

CCRP Working Paper Series

***Government Expenditures on Education, Health and Infrastructure: A
Naïve Look at Levels, Outcomes, and Efficiency***

Antonio Estache, Marianela Gonzalez & Lourdes Trujillo

Centre for Competition and Regulatory Policy (CCRP), The City University
The World Bank
ECARES, Universite Libre de Bruxelles
DAEA, Universidad de Las Palmas de Gran Canaria

CCRP Working Paper No 6

Government Expenditures on Education, Health, and Infrastructure: A Naïve Look at Levels, Outcomes, and Efficiency

Antonio Estache

The World Bank and
ECARES, Université Libre de Bruxelles

Marianela González

DAEA, Universidad de Las Palmas de Gran Canaria

Lourdes Trujillo*

DAEA, Universidad de Las Palmas de Gran Canaria and
CCRP, City University of London

Abstract

All interested parties seem to agree that it's important to be able to monitor public-sector performance at the sectoral level, but most current work based on multicountries databases doesn't lend itself to country-specific conclusions. This is due to a large extent to major data limitations both on sectoral expenditures and on sectoral outcomes. This paper discusses the related issues and shows what we can do with the current data in spite of the drastic limitations. The main conclusion of the paper is that any efforts to assess country specific performances in relative terms are likely to be difficult in view of the data problems. A rough sense of performance across sectors can be estimated for groups of countries, allowing some modest benchmarking exercises. These estimates show that low income countries generally lag higher income countries significantly. However, efficiency has improved during the 1990s in energy and education but has not improved significantly in transport.

*Correspondent author: Lourdes Trujillo CCRP, City University. Northampton Square, EC1 V0HB, London, email: ltrujillo@daea.ulpgc.es. We are grateful to Tim Coelli, Ariel Fiszbein, Ana Goicoechea, Pablo Gottfret, Santiago Herrera, Kai Kayzer, Eduardo Ley, Harry Patrinos, Sergio Perelman, Pierre Pestiau, Anand Rajaram, David Saal, Philippe Vanden Eeckaut, and participants at a seminar sponsored by the U.K. Department for International Development for useful discussions and reactions at various stages of the preparation of this paper. These results represent our views and should not be attributed to the institutions with which we are affiliated. Any mistakes of fact or interpretation in this paper are ours alone.

Introduction

Fiscal adjustment programs always press governments to increase productivity. That pressure is generally summed up in two questions:

- *How much less could the government spend to achieve current sector-specific outcomes—namely, targets for water- or energy-access rates, infant mortality rates, or academic test scores?*
- *How much better could these outcomes be, given the current levels of sector-specific public expenditures?*

These two questions are probably central to any diagnostic effort to determine the quality of public sector expenditures.¹ Unfortunately, no country, rich or poor, seems to be able to provide a precise quantitative answer to these basic questions at the broad level of aggregation used for sectorwide assessments of needs and performance.²

Until recently, accountability for the match between sector specific expenditures and the associated education, health and infrastructure service deliveries had not appeared to be a major concern for finance ministers or watchdogs of public sector performance, as long as total expenditure levels were assessed to be sustainable from a macroeconomic perspective.³ This is obvious in the academic literature. Since the seminal paper by Devarajan, Swaroop, and Zou (1996), few empirical papers have been published on the sectoral composition of public expenditure in developing countries—in spite of a growing interest among political scientists to document the impact of globalization on the behavior of governments.⁴

Things are changing. There is indeed a growing concern for accountability, which in turn is stimulating academic and policy interest in detailed budgeting processes and hence on

¹ See Pradhan (1996), for example.

² The equivalent questions are the bread and butter of a project-level cost-benefit analysis. But even then, the quantitative assessments in most developing countries often fall short of expectations in most of the sectors covered here.

³ This lack of accountability for the level and effectiveness of service delivery also limits the ability of most private companies that are charged with delivering education, health, and infrastructure services (or any other good or service) to raise significant funds from the financial markets.

⁴ There is, of course, a voluminous literature on sector-specific public expenditure—particularly in health and education and in some dimensions of infrastructure, mostly telecoms. Although less vast, the literature on the distribution of expenditures between recurrent and capital expenditures is also illuminating. Among the few papers discussing allocation across sectors, the emphasis is on the distinction between productive and social sectors, focusing on some proxy for productive expenditure ignoring the relevance of the heterogeneity of the productive sector.

sectoral resource-allocation decisions.⁵ In the last three to five years, various governments within the Organization for Economic Co-operation and Development (OECD)—particularly the Australian, British, Dutch, Finnish, and Norwegian governments—have made significant progress toward linking expenditures and outcomes as part of their efforts to increase public-sector accountability to taxpayers.⁶ Most of these countries have produced guidelines for their civil servants with varying amounts of sectoral detail but with enough incentives to begin measuring their activities as any business unit would. The public sector is importing long-tested cost-accounting rules from the private sector—at least in some part of the activities of the public sector such as regulated industries.

But progress is slow. Those leading the charge for change recognize quite explicitly (in widely available public documents) the limitations of their efforts.⁷ Although it would be overly simplistic, one could nevertheless argue that the two barriers to the implementation of new guidelines are posed, first, by governments that have never properly measured sectoral expenditures, and, second, by governments that have never measured sectoral outcomes well either! The problem is *this* basic, even in countries trying to produce quantitative assessments of public-sector performance.

This is not to say there are no data. Every government has a budget. In fact, most governments also have “realized” budgets, reporting *ex post* their actual expenditures. But these budgets are usually intended only to ensure the macroeconomic sustainability of total government expenditures. They are not really designed to track down specific sectoral expenditures—or many other resources-allocation decisions for that matter, which is to say, the distribution between capital and recurrent expenditures. That is a problem for anyone trying to assess the cost-effectiveness of sectoral policies.

The problem is not, however, only with expenditure or cost data. Measuring outcomes is also a serious problem. Once more, it is not that countries do not generate data on outcomes; it is just that the data are not really that helpful to get a sense of the effectiveness of sectoral resource allocation. The coverage of sectoral outcomes is always systematically partial both in terms of indicators of performance or results (e.g., the diversity of transport, health, or education services vs. infrastructures) and in terms of the dimensions of performance (e.g., performance should not only be measured in terms of quantity but also in terms of quality and affordability). For instance, it is surprising to see that about 20 years

⁵ To be fair, this is not the first attempt at change. In the United States, for instance, the Government Performance and Results Act of 1993 featured many aspects of legislation enacted over the past five years in a number of OECD countries. But the Act never really delivered much in terms of measurement or accountability.

⁶ For an overview of OECD initiatives, see OECD, Country Facts Sheets, GOV/PGC (2006), April 6, JT03207081. The *Atkinson Review* may be the best documented such initiative, with its own website http://www.statistics.gov.uk/about/data/methodology/specific/PublicSector/Atkinson/final_report.asp

⁷ These government websites are the strongest evidence available to the international community.

after the so-called Water Decade was announced, no one is reporting the average quality of water delivered or the number of hours per day that water is available in given countries, or even in a representative sample of cities.

Despite the data limitations, international benchmarking of sectorwide performances based on partial expenditure and outcome indicators is increasingly popular in the literature as a way to increase accountability. Most of the time, such benchmarking takes the form of ranking and country-specific performance assessments at the operator level—utilities, hospitals, schools, and so forth. The approach is quite common in the education and health fields.⁸ It is less common in the infrastructure sector, except perhaps for electricity, rail, and telecoms, where the availability of data has stimulated a wide range of publications.⁹

The popularity of international benchmarking is, however, not universally supported, in particular when it is used to assess performance at the sectoral level. The critics are indeed strong.¹⁰ Some have very valid conceptual reservations. But most are concerned that the international benchmarking of sector-specific performance leads to the compilation of country-level data that can be used and abused when policies are formed.

In this paper we try to generate some efficiency benchmarks—although at a higher level of aggregation than earlier efforts (which also sought to minimize the risks of misinterpretation identified by the critics).¹¹ Based on the partial information available, we offer a rough answer to the second of the two core questions posed at the outset—how much more could governments achieve at their current level of expenditure?¹² In the process, we document and discuss in some detail the data problems facing anyone seeking to make broad-brush comparisons across country groups over time in each sector. More specifically, our paper makes three main contributions:

- It provides an overview of sectoral public expenditures and outcomes by sector.
- It gives a sense of the current levels of expenditure and of their efficiency for broad income groups of countries—low, lower-middle, upper-middle—in both developing countries and developed countries.
- And it identifies areas in which government could improve its monitoring of public-sector performance in education, health, and infrastructure.

⁸ This would call for a long bibliography. A good substitute is to go through the Asian Development Bank-ADB- (2006), which provides a very nice overview of both efficiency measures on health and education and of the harsh criticism of these measures.

⁹ See Coelli et al. (2003).

¹⁰ See Ravallion (2005) for a good and incisive methodological critique of this literature.

¹¹ See ADB (2006) and Coelli et al., (2003) for an overview of the previous research

¹² A precise answer is almost impossible to supply, as explained by Ravallion (2005) in his very tough essay on any effort to measure the efficiency of public expenditures.

Our approach is somewhat mechanical. Indeed, our main concern was to be as systematic as possible in collecting and validating all the cross-country data on the level and efficiency of public expenditure in the three sectors over 10–15 years—a relatively long period of time. Because the data on expenditure and outcomes are not generated from the same sources and because they cover different samples of countries, one of the main and most time-consuming challenges was to define expenditure and outcome indicators in a way that would allow us to maximize the number of countries and the number of years represented in the data base. We discuss this in detail later, but it is essential to keep in mind that the binding constraint on the coverage was the availability of data on public expenditure and *matching data* on outcomes. A country is included in the data base only if those data are available. The meagerness of this basic information is not the only issue. For a number of reasons discussed later, the exercise required some assumptions, in particular when measuring efficiency. The reassuring fact is that once all the data were processed, an interesting story emerged, even if it is a somewhat naïve story in view of the limitations just described.

The paper is organized as follows. Section 2 provides an overview of the data available on expenditure and outcomes. Section 3 reports the main comparative facts on expenditure levels and outcomes that could be generated from a basic statistical analysis of the data. Section 4 presents a rough assessment of the evolution of sector-specific efficiency levels for the four main groups of countries of concern here. The main objective of this section is to get as close as possible to benchmarking cross-country performances given the scanty data. Section 5 discusses the many limitations of the analysis while pointed to some of the potential uses of the available information. Section 6 concludes with some suggestions for further work and discusses specific information gaps the international community will have to address in its standardized international datasets.

How much data are there, and what can you do with them?

This section discusses the information we used to conduct large-scale cross-country and cross-sectoral performance comparisons of public expenditures. We conclude that these cross-country assessments cannot match the quality of the analysis that could be achieved from country- or sector-specific studies conducted at the level of the business unit—central or local administration, school system, hospital, or water or electricity distribution company. Why? Because the data are simply not good enough to generate much more than “big picture” comparative assessments. The country coverage varies significantly across sectors and across time. The data are, however, good enough to give one a good sense of the evolution of the sectoral allocation of resources across broad, income-based country groupings over time.

The basic data issues can be summarized as follows. In an ideal world, we would have had access to long-term and detailed public and private expenditure data on each subsector,

with matching data on all the relevant outcomes as well as all the relevant institutional variables, such as the degree of decentralization, the financing options, and other characteristics of the initial conditions with respect to the outcomes in each subsector. Those data do not exist.

In the real world, the immensity of data problems is overwhelming. In fact, it proved to be impossible to produce a simultaneous assessment of multiple categories of public expenditures over a long period of time. First, such an assessment would require data on expenditure that would be consistent across sectors. The IMF Government Finance Statistics database (GFS) proved to be the only data source that achieved this consistency across sectors. GFS is the main data source for most empirical cross-country studies on government expenditures¹³; it is the only database offering comparable data on public expenditure for all countries in the world, including developing countries.

This does not mean that the data are good. For instance, the GFS ignores a wide range of sectoral expenditures by public enterprises and other alternative public providers. It tends to have much better data on the central government than on subnational governments. For most countries, the data reported in this widely quoted publication tend to underestimate actual expenditure levels. In addition to the gaps in coverage, the GFS does not credit specific sectors for subsidies debited from general government expenditures to support sector-specific activities managed from the sectoral budget. In sum, any study relying on that data will be biased. Yet everyone uses the data, simply because they are the only set available.¹⁴

Second, many of the data are collected sporadically. The time coverage of expenditures and outcomes differs for many countries; in other words, inputs are available for some years and outputs for others. There are hardly any relevant data on sector-specific institutional characteristics. We thus had to pick between (i) maximizing the number of variables used, irrespective of time-coverage compatibility across countries and across subsectors, so as to minimize the loss of sample size, and (ii) getting the cleanest possible dataset, irrespective of the impact on the sample size. We adopted a hybrid solution. We gave up on presenting a number of outcomes available for only a few countries or years in favor of maintaining a reasonable sample size while accepting that the data quality and comparability would not be free from criticism.

Third, because we were unable to generate useful country-specific data given the limited data coverage for each country, we focused on generating *enough* country coverage in four country-income groups: low, lower-middle, upper-middle, and higher. These

¹³ The OECD and European Union have their own databases, which are widely used in the empirical literature but are of limited use here because there are no comparable data for developing countries.

¹⁴ Note that the World Health Organization has started to generate its own expenditure series on health. It is more detailed and covers almost the whole world, but it does so only since 1998]

classifications turned out to be broad enough not to be influenced by any remaining country-specific problems. The breadth also gives a sense of the gap that can exist among countries at different stages of development. The availability of time series on these broad groups was essential in the choice of countries and indicators because it also allowed us to track the evolution of this gap among country groups.

With these limitations in mind, we were able to whittle the challenges down to the sector-specific issues. Relying on the IMF data for health and education is not a major problem because they explicitly identify these two categories. The problem is more serious for infrastructure, an expenditure category that does not exist in the GFS, which uses instead the categories of transport and “fuel and energy.” The sum of these two categories would be the closest to what is widely known as infrastructure but still would exclude the water and sanitation sectors as well as the telecom sector. The omission of the latter sector in this analysis is not a major problem because in most countries the associated expenditures are now broadly the private sector’s responsibility. The omission of water and sanitation is much more problematic, but it is not the only issue. Fuel and energy combines anything that would have to do with fuel exploration costs and leaves out the very significant role of utilities in the sector. The upshot is that the data allow only for a very partial view of the infrastructure sector.

The time span is 1990 to 2002, although data are not available for every year of this period for every country. The data coverage available for each infrastructure subsector that can be matched with outcome data during that period is shown in table 1. The table distinguishes between general government and central government only to show that any effort to account for all expenditure sources has an overwhelming cost in terms of country coverage. It also shows how much more the coverage would drop if we wanted to cover the same specific countries for the two subsectors. Working with the central government data is a problem because the data on outcomes are poor. These data series are not ideal, but they are the only ones available to maintain a relatively large data sample at this stage.

| | General government | | Central government | |
|-------------------------------------|------------------------|---------------------|------------------------|---------------------|
| | Number of observations | Number of countries | Number of observations | Number of countries |
| Transport | 50 | 21 | 606 | 84 |
| Energy and fuel | 35 | 21 | 511 | 71 |
| Health | 18 | 3 | 147 | 61 |
| Education | 9 | 9 | 142 | 45 |
| Memo items: | | | | |
| Energy + Transport = Infrastructure | 24 | 12 | 75 | 13 |
| Health + Education = Social | 2 | 2 | 49 | 36 |
| Social + infrastructure | 0 | 0 | 5 | 5 |

Source: IMF Government Finance Statistics database.

Looking at table 1, we reached the initial conclusion that we would be better off assessing the partial performance of governments in each subsector for as many of the years for which we have information.¹⁵ For instance, given the outcome data we have, efforts to cover energy and transport in the same country for the same years would limit our sample size to 13 countries. For health and education, the sample size would be limited to 36 countries. Looking at sectors one by one allows us to work with at least 45 countries (education) and up to 84 countries (transport). The specific countries covered for each sector are reported in the appendix. With this approach, there is enough data to generate the “big picture” country groupings (see table 2).

| Country income category | Transport | Energy and fuel | Health | Education |
|-------------------------|-----------|-----------------|--------|-----------|
| High-income | 225 | 216 | 244 | 244 |
| Low-income | 179 | 159 | 142 | 142 |
| Lower-middle-income | 198 | 199 | 214 | 214 |
| Upper-middle-income | 167 | 162 | 208 | 208 |
| Total | 769 | 736 | 808 | 808 |

The problems with the data have by now been well established. Still, we should report additional issues that emerged during our data-cleaning exercise. First, on the outcome side,

¹⁵ Another way to increase the sample size would be to look at blocks of years or to work with moving averages, recognizing that data are not available for every year. Another option could be to do some more sophisticated smoothing, involving Kalman filters or kernel estimators, but this would be a major project.

the international databases provide some of the outcome data in flows of services and some in stocks. Taking the changes in stocks (as in the case of roads) as an indication of services is, however, of limited interest because they do not necessarily reflect the services associated with the stock. Second, matching outcomes with expenditures is not as simple as it sounds. Capital expenditures are typically “oversized” in the short run because they are designed to meet long-term demand. We have no way of knowing, however, the average lead time of demand built in to the existing capital stock in any given year. There is thus a risk of overestimation of the efficiency problems simply because we are not picking up the timing issue properly. Third, the expenditure data are clearly problematic as well. To have a reasonable sample size, we need to work with central government expenditure data. We have no way of assessing correctly how much public-sector expenditure by subnational governments this assumption ignores. It should not be a major problem for infrastructure stocks because such a large portion is generally financed by the central government, not the entire government. It could be a much bigger problem for services, however, because subnational governments often take the responsibility for a fair share of those.

The relevance of the private sector’s role is also ignored by the approach followed here because nobody really monitors this role in any of the sectors.¹⁶ Finally, there are very basic data issues. For many countries, we found data that did not pass some basic common-sense tests, let alone basic statistical tests, whether on expenditure or outcome. We ended up dropping many countries because of this concern about data quality. We dropped a few more after the conversion of all the monetary data into purchasing power parity (PPP) because it also created outliers. The conversion to constant prices of data on energy and fuel was particularly problematic since a large proportion of the changes in expenditure levels can be traced to prices rather than quantities.

Emerging facts on the evolution of central government expenditure

Here we summarize the basic information on central government expenditures for the various sectors and the various country groupings. The country coverage for each subsector is shown in table 3. The distribution of countries across income groups is reasonably even, with 20–25 countries for almost each group. The sample size for low-income countries is somewhat smaller (with 16 countries), but it is still statistically significant.

¹⁶ WHO may have the best set of data on the role of the private sector (in health), and Calderon and Servén (2004) have the best data set, but only for investment in Latin America. There is no other quantitative cross-country information on the relative importance of the private sector in any of the sectors covered here. There is actually very little qualitative information on this topic.

| Country income category | Transport | Energy and fuel | Health | Education |
|-------------------------|-----------|-----------------|--------|-----------|
| High-income | 25 | 23 | 24 | 23 |
| Low-income | 16 | 15 | 16 | 16 |
| Lower-middle-income | 23 | 22 | 23 | 23 |
| Upper-middle-income | 22 | 20 | 23 | 23 |
| Total | 86 | 80 | 86 | 85 |

The sectoral expenditure levels for each category are presented in figure 1. It is important to keep in mind that these levels are at the lower bounds because they do not pick up any expenses by subnational governments or public enterprises. Figure 1 shows that for all developing-country groups, education represents the largest share of the central government expenditure analyzed here, with an average share of 3.1 percent—and up to 3.6 percent for the lowest-income groups. For high-income countries, health expenditure represents the largest share at 3.9 percent of GDP for our sample of high-income countries—the average for the full sample is 2.6 percent. The transport sector represents a nonnegligible interest of the public sector in all country groups, with an average across countries of about 1.5 percent of GDP.

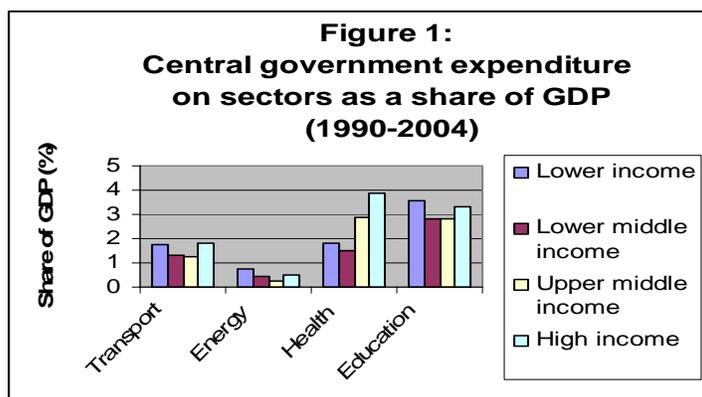
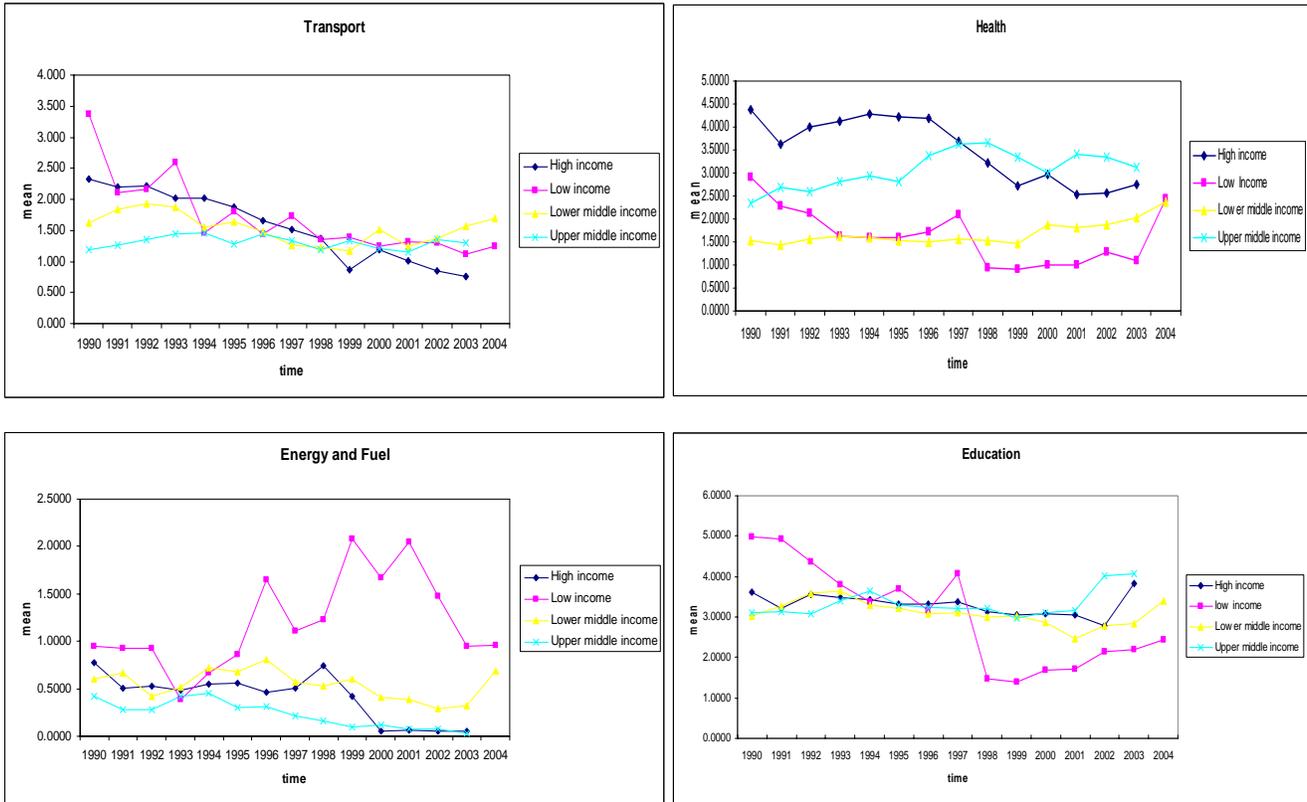


Figure 2 provides additional insights on the evolution of these shares over the past 15 years or so. The emerging story is quite heterogeneous across sectors and income groups. Transport has seen its share of central government expenditure decline in all country groups, except upper-middle-income countries, where it has remained fairly stable following a slow upward trend. Fuel has followed a major downward trend except in low-income countries, where rising oil costs have taken a toll on the budget. Education and health have been relatively stable except in low-income countries, where, after a decline, they have started to recover slowly.

Figure 2 Evolution of central government expenditure (share of GDP at constant 2000 prices)



It is clearly difficult to offer much more than a basic description of trends across categories. A much better analysis would consider all the major changes that have taken place in the organization of government in each country but this data is not available although we know it is relevant. Indeed, clear potential explanations for any changes observed over the past 15 years are the increased involvement of the private sector and of subnational governments in all of these sectors. Among national policies that might explain the changes are the deregulation or regulation of the sector and increased (or changed) demand for subsidies directed at the poor. However, very little data on these two dimensions has been collected in a comparable way across sectors.

Supranational changes, too, would seem to be relevant in this regard; shifting priorities among bilateral and multilateral donors would probably contribute as much as shifts in national policies—some would argue that there is a strong correlation between the national and international changes. The supranational changes are particularly important for the poorest small countries, because such a large share of many of their investment programs is financed by donor loans and grants.

The main point to be made here is that any assessment of the performance of governments based on international datasets is going to be limited by the lack of documentation of these sorts of institutional changes. Measuring reforms is as important as

measuring expenditures and outcomes and relating the results to a benchmarking exercise that compares the relative performance of governments. Nevertheless, we tried to ignore this limitation to get a sense of what the existing data could reveal in spite of the limitations.

The emerging facts on outcomes

To match the analysis of the basic statistics on sectoral expenditures, it may be useful to give a sense of the evolution of the outcome indicators for each of the subsectors. Table 4 offers a brief look at the specific outcome indicators with which we worked. As in the case of public expenditure, the data had great limitations. Ideally, we would have had a long list of outputs associated with each category of expenditure and a good idea of the average time lag between expenditures and expected outcomes. That kind of information is not available.

| Table 4 Outcome indicators | | | |
|-----------------------------------|---|---|---|
| Transport | Energy | Health | Education |
| Length of road network (km) | Energy use per PPP GDP (kg oil equivalent per constant PPP\$) | Infant mortality rate (per 1,000 live births) | Primary completion rate (percent of relevant age group) |
| Aircraft departures (number) | Electric power consumption (Kwh/per capita) | Under-5 mortality rate (per 1,000) | School enrollment, primary (percent net) |
| | | Immunization, DPT (percent of children ages 12-23 months) | School enrollment, secondary (percent net) |
| | | | Repetition rate, primary (percent of total enrollment) |

Note: PPP = purchasing power parity.

The solution we adopted was again very pragmatic. We looked for outcome or output indicators that could be viewed as relatively useful proxies and for which data were available for the same countries and years for which we had expenditure data. The need to match country coverage with expenditures for all sectors for as many years as possible is what limited the sample sizes the most. Figure 3 provides a visual sense of the evolution of these indicators across income groups.

The figures are clearly intended to give only a very general sense of the evolution of outcomes.¹⁷ In fact, this evolution and the relative position of each income group may be the only two useful contributions of these figures. Both are consistent with expectations—that is, higher-income groups do better than lower-income groups in each sector, and in general outcomes are improving over time across income groups. There are clearly problems with

¹⁷ An important detail to keep in mind is that some of the outcomes are defined in positive terms (more is better, e.g., completion rates) and some in negative terms (less is better, e.g., mortality).

some of the data series in some of the years; in addition, some of the significant increases in coverage that seem to take place in one year are hard to explain unless something is wrong with the datasets for those specific years. These problems are not damaging enough to question the emerging big picture on the evolution of outcomes per income groups.

Figure 3 Evolution of sectoral outcome indicators by country groups, 1990–2003



Some rough efficiency measures

This section provides some illustrative estimates of the efficiency levels achieved by the central government in delivering sectoral services. We report only the results that are reasonably robust. We tried many combinations of sectors. In particular, we tried to assess the multiple output responsibilities of government (education, health, and infrastructure). Unfortunately, the data problems prevented us from producing the multisector analysis we had hoped for.

We ended up focusing on four sector specific panels of data (table 5). For each panel, we estimated two models. The first is the simplest possible distance function. The second is a model corrected to account for institutional variables such as the quality of the bureaucracy and the level of corruption in the country. The models generate a measure of efficiency at the country level, but because we believe that country-specific data often are not reliable, we only report the aggregated data at the country-group level.

| Table 5 Panel used in the estimation for each sector | | | | | |
|---|-----------|---------------------|------------------------|----------------------------|-------------------------------|
| Panel | Sector | Number of countries | Number of observations | Number of outputs/outcomes | Number of inputs ^a |
| 4 | Transport | 81 | 547 | 2 | 1 (4) |
| 5 | Energy | 66 | 481 | 2 | 1 (3) |
| 7 | Health | 60 | 146 | 2 | 1 (1) |
| 9 | Education | 36 | 103 | 3 | 1 (1) |

Note: Any difference with table 1 is due to the fact that table 1 is based on a matching of outputs and inputs, whereas table 5 adds matching for the availability of data on corruption and bureaucratic quality.

a. The first number is the number of expenditure categories; the second (in parentheses) is the number of control variables used in a second group of regressions (i.e., corruption, quality of bureaucracy, GDP, population, GDP/capita.)

These datasets are used to estimate a distance function, which is then used to generate the efficiency estimates for each sector. The empirical application of a distance function requires the specification of an appropriate functional form. We rely on a translogarithmic functional form (hereinafter *translog*) because it is flexible, easy to calculate, and allows the imposition of the homogeneity condition for outputs (homogeneity of level 1 in outputs). It offers a flexible functional form providing a local second-order approximation to an unknown functional form. In other words, we made no assumptions regarding *a priori* restrictions about production technology. Finally, we specify an output orientation in the results reported here. In other words, we try to assess the extent to which with the current level of public expenditures (the input), we can achieve more outputs.

Formally, the function can be expressed as follows:

$$\begin{aligned}
\ln D_O = & \alpha_0 + \sum_{m=1}^M \alpha_m \ln y_{mit} + 1/2 \sum_{m=1}^M \sum_{n=1}^M \alpha_{mn} \ln y_{mit} \ln y_{nit} + \\
& \sum_{k=1}^K \beta_k \ln x_{kit} + 1/2 \sum_{k=1}^K \sum_{l=1}^K \beta_{kl} \ln x_{kit} \ln x_{lit} + \sum_{k=1}^K \sum_{m=1}^M \delta_{km} \ln x_{kit} \ln y_{mit} + \\
& \sum_{h=1}^H \psi_h d_h + \sum_{t=1}^T \gamma_t f_t + \varepsilon_{it}
\end{aligned} \tag{1}$$

Where y is a vector of M outputs, x is a vector of K factors, i relates to the i -th country, t relates to the time trend, h refers to the environmental/institutional variables, Ψ_t is the coefficient of the environmental/institutional dummy variables d , γ_t is the coefficient for the time dummy f , and ε_{it} is an error term. Variables are expressed in relation to their deviation from the geometric mean; therefore, the estimated coefficients can be construed as elasticities at the sample mean.

In order to determine the frontier, D_O needs to be equal to 1. When that is the case, the left side of the equation will equal zero (i.e., $\ln 1=0$). This is why we need to assume homogeneity of degree 1. This condition is imposed by normalizing the distance function with one of the outputs.¹⁸ This starts from the assumption that homogeneity implies that:

$$D_O(x, wy) = w D_O(x, y) \tag{2}$$

for any $w>0$. The output chosen does not influence the results (Cuesta and Orea 2002).

If in a translog distance function any output is chosen, say y_M , so that $w = 1/ y_M$, then the following expression results:

$$\ln(D_O/y_M) = TL(x_{it}, y_{it}/y_{Mit}, \alpha, \beta, \delta) \tag{3}$$

yielding the final expression:

$$-\ln(y_{Mit}) = TL(x_{it}, y_{it}/y_{Mit}, \alpha, \beta, \delta) - \ln(D_O) \tag{4}$$

The $-\ln(D_O)$ term can be interpreted as an error term that captures technical inefficiency.

¹⁸ This approach is quite common now. It has been applied recently by Coelli and Perelman, 1999, 2000 or Orea, 2002, among others.

The distance function estimated is stochastic. We decided to rely on an econometric estimation rather than on a nonparametric approach simply because the quality of the data was so poor and the number of outliers so great that we needed to be able to conduct a minimum number of tests to check the robustness of the results. Most of these tests are not available for the nonparametric approaches. Moreover, there is ample evidence to suggest that the nonparametric models are much more sensitive to outliers in the assessment of efficiency levels.

To estimate equation (4), the random disturbance term needs to be determined. The most common method to do so was developed by Battese and Coelli (1988) for panel data. It applies an additive term as suggested by Cuesta and Orea (2002) to account for the fact that we are estimating an output-oriented distance function. The error terms thus have the following form

$$v_{it} + u_i \tag{5}$$

where v_{it} is a symmetrical error term, iid with a zero average (which represents the random variables uncontrollable by the government), and u_i is a one-sided negative error term (which measures the technical inefficiency of each operator that is constant over time) and is distributed independently of v_{it} .

Applied to the distance function, this yields

$$-\ln(y_{Mit}) = TL(x_{it}, y_{it} / y_{Mit}, \alpha, \beta, \delta) + v_{it} + u_i \tag{6}$$

This equation can be estimated by the maximum-likelihood method, which requires distributional assumptions concerning the random shock. This assumes that v_{it} follows a $N(0, \sigma_v^2)$ distribution and u_i follows a $|N(0, \sigma_u^2)|$ distribution (Ritter and Simar 1997). We estimated the models with and without the environmental and institutional variables, but neither made a difference to the estimated efficiency levels. In other words, our proxies for the quality of bureaucracy and corruption were not statistically significant. The most interesting result is that while the expenditure levels are significant in all our final models, the associated elasticities are all quite low (below 0.05). The trend variables, however, are statistically significant. The trend variable is surprisingly negative in transport but positive in energy and insignificant in health and education, illustrating the differences in the importance of technological change across sectors.

Table 6 presents a summary of the average efficiency levels estimated for each sector for the 1990–2002 period. It does not make much sense to compare sectors or to try to infer meaning from the efficiency estimates because the specific levels in each sector are largely

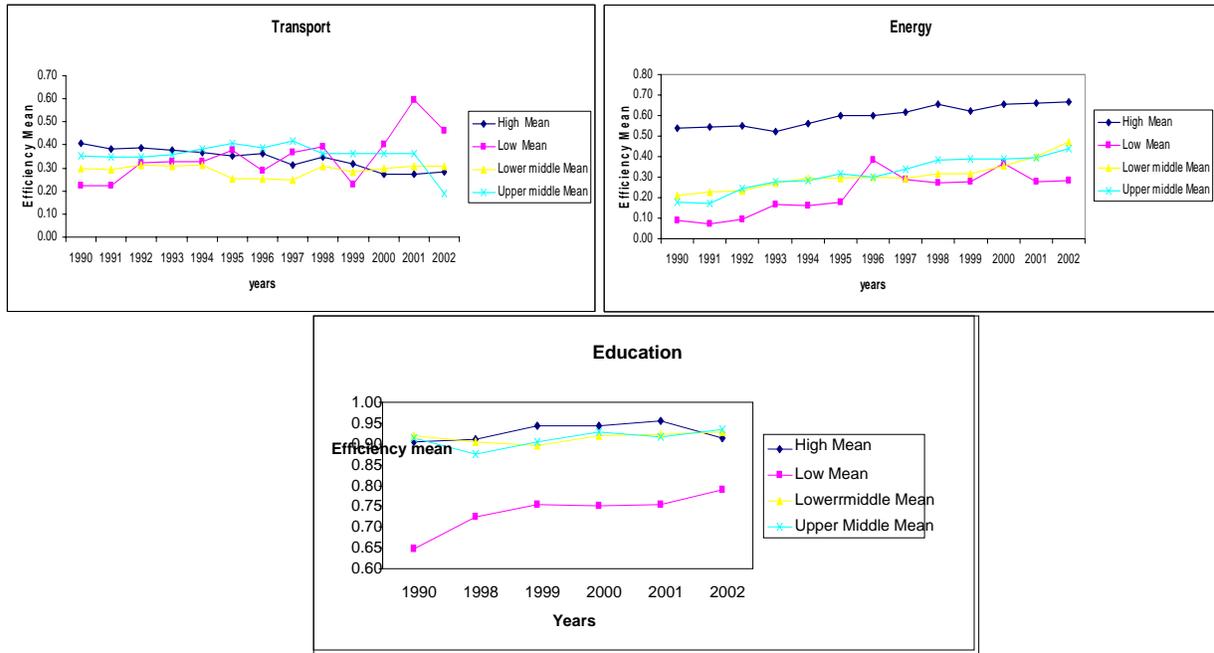
driven by the quality of the data and the specific measures of outputs and inputs used to estimate the model. In sum, the models are not really that useful in answering the original questions about the scope for improvement in outputs or in costs across sectors.

| Country income categories | Transport | Energy | Health | Education |
|----------------------------------|------------------|---------------|---------------|------------------|
| High-income | 0.36 | 0.59 | n.a. | 0.93 |
| Low-income | 0.33 | 0.23 | n.a. | 0.74 |
| Lower-middle-income | 0.29 | 0.30 | n.a. | 0.92 |
| Upper-middle-income | 0.37 | 0.32 | n.a. | 0.91 |

Because we could not really come up with a reasonably good model for health, we do not report the results from the time-series analysis, although we will discuss results from a cross-section of 192 countries at the end of this section.¹⁹ Within each of the other sectors, however, the supporting econometric results are robust enough, and a story can be squeezed out for each subsector. That story differs across the subsectors. A discussion of the evolution of each of these for each country group generates a fuller picture of the global evolution of government expenditure. Figure 4 provides a graphical presentation of this evolution. The main potential use of these results is to get a sense of the relative efficiency levels, even if it is sometimes useful to ignore some of the outlier years.

¹⁹ The estimated model has the wrong signs on some of the parameters.

Figure 4 Evolution of levels of efficiency of government expenditure, 1990–2002



In transport, the story is driven by what happens in the road sector. Given that support for the roads sector is one of the largest sources of public expenditure in transport, the general results should not be too far from what would have been generated using a full set of information on public expenditures. The main story emerging is that there are no statistically significant differences across income groups. The richer countries do a little bit better, but this is to be credited mostly to the much higher level of air traffic picked up in our second output for transport.

In energy, the story is very different. The performance levels differ sufficiently across income groups to imply some relevance. On the one hand, the poorer the countries, the more likely it is that the expenditure composition will be dominated by investment. On the other hand, the richer the country, the more likely it is that operation and maintenance expenditures will drive the story. Ideally, we should be able to assess the efficiency of capital expenditures (Capex) and operational expenditures (Opex) separately, as is done for private utilities under a price cap regime, but the data needed to do that do not exist at this level of aggregation. There are other problems as well. For instance, the fact that the cost of fuel is picked up in the assessment may distort the story, in particular given that the expenditure levels are expressed in constant prices. But this should not be sufficient to explain the nearly three times difference between the poorest and richest countries. In spite of the data problems, the upshot of our results is that even if there seems to be a clear scope for improvements in efficiency, the poorest countries are catching up to some degree.

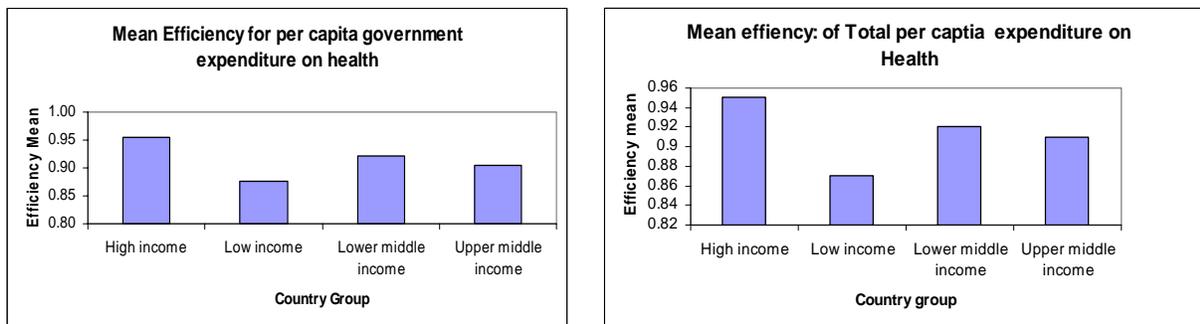
For education, there are differences across income groups, but these are predictable given the indicator used and that quality considerations are not well modeled. The most

interesting contribution of this naïve analysis may simply be the visual confirmation that the poorest country group shows signs of catching up. The trend in efficiency is indeed increasing significantly over the 12-year period covered by the sample.

Although the panel data and models for health proved to be too unreliable to deserve much coverage here, we relied on the latest data offered by WHO (2005) to generate a comparison across income groups. We used the general theoretical production function as with the panel data, but focused on different inputs and outputs. We considered two outputs, life expectancy at births for males and for females, expressed in years. For inputs, we relied on one of the following variables: per capita government expenditure on health or total expenditure on health. The data on output were for 2004; on inputs, for 2003. All monetary variables are expressed in U.S. dollars and PPP. The results are not significantly affected by the choice of the input variable, as seen in figure 5.

Once more the average efficiency level estimated for the sector is not that helpful, either because of its specificity to the model or the gaps in relevant information, but the relative position of the various income groups is interesting and has proven quite robust to the various models we tried (as it was with the models relying on panel data). The only problem is that we cannot track the evolution of efficiency. As in energy or education, the results are expected. The highest-income countries have the highest efficiency levels, and the lowest-income countries the worst. Somewhat surprisingly, income level is not a perfect predictor of efficiency as defined here, since the upper-middle-income countries did somewhat worse than those in the lower-middle-income group in 2004. The important gap, however, is that between the lowest income groups and the other groups.

Figure 5 Efficiency of health expenditure in 2004



Overall, this section tells us simply that in the poorest countries there is more scope for performance improvements in all sectors than in higher-income countries. This is not really news. Moreover, because the specific performance levels are driven by the choice of models and outputs, they should not be taken at face value. The only thing that comes out strongly enough is that the differences across income groups are significant and that in education and in energy at least, the lower-income countries are not catching up very quickly, at least for the outcome measures with which we worked.

Concluding comments

It may be useful to conclude by returning to the two questions presented at the outset of this paper. Roughly, they are: How much more can the government do in each sector with its current resource allocation? And how much cheaper could today's level of intervention be? Based on the data and analytical work reported here, the precise answer to these questions remains: we do not know. A more imprecise, but probably still safe, answer would focus on the wedge between the poorest countries and the others, arguing that there is evidence of significant gains to be achieved considering the differences in efficiency levels across income groups. With the models we used, it was in energy and education that the performance of the low-income countries fell furthest below the benchmark set by the higher-income groups.

By how much could performance improve? Once more, the honest answer is: we do not know—not much help to policy makers. In these days in which quantitative targets of any type seem to be popular—from the Millennium Development Goals to the number of rules and regulations thought to impede business—this is probably not an acceptable reply for the donor community. It may be useful to consider setting a target for efficiency improvements—making the strong assumption that the international community can manage to agree on the inputs and outputs to consider and then to collect the relevant data. A look at the numbers would suggest that a short-to medium term improvement in both sectors of 25 to 35 percent over current levels would not be an unrealistic lower-limit target, at least in terms of the outputs identified here.²⁰

At the sectoral level, the scope for improvement across income groups seems to be largest in infrastructure sectors. This is particularly clear for transport, where the use of a stock variable (instead of a services variable) is a clear problem when setting targets. Investments in these sectors are amortized in a much slower way than in education or health. Larger initial investments appear to be less efficient early in the life of a project. Because we do not know the average age of the capital stock in question, it is probably safe to stick to a lower limit for any target. This is what justifies a 25 to 35 percent improvement even when current efficiency levels appear to be only around 30 percent in energy and transport and it would seem possible to triple current outcomes.

Ultimately, these numbers are presented as a conservative lower boundary to get a global sense of where country groups are. But they can hardly be used to make country-specific recommendations. The technical problems underlying these numbers are not minor, and a lot more work would be needed before a more robust “guestimate” could be generated from the current data. In spite of all the good research that has been published, very little has

²⁰ This is clearly an arbitrary target but it offers a rough realistic order of magnitude.

offered quantitative evidence suitable for use in fiscal diagnostics. A number of technical directions should be explored to refine significantly the results presented here.

The simplest solution would be to test the model on a cross-section dataset made up of the latest data available—as we did for health with the WHO data. By doing this, one would obtain a useful snapshot from relatively large datasets. But it would not say much about the extent to which the poorest are catching up and at what speed. If we want to get a sense of the speed at which things are changing or could change, we need dynamic models. To meet that requirement we could use, for the four sectors, a pure output-expansion model in which the inputs would be used in a second stage to explain differences in efficiency. If a weak or negative relation appeared, we would need to test some alternatives—for example, to group a year in windows analysis (for instance, the mean of year 1 and 2 compared with the mean output of year 2 and 3). If this alternative setting were again inconclusive, it is unlikely that much more could be achieved than what has already been done here. Alternative solutions along the same line would be to look for cost drivers or to work with national accounts data to increase cross-country and intrasector consistency, since the omission biases are likely to be less dramatic with those datasets. The challenge there would be to unbundle the public- from the private-sector providers.

In sum, as evidenced by the long list of papers included in the bibliography, some things can be done better than what we presented here, at least for education and possibly for health. Unfortunately, any recommended use of these sorts of assessments for country-specific policy purposes will have to be presented with extremely heavy words of caution. Ultimately, policy-relevant work should be conducted at the level of the business unit in each country, but this requires a lot more data than are currently available because, historically, monitoring the performance of the government at the sector level has not really been anyone's top priority. Until the international community begins to recognize that the sectoral allocation of public expenditure matters, today's data gaps are unlikely to disappear and the monitoring of the performance of the public sector will continue to be only moderate at the aggregate level and superficial at best at the sectoral level.

Appendix Countries included in sector samples

| Transport | Energy and fuel | Health | Education |
|--------------------|--------------------|--------------------|----------------------|
| Albania | Argentina | Albania | Albania |
| Argentina | Australia | Argentina | Azerbaijan |
| Australia | Austria | Azerbaijan | Bangladesh |
| Azerbaijan | Azerbaijan | Bangladesh | Belarus |
| Bahrain | Bahrain | Belarus | Bolivia |
| Belarus | Bangladesh | Bolivia | Bulgaria |
| Bolivia | Belarus | Brazil | Chile |
| Brazil | Bolivia | Bulgaria | Costa Rica |
| Bulgaria | Brazil | Burundi | Croatia |
| Burundi | Bulgaria | Cameroon | Dominican Republic |
| Cameroon | Cameroon | Chile | El Salvador |
| Canada | Canada | China | Estonia |
| China | Costa Rica | Costa Rica | Georgia |
| Costa Rica | Croatia | Croatia | Hungary |
| Croatia | Czech Republic | Czech Republic | Indonesia |
| Cyprus | Denmark | Dominican Republic | Jamaica |
| Czech Republic | Dominican Republic | Egypt, Arab Rep. | Kazakhstan |
| Denmark | Egypt, Arab Rep. | El Salvador | Latvia |
| Dominican Republic | El Salvador | Estonia | Lithuania |
| Egypt, Arab Rep. | Estonia | Gambia, The | Mauritius |
| El Salvador | Finland | Georgia | Mexico |
| Estonia | France | Hungary | Moldova |
| Finland | Germany | India | Mongolia |
| France | Hungary | Indonesia | Morocco |
| Georgia | Iceland | Iran, Islamic Rep. | Panama |
| Germany | Indonesia | Jamaica | Peru |
| Hungary | Iran, Islamic Rep. | Kazakhstan | Poland |
| Iceland | Ireland | Latvia | Romania |
| India | Israel | Lebanon | Seychelles |
| Indonesia | Jamaica | Lesotho | Slovak Republic |
| Iran, Islamic Rep. | Kazakhstan | Lithuania | Syrian Arab Republic |
| Ireland | Kuwait | Madagascar | Tajikistan |
| Israel | Latvia | Malaysia | Tunisia |
| Jamaica | Lebanon | Mauritius | Ukraine |
| Japan | Lithuania | Mexico | Uruguay |
| Kazakhstan | Luxembourg | Moldova | Venezuela, RB |
| Kuwait | Malta | Mongolia | |
| Latvia | Mexico | Morocco | |
| Lebanon | Moldova | Nepal | |
| Lesotho | Morocco | Nicaragua | |
| Lithuania | Netherlands | Pakistan | |
| Madagascar | Nicaragua | Panama | |
| Malaysia | Norway | Peru | |
| Malta | Pakistan | | |

| | | |
|----------------------|----------------------|----------------------|
| Mauritania | Panama | Poland |
| Mauritius | Peru | Romania |
| Mexico | Poland | Russian Federation |
| Moldova | Romania | Seychelles |
| Mongolia | Russian Federation | Slovak Republic |
| Morocco | Slovak Republic | South Africa |
| Nepal | Slovenia | Syrian Arab Republic |
| Netherlands | South Africa | Tajikistan |
| Nicaragua | Sweden | Thailand |
| Norway | Switzerland | Trinidad and Tobago |
| Pakistan | Syrian Arab Republic | Tunisia |
| Panama | Thailand | Turkey |
| Peru | Trinidad and Tobago | Ukraine |
| Poland | Tunisia | Uruguay |
| Romania | Turkey | Vanuatu |
| Russian Federation | Ukraine | Venezuela, RB |
| Seychelles | United Arab Emirates | Zimbabwe |
| Singapore | United Kingdom | |
| Slovak Republic | United States | |
| Slovenia | Uruguay | |
| South Africa | Venezuela, RB | |
| Sweden | Zimbabwe | |
| Switzerland | | |
| Syrian Arab Republic | | |
| Tajikistan | | |
| Thailand | | |
| Trinidad and Tobago | | |
| Tunisia | | |
| Turkey | | |
| Ukraine | | |
| United Arab Emirates | | |
| United Kingdom | | |
| United States | | |
| Uruguay | | |
| Vanuatu | | |
| Venezuela, RB | | |
| Zimbabwe | | |

References

- Afonso, A., L. Schuknecht, and V. Tanzi. 2006. "Public sector efficiency: Evidence for new EU member states and emerging markets." European Central Bank, Working Paper 581.
- . 2005. Public sector efficiency: An international comparison." *Public Choice* 123 (3–4): 321–347.
- . 2006. "Quality of public finances and growth." European Central Bank, Working Paper 538.
- Afonso, A., and [[Name or initial dropped here?]] St. Aubyn. 2005. "Non-parametric approaches to education and health efficiency in OECD countries." *Journal of Applied Economics* 8 (2): 227–246.
- Agell, J., T. Lindh, and H. Ohlsson. 1997. "Growth and the public sector: A critical review essay." *European Journal of Political Economy* 13 (1): 33–52.
- . 1997. "Growth and the public sector: A reply." *European Journal of Political Economy* 15 (2): 359–366.
- Al-Samarrai, S. Forthcoming. "Achieving education for all: How much does money matter?" *Journal of International Development*.
- Anand, S., and M. Ravallion. 1993. "Human development in poor countries: On the role of private incomes and public services." *Journal of Economic Perspectives* 7 (1): 133–150.
- Asian Development Bank. 2006. *Key indicators 2006: Measuring policy effectiveness in health and education*. Manila
- Atkinson, A. 2005. *The Atkinson Review: Final Report: Measurement of government output and productivity for the national accounts*. Basingstoke, Hampshire, and New York: Palgrave Macmillan. http://www.statistics.gov.uk/about/data/methodology/specific/PublicSector/Atkinson/final_report.asp
- Battese, G. E., and T. J. Coelli . 1995. "A model for technical inefficiency effects in a stochastic frontier production for panel data." *Empirical Economics* 20: 325–333.
- . 1988. "Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data." *Journal of Econometrics* 38: 387–399
- Battese, G. E., and G. S. Corra. 1977. "Estimation of a production frontier model: with application to the pastoral zone of eastern Australia." *Australian Journal of Agricultural Economics* 21: 169–179.

- Bidani, B. and M. Ravallion. 1997. "Decomposing social indicators using distributional data." *Journal of Econometrics* 77: 125–139.
- Bose, N. M., Emranul Haque, and D. R. Osborn. 2004. "The composition of public expenditure and economic growth in developing countries." *Global Journal of Finance and Economics* 1 (1): 97–117.
- Calderon, C. and L. Servén. 2004. "Trends in infrastructure in Latin America, 1980-2001," Policy Research Working Paper Series 3401, The World Bank.
- Cherchye, L., and W. Moesen. 2004. "Institutional infrastructure and economic performance: levels vs. catching up and frontier shifts." Mimeo, Catholic University of Leuven, Campus Kortrijk.]
- Coelli T., D. Rao, C. J. O'Donnell, and G. E. Battese. 2005. *An introduction to efficiency and productivity analysis*, second ed. Springer. [[pls provide city of publication.]]
- Coelli, T., A. Estache, S. Perelman and L. Trujillo 2003. *A Primer on Efficiency Measurement for Utilities and Transport Regulators*, World Bank Institute Publications, Studies in Development Series, Washington, DC
- Coelli and Perelman. 2000. "Technical efficiency of European railways: A Distance Function Approach", *Applied Economics* 32(15): 1967-76
- Coelli and Perelman 1999 "A comparison of parametric and non-parametric distance functions: with application to European Railways", *European Journal of Operations Research* 117(2): 1967-76
- Cuesta, R. A., and L. Orea . 2002. "Mergers and technical efficiency in Spanish saving banks: a stochastic distance function approach." *Journal of Banking and Finance* 26: 2231–47.
- Devarajan, S., V. Swaroop, and H. Zou. 1996. "The composition of public expenditure and growth." *Journal of Monetary Economics* 37: 313–344.
- Dreher, A., J. E. Sturm, and H. Ursprung. 2006. "The impact of globalization on the composition of government expenditures: Evidence from panel data." CESifo Working Paper 1755.
- Fried, H., K. Lovell, and S. Schmidt, eds. 1993. *The measurement of productive efficiency: Techniques and applications*. Oxford: Oxford University Press.
- Gupta, S., and M. Verhoeven. 2001. "The efficiency of government expenditures: Experiences from Africa." *Journal of Policy Modeling* 23: 433–67

- Gupta, S., M. Verhoeven, and E. R. Tiongson. 2002. "The effectiveness of government spending on education and health care in developing and transition economies." *European Journal of Political Economy* 18 (4): 717–37.
- Karras, G. 1996. "The optimal government size: further international evidence on the productivity of government services." *Economic Inquiry* 34 (2): 193–203.
- Herrera, S., and G. Pang. 2005. "Efficiency of public spending in developing countries: An efficiency frontier approach." World Bank, Policy Research Working Paper 3645.
- Jayasuryia, R., and Q. Wodon. 2002. *Efficiency in reaching the Millennium Development Goals*. World Bank Working Paper 9, Washington, DC.
- Mayne, J. and E. Zapico-Goni, ed. 1997. *Monitoring performance in the public sector—Future directions from international experience*. New Brunswick, NJ: Transaction.
- Musgrove, P. 1996. "Public and private roles in health: Theory and financing patterns, World Bank Discussion Paper 339, Washington, DC.
- OECD. 2003. "Enhancing the cost effectiveness of public spending." *Economic Outlook*, 2003/02 (74). December.
- Pestieau, P. 2006. "Assessing the performance of the public sector." Unpublished paper, Poverty Reduction and Economic Management Vice Presidency, World Bank, Washington, DC.
- Pradhan, S. 1996. "Evaluating public spending." World Bank Discussion Paper 323, Washington, DC.
- Ravallion, M. 2005. "On measuring aggregate 'social efficiency.'" *Economic Development and Cultural Change* 53 (2): 273–292.
- Ritter, C. and L. Simar 1997. "Pitfalls of normal-gamma stochastic frontier models", *The Journal of Productivity Analysis*, 8, 167-182
- Sanz, I., and F. J. Velazquez. 2001. "The composition of public expenditure and growth: difference models of government expenditure distribution by functions." Discussion Paper 0115, Department of Economics, University of Otago, New Zealand.
- Shephard, R. W. 1953. *Cost and production functions*. Princeton, NJ: Princeton University Press.
- . 1970. *Theory of cost and production functions*. Princeton, NJ: Princeton University Press.
- Vandenberghe, V., and S. Robin. 2004. "Evaluating the effectiveness of private education across countries: a comparison of methods." *Labour Economics* 11: 487–506.

WHO. 2005. WHO Statistical Information System (WHOSIS), available at http://www3.who.int/whosis/core/core_select.cfm