great Reset: 2021 European Public Investment Outlook*, OpenBook Publishers.

Calderón, César, Enrique Moral-Benito, and Luis Servén (2015) "Is infrastructure capital productive? A dynamic heterogeneous approach," *Journal of Applied Econometrics*, 30(2): 177–198.

Canning, David (1998) "A Database of World Stocks of Infrastructure, 1950-95," *World Bank Economic Review*, 12(3): 529-547.

Cavalieri, Marina, Calogero Guccio, and Ilde Rizzo (2017) "On the role of environmental corruption in healthcare infrastructures: An empirical assessment for Italy using DEA with truncated regression approach," Health Policy, 121: 515-524.

Coelli, Timothy, and Sergio Perelman (1999) "A comparison of parametric and non-parametric distance functions: With application to European railways," *European Journal of Operational Research* 117: 326-339.

Coelli, Timothy, and Sergio Perelman (2000) "Technical efficiency of European railways: A distance function approach," *Applied Economics*, 32(15): 1967-1976.

Coelli, Timothy, Dodla Sai Prasada Rao, Christopher J. O'Donnell, and George Edward Battese (2005) *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

D'Agostino, Ralph, Albert Belanger, and Ralph D'Agostino Jr. (1990) "A suggestion for using powerful and informative tests of normality," *American Statistician*, 44(4): 316–321.

Estache, Antonio and Marianela González (2002) "Efficiency gains from port reform and the potential for yardstick competition: Lessons from Mexico," *World Development*, 30(4): 545-560.

Farrell, Michael J. (1957) "The measurement of productive efficiency," *Journal of the Royal Statistical Society*, 120(3): 253–290.

Fritzsche, Carolin (2019) "Analyzing the efficiency of county road production: evidence from Eastern German counties," *German Economic Review* 20(4): e415–e435.

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

Grigoli, Francesco (2014) "A hybrid approach to estimating the efficiency of public spending on education in emerging and developing economies," IMF Working Paper N 14/19.

Grigoli, Francesco and Javier Kapsoli (2018) "Waste not, want not: The efficiency of health expenditure in emerging and developing economies," *Review of Development Economics* 22(1): 384-403.

Growitsch, Christian, Tooraj Jamasb and Michael Pollitt (2009) "Quality of service, efficiency and scale in network industries: an analysis of European electricity distribution," *Applied Economics*, 41(20): 2555-2570.

Gupta, Sanjeev and Marijn Verhoeven *(*2001) "The efficiency of government expenditure: experiences from Africa" *Journal of Policy Modeling*, 23(4): 433-467.

Hepburn, Cameron, Brian O Callaghan, Nicholas Stern, and Dimitri Zenghelis (2020) "Will COVID-19 fiscal recovery packages accelerate or retard progress on climate change?" *Oxford Review of Economic Policy,* 36 (20): 1–48.

Herrera, Santiago, and Gaobo Pang (2005) "Efficiency of public spending in developing countries: An efficiency frontier approach," World Bank Policy Research Working Paper N 3645.

International Monetary Fund (2015) "Making public investment more efficient," IMF Board Paper.

_____ (2020) "Fiscal Monitor: Policies for the Recovery," IMF, Washington DC, October 2020.

_____ (2022a) "Fiscal Monitor: Helping People Bounce Back," IMF, Washington DC, October 2022.

_____ (2022b) "Saudi Arabia—Selected Issues," IMF, Washington DC, August 2022.

_____ (2022c) "World Economic Outlook: Countering the Cost-of-Living Crisis," IMF, Washington DC, October 2022.

Kapsoli, Javier, and Iulia Ruxandra Teodoru (2017), "Benchmarking social spending using efficiency frontiers," Working Paper WP/17/197. International Monetary Fund.

Kalb, Alexander (2014) "What determines local governments' cost-efficiency? The case of road maintenance" *Regional Studies*, 48(9): 1483–1498.

Kneip, Alois, Byeong Park, and Léopold Simar (1998) "A note on the convergence of nonparametric DEA estimators for production efficiency scores." *Econometric Theory*, Vol. 14, N 6, pp. 783-93.

Kyriacou, Andreas P., Leonel Muinelo-Gallo, and Oriol Roca-Sagalés (2019) "The efficiency of transport infrastructure investment and the role of government quality: An empirical analysis," *Transport Policy*, 74: 93–102.

Lozano, Sebastián and Ester Gutiérrez (2011) "Efficiency analysis and target setting of Spanish airports," *Networks and Spatial Economics*, 11:139-157.

Lv, Kangjuan, Anyu Yu, and Yiwen Bian (2017) "Regional energy efficiency and its determinants in China

during 2001–2010: a slacks-based measure and spatial econometric analysis," *Journal of Productivity Analysis,* 47:65–81.

Lorenzo, José Manuel Prado and Isabel María García Sánchez (2007) "Efficiency evaluation in municipal services: an application to the street lighting service in Spain," *Journal of Productivity Analysis* 27:149-162.

Marcelo, Darwin, Cledan Mandri-Perott, Schuyler House, Jordan Schwartz (2016) "Prioritizing infrastructure investment: A framework for government decision-making," Policy Research Working Paper No. 7674, World Bank, Washington DC.

Martín, Juan Carlos, Concepción Román, and Augusto Voltes-Dorta (2009) "A stochastic frontier analysis to estimate the relative efficiency of Spanish airports" *Journal of Productivity Analysis* 31: 163–176.

Moszoro, Mariano and Mauricio Soto (2022) "Road quality and mean speed score," IMF Working Paper WP/22/95, Washington DC.

Murillo-Zamorano, Luis (2004) "Economic efficiency and frontier techniques," *Journal of Economic Surveys*, 18(1): 33-77.

Obstfeld, Maurice and Kenneth Rogoff (1996) *Foundations of International Macroeconomics*, Cambridge, MA: MIT Press.

Pina, Vincente, and Lourdes Torres (2006) "Public-private efficiency in the delivery of services of general economic interest: The case of urban transport" *Local Government Studies*, 32(2): 177 – 198.

Rayp, Glenn and Nicolas Van De Sijpe (2007) "Measuring and explaining government efficiency in developing countries," *Journal of Development Studies*, 43(2): 360-381.

Simar, Léopold, and Paul W. Wilson (1998) "Sensitivity analysis of the efficiency scores: How to bootstrap in nonparametric frontier models," *Management Science*, 44(1): 49–61.

_____ (2000) "A general methodology for bootstrapping in non-parametric frontier models," *Journal of Applied Statistics*, 27(6): 779–802.

Tanzi, V., and H. Davoodi (2002) "Corruption, public investment, and growth." In Governance, Corruption, and Economic Performance*,* edited by G. Abed and S. Gupta, 280–299. Washington: International Monetary Fund.

Wanke, Peter F. (2013) "Physical infrastructure and shipment consolidation efficiency drivers in Brazilian ports: A two-stage network-DEA approach," *Transport Policy* 29: 145-153.

Wheat, Phill, Alexander D. Stead, William H. Greene (2019) "Robust stochastic frontier analysis: a Student's t-half normal model with application to highway maintenance costs in England," *Journal of Productivity Analysis*, 51: 21–38.

Xiao, Yuan, David Amaglobeli, and Riki Matsumoto (2021) "IMF Investment and Capital Stock Dataset (ICSD) 2021: Manual & FAQ - Estimating Public, Private, And PPP Capital Stocks," (mimeo).

PUBLICATIONS

Benchmarking Infrastructure Using Public Investment Efficiency Frontiers

Working Paper No.WP/2023/101