EFFICIENCY OF PUBLIC SPENDING IN DEVELOPING COUNTRIES: AN EFFICIENCY FRONTIER APPROACH

Santiago Herrera and Gaobo Pang[*]

1. Introduction

Governments of developing countries typically spend resources equivalent to between 15 and 30 per cent of GDP. Hence, small changes in the efficiency of public spending could have a significant impact on GDP and on the attainment of the government's objectives whichever these are. The first challenge faced by stakeholders is measuring and scoring efficiency. This paper attempts such quantification. Additionally it verifies statistically some empirical regularities that describe the cross-country variation in the estimated efficiency scores.

The paper has four chapters following this Introduction. The first one presents the methodology that defines efficiency as the distance from the observed input/output combinations to an efficient frontier. This frontier, defined as the maximum attainable output for a given input level, is estimated using the Free Disposable Hull (FDH) and Data Envelopment Analysis (DEA) techniques. The exercise focuses on health and education expenditure because they absorb the largest share of most countries' budgets, and because of lack of data availability for international comparisons in other types of expenditures.

The second chapter estimates the efficiency frontiers for nine education output indicators and four health output indicators, based on a sample of 140 countries and data for 1996-2002. Both input efficiency (excess input consumption to achieve a level of output) and output efficiency (output shortfall for a given level of inputs) are scored. The chapter presents both the single input/single output and the multiple inputs/multiple outputs frameworks. In addition, this chapter explores how expenditure efficiency has changed over time.

The third chapter seeks to identify empirical regularities that explain cross-country variation in the efficiency scores. Using a Tobit panel approach, this chapter shows that higher expenditure levels are generally associated with lower efficiency scores. Similarly, countries in which the wage bill is a larger share of the total budget tend to have lower efficiency scores. Three other variables that explain

[*] World Bank.

The paper is a draft of work in progress that carries the names of the authors and should be cited accordingly. The findings, interpretations and conclusions are the authors' responsibility and do not necessarily represent the view of the World Bank or its Executive Directors. The authors thank Antonio Estache, Jorge Garcia, Ulrich Lachler, Vikram Nehru, participants of the Economic Policy and Debt seminar, and participants of the Banca d'Italia Workshop on Public Finance for helpful comments on earlier drafts of the paper. The dataset, specific-country results, and programs are available in the World Bank PRMED (Economic Policy and Debt Department) website http://intranet.worldbank.org/WBSITE/INTRANET/SECTORS/INTECONOMICPOLICY/INTQFA/0,,menuPK:369265~pagePK:151716~piPK:176772~theSitePK:369258,00.html

the cross country variation in efficiency scores are the degree of urbanization (positively correlated with efficiency, the prevalence of the HIV/AIDS epidemic (negatively associated with efficiency scores), and inequality in income distribution (higher inequality associated with lower efficiency).

The fourth and last chapter summarizes the conclusions.

2. Measuring efficiency: methodologies and overview of the literature

The object of this chapter is to briefly describe the specific empirical methods applied in this paper to measure efficiency and to survey the literature more directly related to the analysis of public expenditure efficiency. Empirical and theoretical measures of efficiency are based on ratios of observed output levels to the maximum that could have been obtained given the inputs utilized. This maximum constitutes the efficient frontier which will be the benchmark for measuring the relative efficiency of the observations. There are multiple techniques to estimate this frontier, surveyed recently by Murillo-Zamorano (2004), and the methods have been recently applied to examine the efficiency of public spending in several counties. These are the topics of the next two sections.

2.1 Methods for measuring efficiency

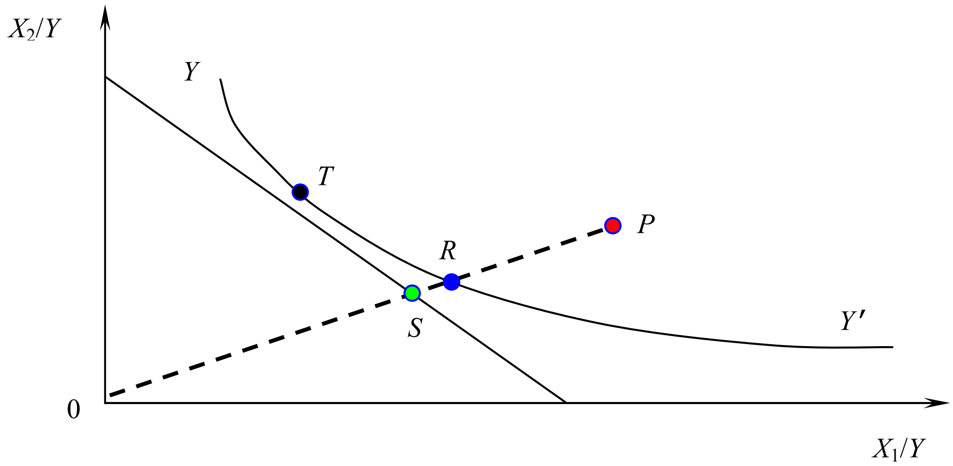
The origin of the modern discussion of efficiency measurement dates back to Farell (1957), who identified two different ways in which productive agents could be inefficient: one, they could use more inputs than technically required to obtain a given level of output, or two, they could use a sub-optimal input combination given the input prices and their marginal productivities. The first type of inefficiency is termed technical inefficiency while the second one is known as allocative inefficiency.

These two types of inefficiency can be represented graphically by means of the unit isoquant curve in Figure 1. The set of minimum inputs required for a unit of output lies on the isoquant curve *YY'*. An agent's input-output combination defined by bundle *P* produces one unit of output using input quantities X_1 and X_2. Since the same output can be achieved by consuming less of both inputs along the radial back to bundle *R*, the segment *RP* represents the inefficiency in resource utilization. The technical efficiency (*TE*), input-oriented, is therefore defined as $TE = OR/OP$. Furthermore, the producer could achieve additional cost reduction by choosing a different input combination. The least cost combination of inputs that produces one unit of output is given by point *T*, where the marginal rate of technical substitution is equal to the input price ratio. To achieve this cost level implicit in the optimal combination of inputs, input use needs to be contracted to bundle *S*. The input allocative efficiency (*AE*) is defined as $AE = OS/OR$.

The focus of this paper is measuring technical efficiency, given the lack of comparable input prices across the countries. This concept of efficiency is narrower

Figure 1

Technical and Allocative Inefficiency



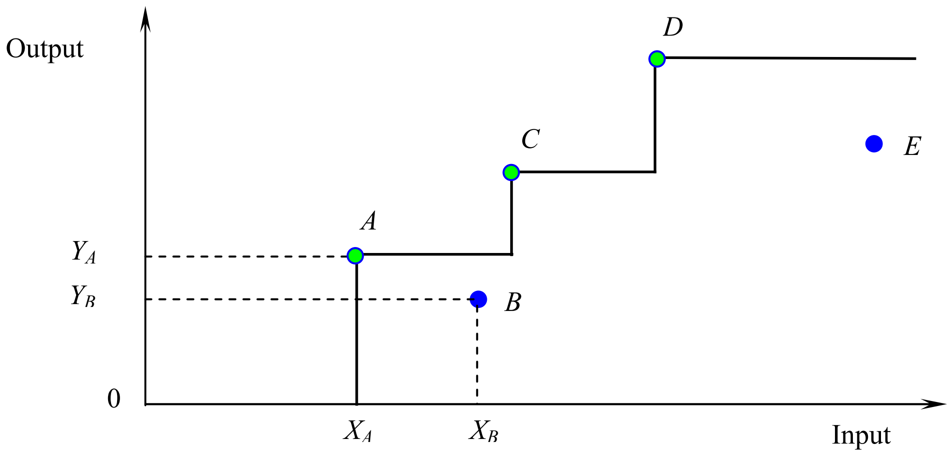
than the one implicit in social welfare analysis. That is, countries may be producing the wrong output very efficiently (at low cost). We abstract from this consideration (discussed by Tanzi, 2004), focusing on the narrow concept of efficiency.

Numerous techniques have been developed over the past decades to tackle the empirical problem of estimating the unknown and unobservable efficient frontier (in this case the isoquant *YY*"). These may be classified using several taxonomies. The two most widely used catalog methods into parametric or non-parametric, and into stochastic or deterministic. The parametric approach assumes a specific functional form for the relationship between the inputs and the outputs as well as for the inefficiency term incorporated in the deviation of the observed values from the frontier. The non-parametric approach calculates the frontier directly from the data without imposing specific functional restrictions. The first approach is based on econometric methods, while the second one uses mathematical programming techniques. The deterministic approach considers all deviations from the frontier explained by inefficiency, while the stochastic focus considers those deviations a combination of inefficiency and random shocks outside the control of the decision maker.

This paper uses non-parametric methods to avoid assuming specific functional forms for the relationship between inputs and outputs or for the inefficiency terms. A companion paper will explore the parametric approach, along the lines proposed by Greene (2003). The remainder of the section briefly describes the two methods: the Free Disposable Hull (FDH) and the Data Envelopment Analysis (DEA).

Figure 2

Free Disposal Hull (FDH) Production Possibility Frontier

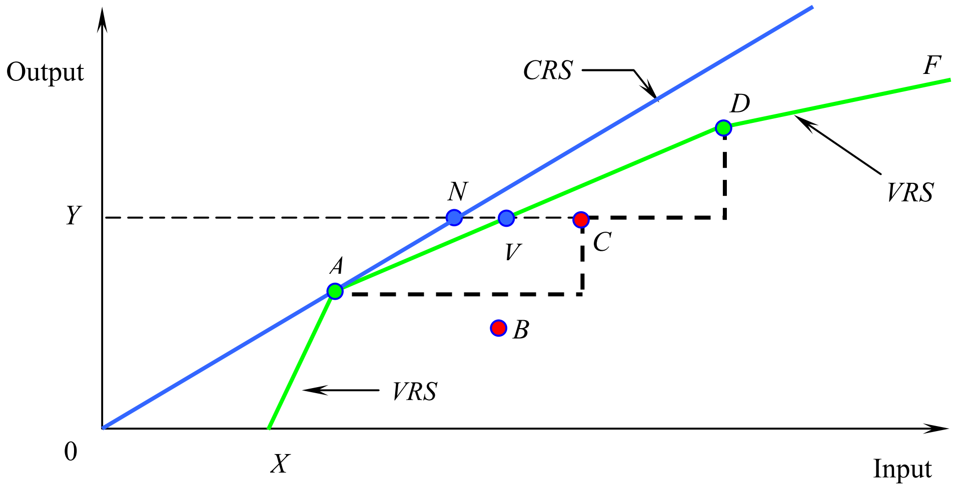


The FDH method imposes the least amount of restrictions on the data, as it only assumes free disposability of resources. Figure 2 illustrates the single input/single output case of FDH production possibility frontier. Countries *A* and *B* use input X_A and X_B to produce outputs Y_A and Y_B, respectively. The *input efficiency* score for country *B* is defined as the quotient X_A/X_B. The *output efficiency* score is given by the quotient Y_B/Y_A. A score of one implies that the country is on the frontier. An input efficiency score of 0.75 indicates that this particular country uses inputs in excess of the most efficient producer to achieve the same output level. An output efficiency score of 0.75 indicates that the inefficient producer attains 75 per cent of the output obtained by the most efficient producer with the same input intake. Multiple input and output efficiency tests can be defined in an analogous way.

The second approach, Data Envelopment Analysis (DEA), assumes that linear combinations of the observed input-output bundles are feasible. Hence it assumes convexity of the production set to construct an envelope around the observed combinations. Figure 3 illustrates the single input/single output DEA production possibility frontier. In contrast to the vertical step-ups of FDH frontier, DEA frontier is a piecewise linear locus connecting all the efficient decision-making units (DMU). The feasibility assumption, displayed by the piecewise linearity, implies that the efficiency of *C*, for instance, is not only ranked against the real performers *A* and *D*, called the peers of *C* in the literature, but also evaluated with a virtual decision maker, *V*, which employs a weighted collection of *A* and *D* inputs to yield a virtual output. DMU C, which would have been considered to be efficient by FDH, is now lying below the variable returns to scale (*VRS*, further defined below) efficiency frontier, *XADF*, by DEA ranking. This example shows that FDH tends to assign

Figure 3

DEA Production Possibility Frontier



efficiency to more DMUs than DEA does. The input-oriented technical efficiency of *C* is now defined by *TE* = *YV*/*YC*.

If constant returns to scale (*CRS*) characterize the production set, the frontier may be represented by a ray extending from the origin through the efficient DMU (ray *0A*). By this standard, only A would be rated efficient. The important feature of the *XADF* frontier is that this frontier reflects variable returns to scale. The segment *XA* reflects locally increasing returns to scale (*IRS*), that is, an increase in the inputs results in a greater than proportionate increase in output. Segments *AD* and *DF* reflect decreasing returns to scale. It is worth noticing that constant returns to scale technical efficiency (*CRSTE*) is equal to the product of variable returns to scale technical efficiency (*VRSTE*) and scale efficiency (*SE*). Accordingly, *DMU D* is technically efficient but scale inefficient, while *DMU C* is neither technically efficient nor scale efficient. The scale efficiency of *C* is calculated as *YN*/*YV*. For more detailed exploration of returns to scale, readers are referred to Charnes, Cooper, and Rhodes (1978) and Banker, Charnes and Cooper (1984), among others.

The limitations of the non-parametric method derive mostly from the sensitivity of the results to sampling variability, to the quality of the data and to the presence of outliers. This has led recent literature to explore the relationship between statistical analysis and non-parametric methods (Simar and Wilson, 2000). Some solutions have been advanced. For instance, confidence intervals for the efficiency scores can be estimated using asymptotic theory in the single input case (for input efficiency estimators) or single output (in the output efficiency) case, given these are shown to be maximum likelihood estimators (Banker, 1993 and Goskpoff, 1996).

For multiple input/output cases the distribution of the efficiency estimators is unknown or quite complicated and analysts recommend constructing the empirical distribution of the scores by means of bootstrapping methods (Simar and Wilson, 2000). Other solutions to the outlier or noisy data consist in constructing a frontier that does not envelop all the data point, building an expected minimum input function or expected maximum output functions (Cazals, Florens and Simar, 2002, and Wheelock and Wilson, 2003). Another limitation of the method, at least in the context in which we will apply it, is the inadequate treatment of dynamics, given the lag between input consumption (public expenditure) and output production (health and education outcomes).

2.2 *Overview of precursor papers*

There is abundant literature measuring productive efficiency of diverse types of decision making units. For instance, there are papers measuring efficiency of museums (Bishop and Brand, 2003), container terminals (Cullinane and Song, 2003), electric generation plants (Cherchye and Post, 2001), banks (Wheelock and Wilson, 2003), schools (Worthington, 2001) and hospitals (Bergess and Wilson, 1998), among others. Few papers, however, analyze aggregate public sector spending efficiency using cross-country data. These are the direct precursors of this paper and are the focus of this section's survey.

Gupta and Verhoeven (2001) employ the input-oriented FDH approach to assess the efficiency of government spending on education and health in thirty-seven African countries in 1984-1995. Using several output indicators for health and education, they construct efficiency frontiers for each of the indicators and for each of the time periods they considered. That is, they used a single input/single output for each time period. They find that, on average, African countries are inefficient in providing education and health services relative to both Asian and the Western Hemisphere countries. They also report, however, an increase in the productivity of spending through time, as they document outward shifts in the efficiency frontier. Finally the authors report a negative relationship between the input efficiency scores and the level of public spending, which leads them to conclude that higher educational attainment and health output requires efficiency improvement more than increased budgetary allocations.

Evans and Tandon (2000) adopt a parametric approach to measure efficiency of national health systems for the World Health Organization, by estimating a fixed effects panel of 191 countries for the period 1993-97. Health output was measured by the disability-adjusted life expectancy (DALE) index, while health expenditures (public and private aggregated) and the average years of schooling of the adult population were considered as inputs. The output efficiency score is defined as the ratio of actual performance above the potential maximum. The authors also introduce the square of the inputs (average years of schooling and expenditure), arguing it's a second-order Taylor-series approximation to an unknown functional form. The fact that the quadratic terms are significant may be an indication of the importance of non-linearity, but may also reflect neglected dynamics or heterogeneity in the sample (Haque, Pesaran and Sharma, 1999), given that both developed and developing nations were included. An interesting contribution of the paper is a construction of a confidence interval for the efficiency estimates through a Monte Carlo procedure. These authors document a positive relationship between their efficiency scores and the level of spending. The more efficient health systems are those of Oman, Chile and Costa Rica. The more inefficient countries are all African: Zimabawe, Zambia, Namibia, Botswana, Malawi and Lesotho.

Jarasuriya and Woodon (2002) also adopt a parametric approach to estimate efficiency of health and education provision in a sample of developing countries. The authors estimate the efficiency frontier by econometric methods. These authors consider separately an educational attainment indicator (net primary enrolment) and a health output indicator (life expectancy) and estimate a functional linear relationship between these output indicators and three inputs: per capita GDP, spending per capita, and the adult literacy rate. Using a panel of 76 countries for the period 1990 to 1998, they found no relationship between expenditure and the educational or health output variables when they include the per capita GDP. This led the authors to conclude that spending more is not a guarantee to obtain better education or health results. The authors do not point at the correlation between the two variables as a possible cause of this problem, which we discuss in the next section. The countries with the lowest efficiency in health indicators are all African (Malawi, Zambia, Mozambique, Ethiopia) as well as in education attainment (Ethiopia, Niger, Burkina Faso).

The authors go further by attempting to explain the cross-country variation in efficiency and find that the degree of urbanization and the quality of bureaucracy are the most relevant variables. To capture possible non-linearity, the authors introduce these variables squared. This stage of their work poses several problems. First, it is possible that the (non-linear) quadratic terms reflect heterogeneity across countries and dynamics across time. As shown by Haque, Pesaran and Sharma (1999), this would produce inconsistent estimates. Second, the authors do not adjust for the fact that the dependent variable (efficiency scores) is censored, given that it can adopt only values between zero and one. And third, the authors do not consider the serial correlation of the efficiency scores (Simar and Wilson, 2004).

Greene (2003a) combines the previous two papers in the sense he concentrated on health efficiency only using the WHO panel data and explained inefficiency scores variation across the sample of counties. Greene’s stochastic frontier estimation is much more general and flexible, as it allows for time variation of the coefficients and heterogeneity in the countries’ sensitivity to the explanatory variables. The author first estimates a health production function using expenditure (public and private together) and education as inputs, and then explains inefficiency with a set of explanatory variables of which the only significant ones are the income inequality measure, GDP per capita and a dummy variable for tropical location.

Afonso, Schuknecht and Tanzi (2003) examine the efficiency of public spending using a non-parametric approach. First, they construct composite indicators of public sector performance for 23 OECD countries, using variables that capture quality of administrative functions, educational and health attainment, and the quality of infrastructure. Taking the performance indicator as the output, and total public spending as the input, they perform single input/single output FDH to rank the expenditure efficiency of the sample. Their results show that countries with small public sectors exhibit the highest overall performance.

Afonso and St. Aubyn (2004) address the efficiency of expenditure in education and health for a sample of OECD countries applying both DEA and FDH. This paper presents detailed results by comparing input-oriented and output-oriented efficiency measurements. The small overlap of the samples limits the comparability of these results with those presented in the next section. An apparently strange result, reported in earlier drafts of the paper, was the inclusion of Mexico as one of the benchmark countries (on the efficiency frontier). The result is strange given that the sample is the OECD countries, and it counterintuitive. This is the result of Mexico having very low spending and low education attainment results, hence it can be considered as the "origin" of the efficiency frontier. The next chapter discusses this topic and reports similar counterintuitive results but for other countries.

3. Empirical results

3.1 Input and output indicators: description, assumptions and limitations

Cross-country comparisons assume some homogeneity across the world in the production technology of health and education. There are two particular aspects in which the homogeneity assumption is important. First, the comparison assumes that there is a small number of factors of production that are the same across countries. Any omission of an important factor will yield as a result a high efficiency ranking of the country that uses more of the omitted input. Second, the comparison requires that the quality of the inputs is more or less the same, with the efficiency scores biased in favor of countries where the quality is of higher grade.

Factor heterogeneity will not be a problem, as long as it is evenly distributed across countries. It will be problematic if there are differences between countries in the average quality of a factor (Farrell, 1957). The exercise that we present suffers from this limitation, given that the main input in both production technologies is used more intensively in richer countries (with higher per capita GDP). The main input is public spending per capita on education and health measured in constant 1995 US dollars in PPP terms. A clear positive association between this variable and per capita GDP can be verified (Figures 4 and 5).

This positive association between expenditure and the level of economic development (as measured by per capita GDP) may be explained by several reasons. One of them could be the Balassa-Samuelson effect, according to which price levels in wealthier countries tend to be higher than in poorer countries.[1] This applies to both final goods and factor prices. Thus price of the same service (health or education, for instance) will be higher in the country with higher GDP. Similarly, wages in the relatively richer counties are higher, given the higher marginal productivity of labor, which will tend to increase costs, especially in labor-intensive activities as health and education.

Figure 4 can be interpreted as evidence of the validity of Wagner's hypothesis at the cross-country level. This hypothesis, postulates that there is a tendency for governments to increase their activities as economic activity increases. Since 1890 Wagner postulated that economic development implied rising complexities that required more governmental activity, or that the elasticity of demand for publicly provided services, in particular education was greater than one. This hypothesis has been tested econometrically (Chang, 2002) in time series and cross-country settings, showing that this is nothing particular of the series used for the present study.

Previous studies that measured the efficiency of public spending recognized the positive association and suggested alternative solutions. One possibility is to split the sample by groups of countries (Gupta and Verhoeven, 2001). We follow this approach by excluding the industrialized nations from the sample, and by presenting most of the results clustered regionally (Africa (AFR), East Asia and Pacific (EAP), Latin America and Caribbean (LAC), Middle East and North Africa (MNA) and South Asia (SAS)). A second alternative incorporates directly the per capita GDP as a factor of production, jointly with expenditure and other inputs (Jarasuriya and Woodon, 2002). The problem with this approach is that it combines variables derived from a production function approach, and hence with clear interpretation, with others (GDP per capita) that are difficult to interpret from any viewpoint. When the two types of variables are combined, their effects cannot be disentangled.

A third option consists in using as an input the orthogonal component of public expenditure to GDP.[2] We scored the efficiency using as input both the original expenditure variable and the orthogonalized variable. The goodness-of-fit of each model was gauged based on the frequency distribution of the inefficiency measures, as suggested by Farrell (1957) and Varian (1990). Comparing the efficiency distributions (Figure 5) it is clear that the orthogonalized expenditure version produces distributions that are not skewed towards extreme inefficient outcomes. On this basis, the paper considered the orthogonal component of expenditure on health and education.

---

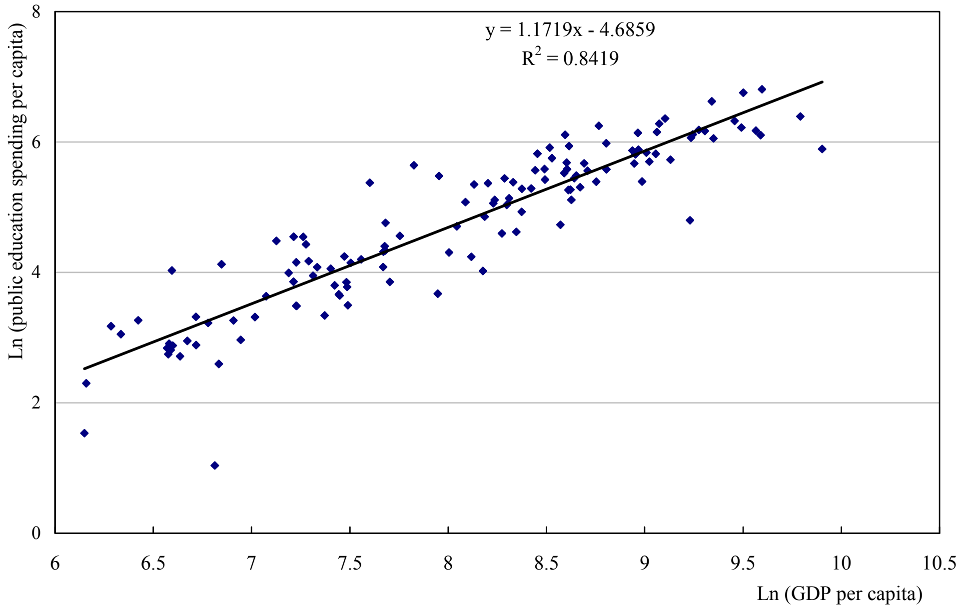
1 The Balassa-Samuelson effect refers to the fact that price levels are higher in richer countries than in poorer countries. It can be shown that relative wages and relative prices are a function of the marginal productivity of labor in the traded goods. Given higher capital abundance in the richer countries, the productivity of labor tends to be higher in these countries, and hence will be wages and prices.

2 The orthogonalized expenditure variable is the residual of the linear regression between pubic expenditure and GDP per capita. Since residuals may take positive and negative values, the variable was right-shifted to avoid negative values to facilitate graphical presentation of the frontiers.

Figure 4

Public Expenditure and GDP (Both per capita and in Logs)

Education Spending vs. GDP per capita



Health Spending vs. GDP per capita

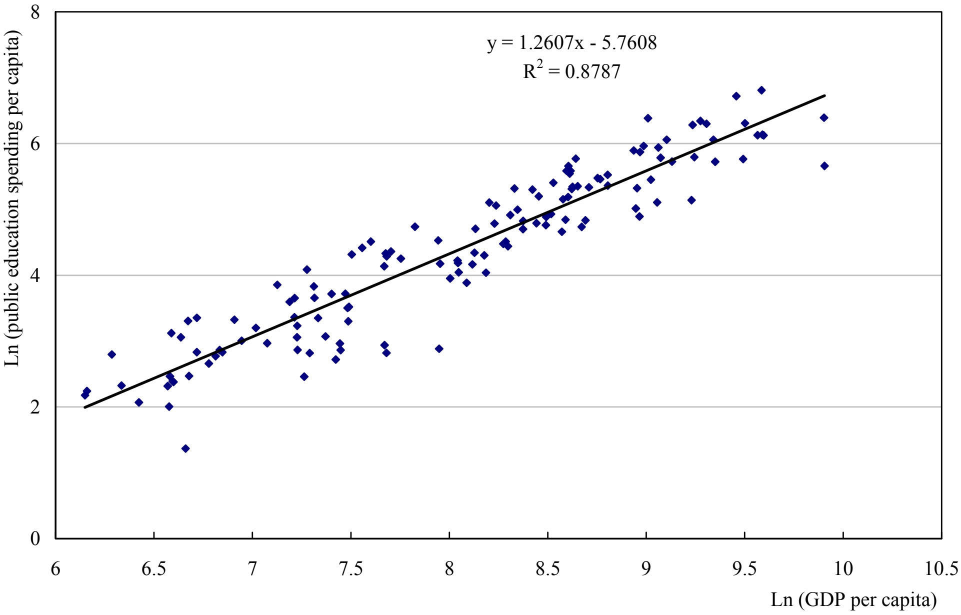
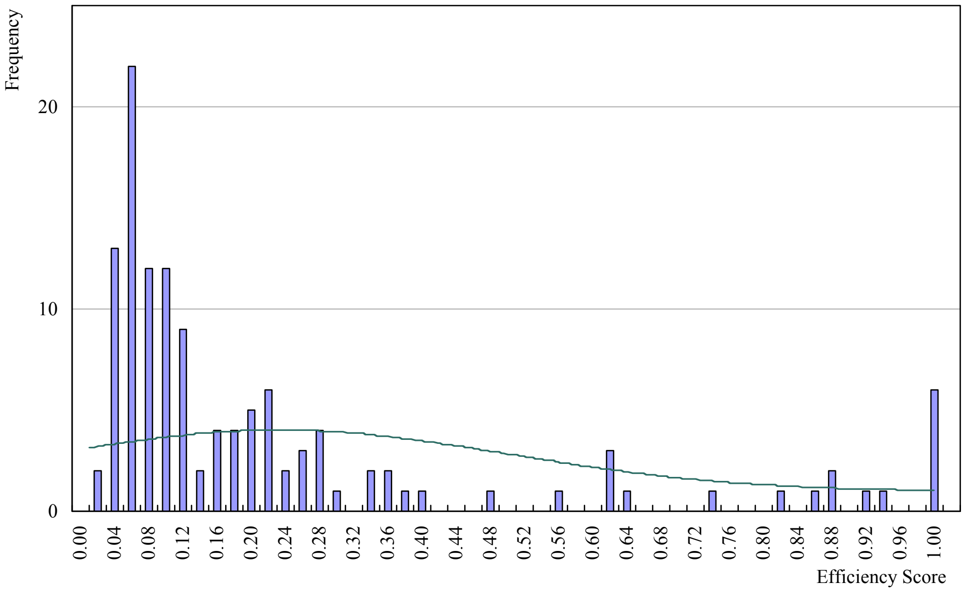


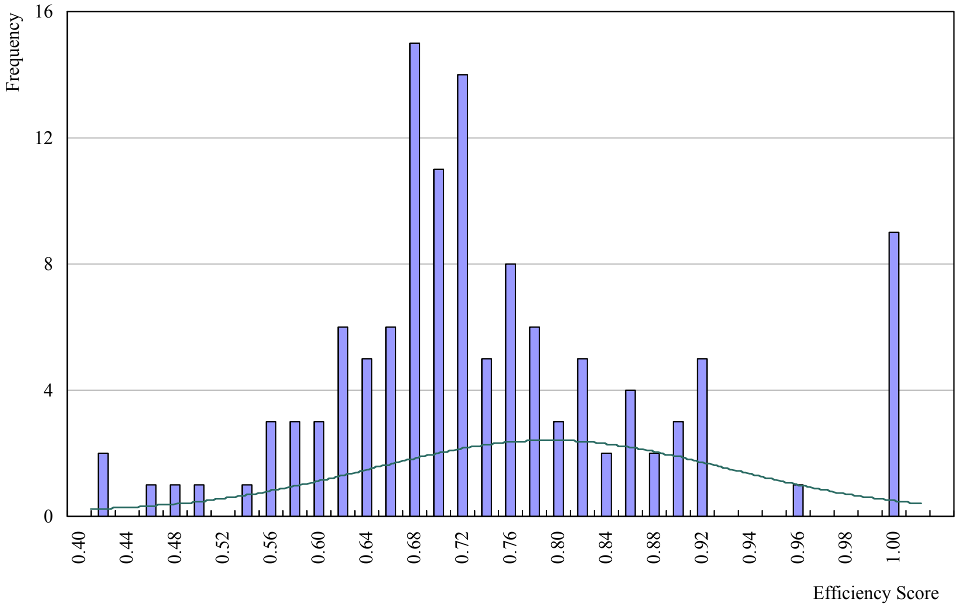
Figure 5

Density of Efficiency Scores – Gross Primary School Enrolment

Unorthogonalized Public Expenditure



Orthogonalized Public Expenditure



This paper uses nine indicators of education output and four indicators of health output.[3] The education indicators are: primary school enrolment (gross and net), secondary school enrolment (gross and net), literacy of youth, average years of school, first level complete, second level complete, and learning scores. Though the ideal educational output indicator are comparable learning scores, international assessments are based on samples mostly composed of developed nations, limiting the applicability to the present paper. However, Crouch and Fasih (2004) recently combined several international assessments to obtain a larger sample of comparable results.[4] Unfortunately they only do it for one period. The correlation between the learning scores and other output variables is high (.81 with net secondary school enrolment and .76 with average years of school), as shown in Figure 6.[5] The health output indicators are: life expectancy at birth, immunization (DPT[6] and measles), and the disability-adjusted life expectancy (DALE).

The cross-country comparisons with this set of indicators assume some form of data homogeneity, which might be problematic given the diversity of counties in the sample considered. Even for a more homogeneous group of countries, such as the OECD, there is call for caution when comparing expenditure levels in member countries (Jounard *et al.*, 2003). There is very little to do to overcome this limitation, except subdivide the sample into different groups. Probably a regional aggregation can be useful, but even at that level there may be extreme heterogeneity.

Other four limitations of the analysis arising from the particular data sources are: first, the level of aggregation. The paper uses aggregate public spending on health and education, while using disaggregate measures of output, such as. primary enrolment or secondary enrolment. Ideally, the input should be use separately public spending in primary and secondary education. Similarly, health care spending could be disaggregated into primary care level care and secondary level. The data can be disaggregated even further, by analyzing efficiency at the school or hospital levels. Second, there are omitted factors of production. This is especially true in education, as the paper did not consider private spending due to data constraints for developing nations. If this factor were used more intensively in a particular group of countries, then the efficiency scores (reported in the next section) would be biased favoring efficiency in that group.

3 The data sources are: the World Bank World Development Indicators (WDI), Barro-Lee database, Crouch and Fasih (2004), and the World Health Organization (Mathers *et al.,* 2000).

4 Crouch and Fasih (2004) consider several international tests of learning achievement in math, science and literacy applied at different levels of the school system. The tests are the following: TIMSS (Third International Mathematics and Science Survey), PIRLS (Progress in International Literacy Study), PISA (Program for International Student Assessment), Reading Literacy Study, LLECE (Laboratorio Latinoamericano de Evaluacion de la Calidad de la Educacion), SACMEQ (Southern Africa Consortium for Monitoring of Education Quality), MLA (Monitoring Learning Achievement). Since the tests have different samples, they converted all test scores through iterative comparisons to a single numeraire.

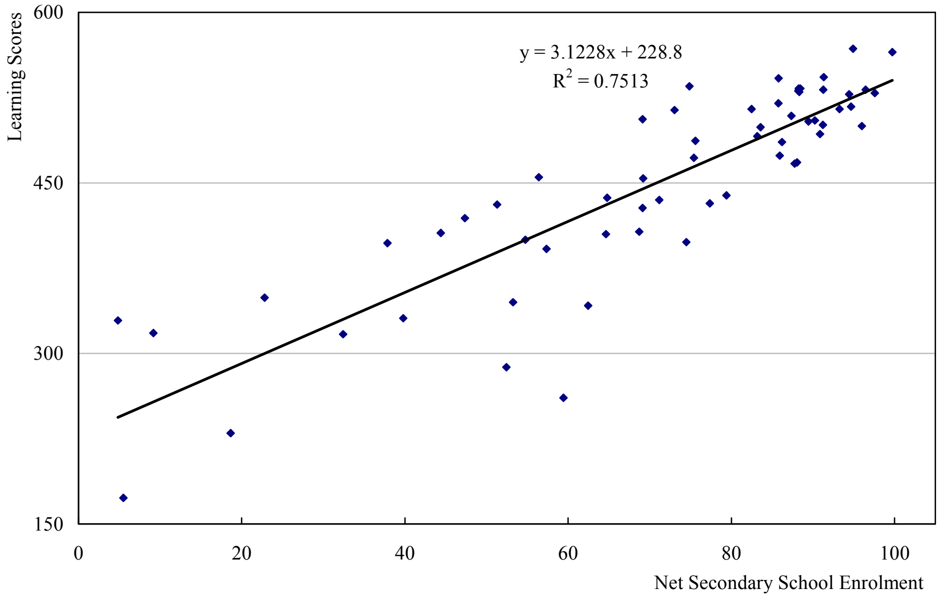
5 The correlation coefficients and Figure 6 exclude developed nations for the Crouch and Fasih (2004) sample.

6 DPT is Diphtheria-Pertussis and Tetanus.

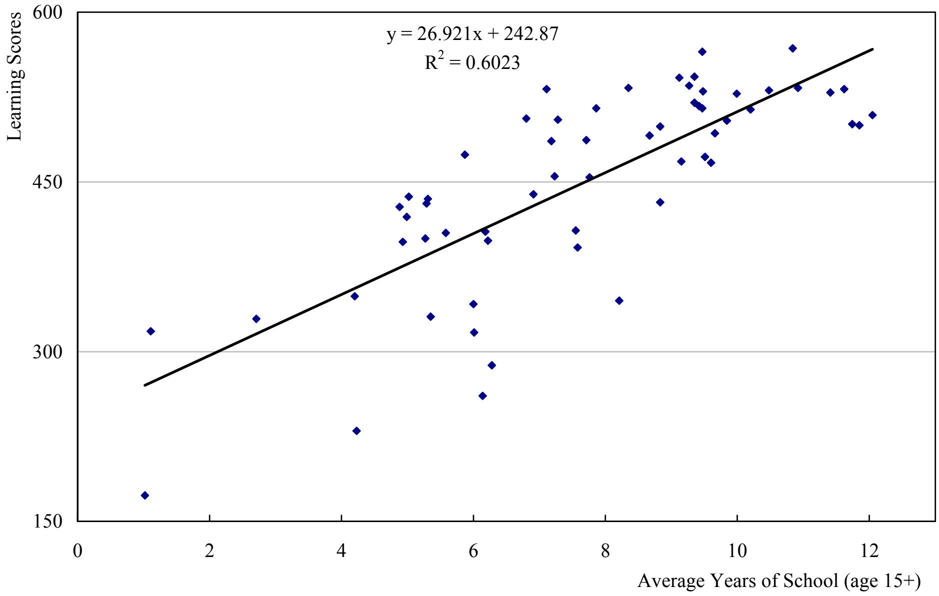
Figure 6

Correlation between Learnings Scores and Other Education Indicators

Correlation between Learning Scores and Net Secondary Enrolment



Correlation between Learning Scores and Average Years of School



Source for the figures on this page: World Bank WDI and Crouch and Fasih (2004).

The third limitation arising from the data is the combination of monetary and non-monetary factors of production. The paper uses together with public expenditure, other non-monetary factors of production such as the ratio of teachers to students, in the case of education, or literacy of adults in the case of health and education. Other factors of production that could have been used were the physical number of teaching hours (in education) or the number of doctors or in-patient beds, as Afonso and St. Aubyn did for the OECD countries. However, inexistent data for a large number of developing countries constrained the options. Fourth, data availability constrained a better differentiation between outputs and outcomes. For instance, most of the indicators of education, such as completion and enrolment rates do not measure how much learning is taking place in a particular country. In education, this paper advances by considering the learning scores as one of the indicators. In health, other outcomes such as the number of sick-day leaves or the number of missed-school days because of health-related causes could be better reflections of outcomes.

3.2 *Single input/output results*

3.2.1 *FDH and DEA analysis: education*

Figures 7a-c show both FDH and DEA estimation of the efficiency frontier for three of the nine output indicators: gross primary school enrolment, first level complete and learning scores.

Figure 7d illustrates the efficiency frontier for the learning scores if the developed countries are included in the sample, demonstrating the sensitivity of the results to the sample definition. This fact is particularly acute in the case of learning scores which capture the quality of education dimension that no other indicator captures. While in the sample of developing countries Chile, Hungary and the Czech Republic are on the frontier; once the developed nations are included they appear as inefficient.[7]

Several results may be highlighted:

a) In general, the rankings are robust to the output indicator selected. This can be can be verified by the Spearman rank-correlation coefficient: all are positive, significant and high. The range oscillates from a minimum of .53 to a maximum of .94, with the mean of .70. This result implies that countries appearing as efficient (or inefficient) according to one indicator, are ranked similarly when other output indicator is used.

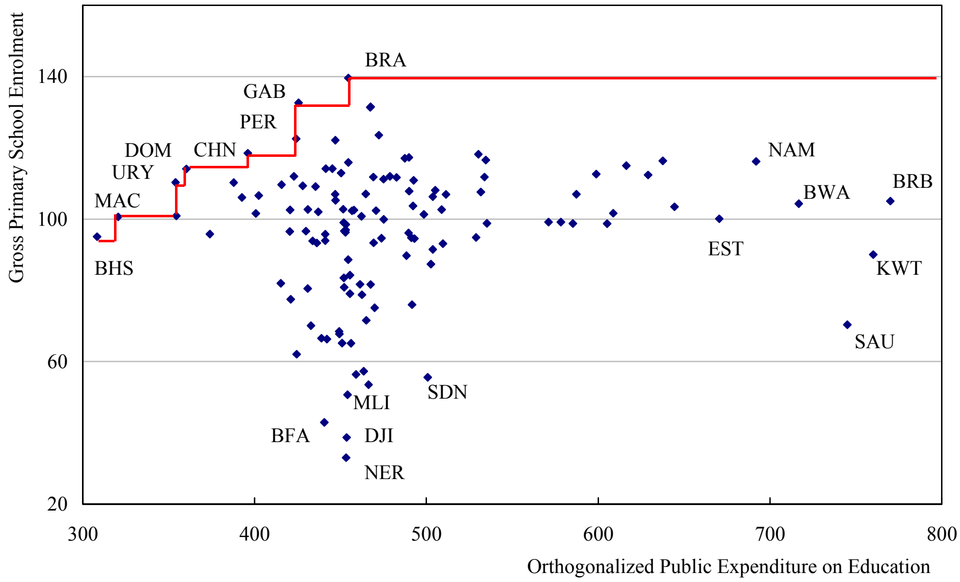
b) Despite the orthogonalization by GDP, the relatively rich countries tend to be in the less efficient group, *i.e.* countries with higher per capita GDP spend more than other countries in attaining similar education outcomes. Higher spending

[7] The frontier depicted in Figure 7d excludes Japan, Korea, Ireland and Belgium to facilitate comparisons with the frontier without developed nations.

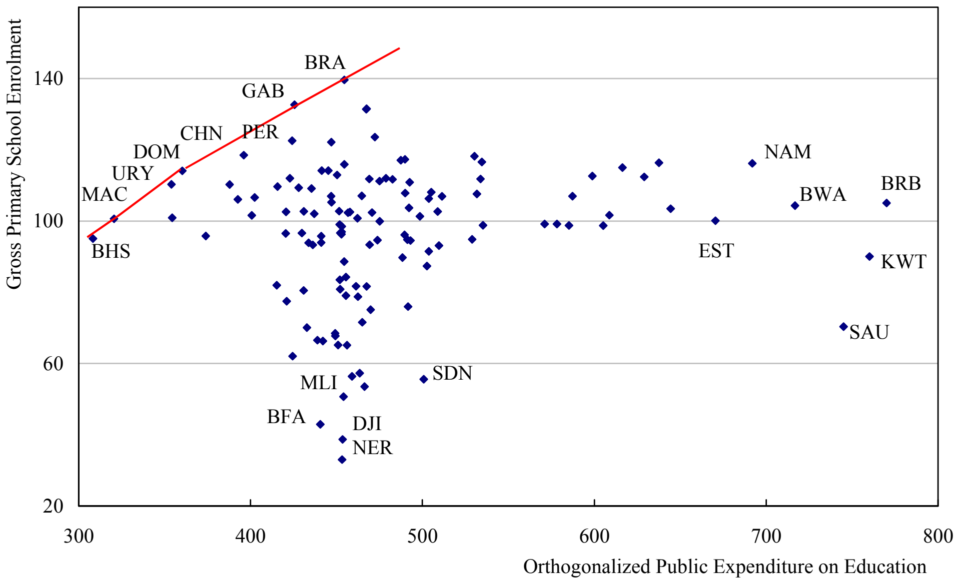
Figure 7

Education Efficiency Frontier: Single Input and Single Output

a.1: Gross Primary School Enrolment vs. Education Expenditure (Free Disposable Hull, FDH)



a.2: Gross Primary School Enrolment vs. Education Expenditure (Data Envelopment Analysis, DEA)

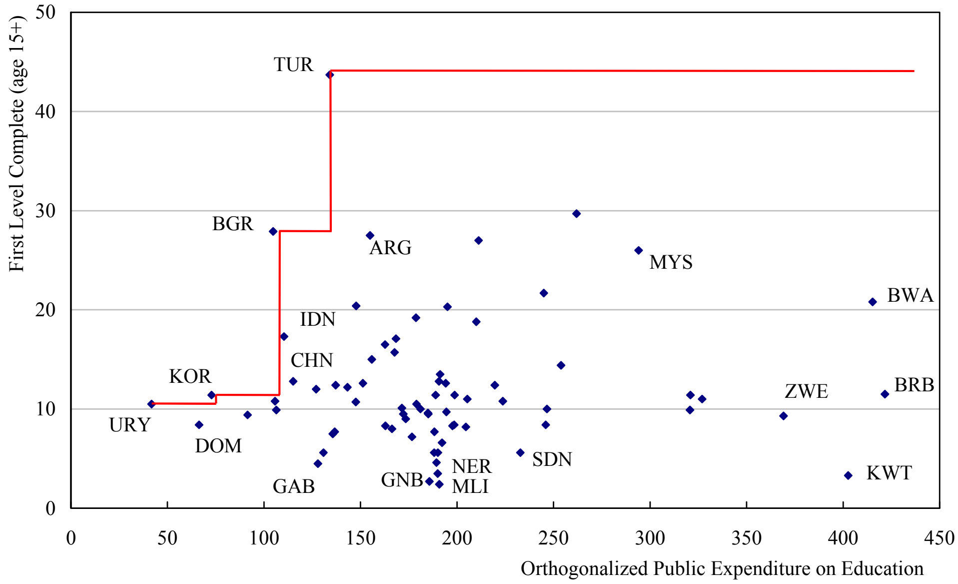


Source for the figures on this page: World Bank WDI.

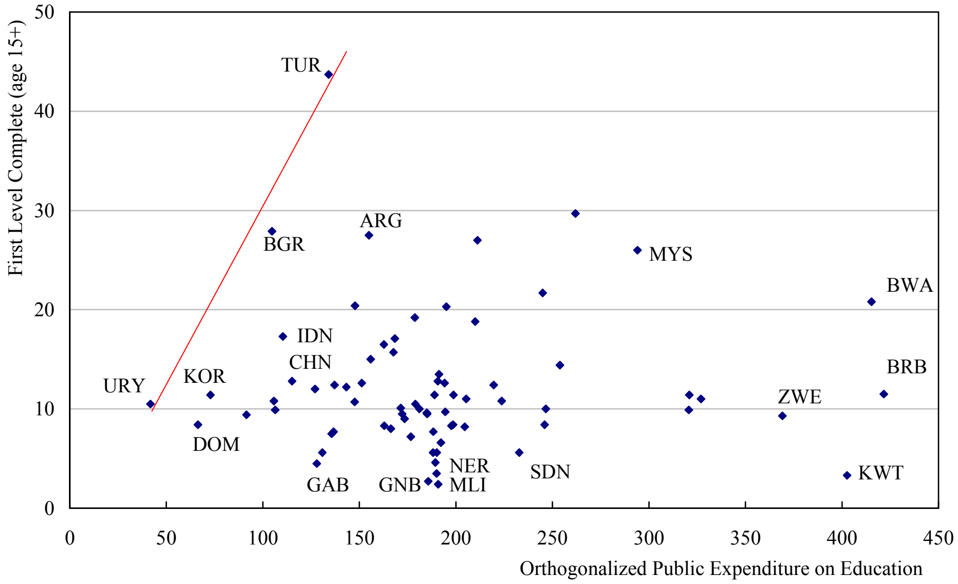
Figure 7 (continued)

Education Efficiency Frontier: Single Input and Single Output

b.1: First Level Complete vs. Education Expenditure (Free Disposable Hull, FDH)



b.2: First Level Complete vs. Education Expenditure (Data Envelopment Analysis, DEA)

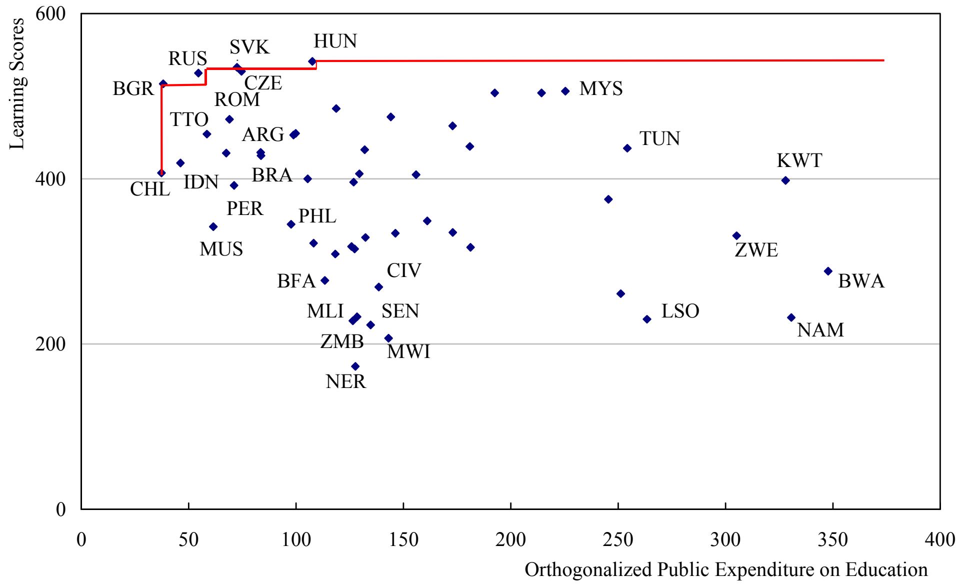


Source for the figures on this page: World Bank WDI, Barro-Lee database.

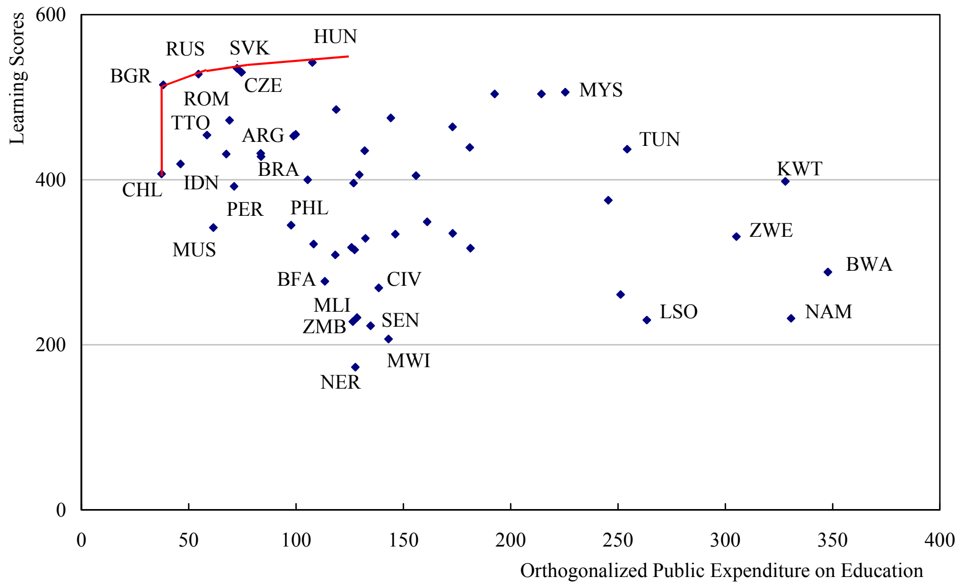
Figure 7 (continued)

Education Efficiency Frontier: Single Input and Single Output

c.1: Learning Scores vs. Education Expenditure (Free Disposable Hull, FDH)



c.2: Learning Scores vs. Education Expenditure (Data Envelopment Analysis, DEA)

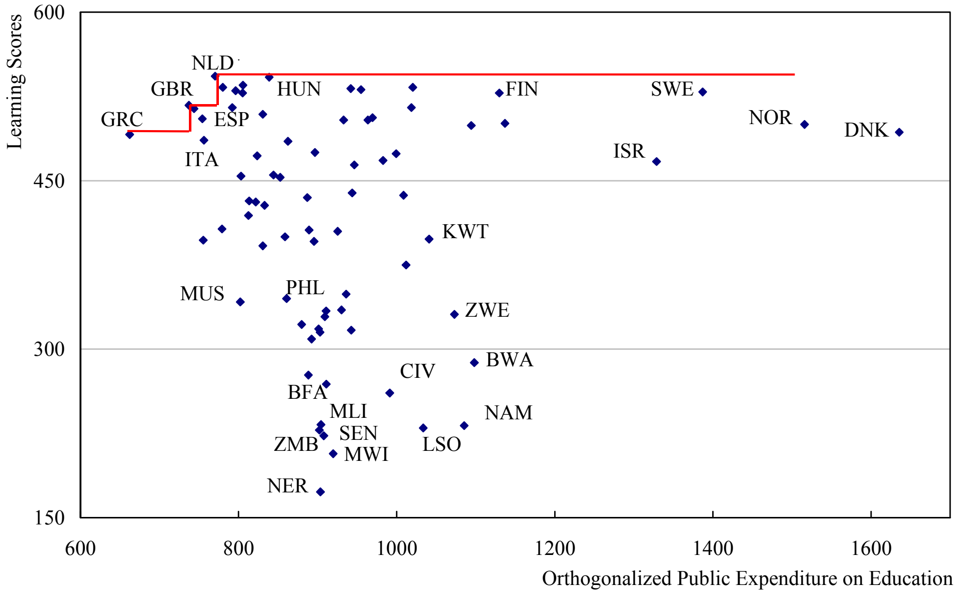


Source for the figures on this page: World Bank WDI and Crouch and Fasih (2004).

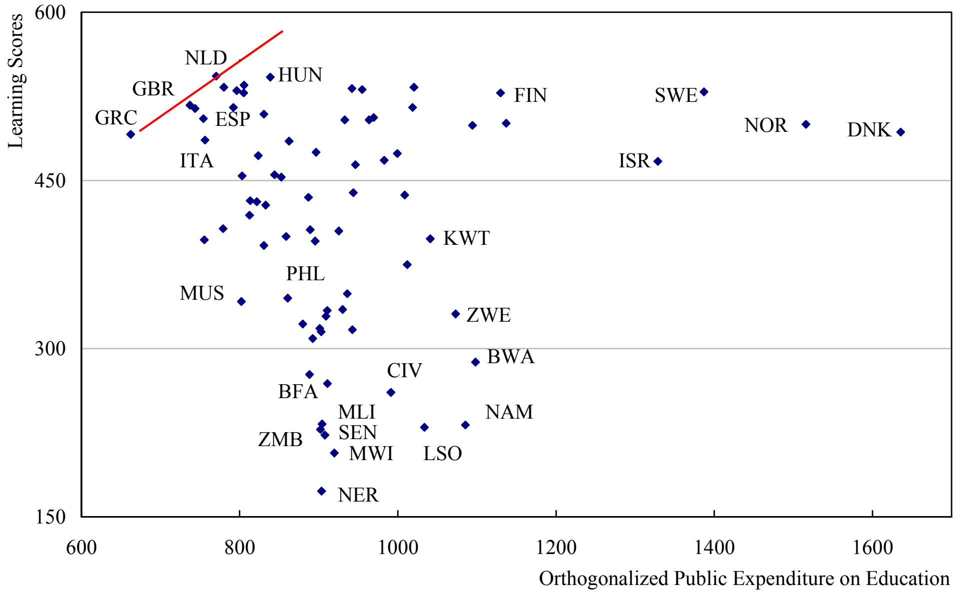
Figure 7 (continued)

Education Efficiency Frontier: Single Input and Single Output

d.1: Learning Scores vs. Education Expenditure (Free Disposable Hull, FDH)



d.2: Learning Scores vs. Education Expenditure (Data Envelopment Analysis, DEA)



Source for the figures on this page: World Bank WDI and Crouch and Fasih (2004).

may reflect the higher cost of tertiary education. This is one factor that may help explain the stand-out of Estonia, Latvia, and Poland. Oil-rich countries, such as Kuwait and Saudi Arabia, tend to be in the group of relatively more inefficient producers.

c) Another group of relatively inefficient producers are those with “average” expenditure levels but extremely low education attainment. Among those are mostly African counties (Angola, Niger, Burkina Faso, Sudan, and Ethiopia), some Middle Eastern countries (Djibouti, and Yemen) and South Asia (Bangladesh and Pakistan).

d) Output efficiency rankings also vary with the selected output indicators. The Spearman correlation coefficient of the output efficiency scores shows that these are robust to the selected indicator, though the mean of the correlation coefficients is lower (.52) and the range is somewhat higher (.30 to .95) than those registered in the input efficiency rankings.

e) In an attempt to identify clusters of more efficient countries and more inefficient countries, the top (and bottom) 10 per cent of the efficiency ranking were selected for each of the indicators. If a country appeared in the efficient (inefficient) tail in three or more of the indicators, it was included in Table 1.

f) This clustering exercise reveals (Table 1) a group of African countries as the most inefficient. Two oil-rich countries are included in this group as well. Among the more efficient group of countries we consistently find Uruguay, Korea, Bahamas, and Bahrain. Explaining why these particular sets of countries appear in each cluster requires more in-depth analysis. The last section of this paper attempts to associate efficiency results with some explanatory variables.

g) To grasp the order of magnitudes of the deviations from the efficiency frontier, we computed an average for all indicators for the inefficient countries. The input efficiency estimations indicate that the most inefficient decile could reach the same educational attainment levels by spending approximately 50 per cent less. The output efficiency estimators indicate that, on average, with the expenditure level this group could reach educational attainment levels four times as high.

h) It is critical to note that even if a country appears as efficient, there might still be a significant discrepancy between the observed output level and the desired or target output level. For instance, Bahamas, Bahrain, Dominican Republic and Guatemala appear as efficient countries on the efficiency frontier or very close to it (Figure 7 a.1). However, these countries are still far away from where Gabon or Brazil are, and could consider desirable to achieve those target enrolment rates. Both Guatemala and Dominican Republic spend 2 per cent of GDP on education but have (net) secondary enrolment rates below 40 per cent. And net primary enrolment is about 80 per cent. It would be difficult to argue that that is a desirable outcome, though it is an efficient one. Similarly, though Chile appears as efficient with learning scores of about 400, the country could still achieve higher learning scores of over 500 points at the cost of additional public spending. The important thing is that the country moves along the efficiency frontier to the higher target output level. Countries can even improve efficiency

Table 1

Education Attainment – Single Input/Single Output

	Input-Efficient	**Output-Efficient**
More efficient	Uruguay, Korea, Dominican Republic, Indonesia, Guatemala, China, Bahamas, Bahrain, El Salvador	Uruguay, Korea, Bahrain,
Least efficient	Botswana, South Africa, Kuwait, Tunisia, Lesotho, Barbados, Saudi Arabia, Zimbawe, Namibia, Malaysia, St. Lucia, Jamaica, St. Vincent, Latvia	Niger, Mali, Tanzania, Burkina Faso, Guinea-Bissau, Ethiopia, Guinea, Burundi, Sudan, Sierra Leone, Chad

by exploiting scale economies if they are operating in the increasing returns to scale zone of the production possibility frontier (output levels smaller than that of point A, Figure 3).

i) The regional aggregation of the efficiency scores by each individual output indicator shows that scores are lower when they are input oriented (Table 2) than when they are output oriented (Table 3).[8] This is especially true for ECA. In general, we observe higher efficiency scores when primary enrolment is considered as the output indicator. Scores are lower for secondary enrolment, especially when output-oriented measures are considered. Africa and MNA have similar levels of input-inefficiency: in most cases, both regions use public spending in excess of 35 per cent than the benchmark cases. EAP, ECA, LAC and SAS spend in excess between 20-30 per cent of the benchmark level. The output efficiency scores are lower in Africa.

3.2.2 FDH and DEA analysis: health

This section considers the case of one input (public expenditure on health per capita in PPP terms) and four alternative output indicators: life expectancy at birth, DPT immunization, measles immunization, and the disability-adjusted life expectancy (DALE) index which takes into account both mortality and illness. The efficiency frontiers for each indicator are computed using both the FDH and DEA methodologies. Figures 8 a-d show the efficiency frontier for one indicator.

---

8 The regional aggregation is for illustrative purposes only and was computed as the simple average of the individual country scores obtained for the whole sample. The scores were not computed by constructing separate efficiency frontiers for each region. Hence, they do not reflect the heterogeneity in the individual country scores and possibly do not reflect adequately variations across regions.

Table 2

Educational Attainment: Input Efficiency Scores by Regions across the World Single Input/Single Output

	AFR	EAP	ECA	LAC	MNA	SAS
Gross Primary Enrolment	.69	.74	.67	.74	.65	.75
Net Primary Enrolment	.68	.78	.72	.77	.68	.71
Gross Secondary Enrolment	.65	.69	.67	.69	.63	.70
Net Secondary Enrolment	.64	.71	.71	.69	.64	.72
Average Years of School	.21	.36	.37	.32	.18	.25
First Level Complete	.21	.43	.48	.36	.20	.26
Second Level Complete	.22	.37	.33	.32	.19	.27
Literacy of Youth	.66	.73	.86	.72	.63	.72

Table 3

Educational Attainment: Output Efficiency Scores by Regions across the World Single Input/Single Output

	AFR	EAP	ECA	LAC	MNA	SAS
Gross Primary Enrolment	.62	.79	.72	.82	.67	.72
Net Primary Enrolment	.64	.93	.90	.93	.79	.78
Gross Secondary Enrolment	.23	.50	.70	.61	.54	.39
Net secondary Enrolment	.26	.58	.84	.66	.60	.44
Average Years of School	.32	.63	.79	.60	.53	.38
First Level Complete	.19	.49	.50	.36	.22	.20
Second Level Complete	.09	.37	.38	.24	.26	.22
Literacy of Youth	.72	.95	.99	.94	.88	.66

Several results may be highlighted:

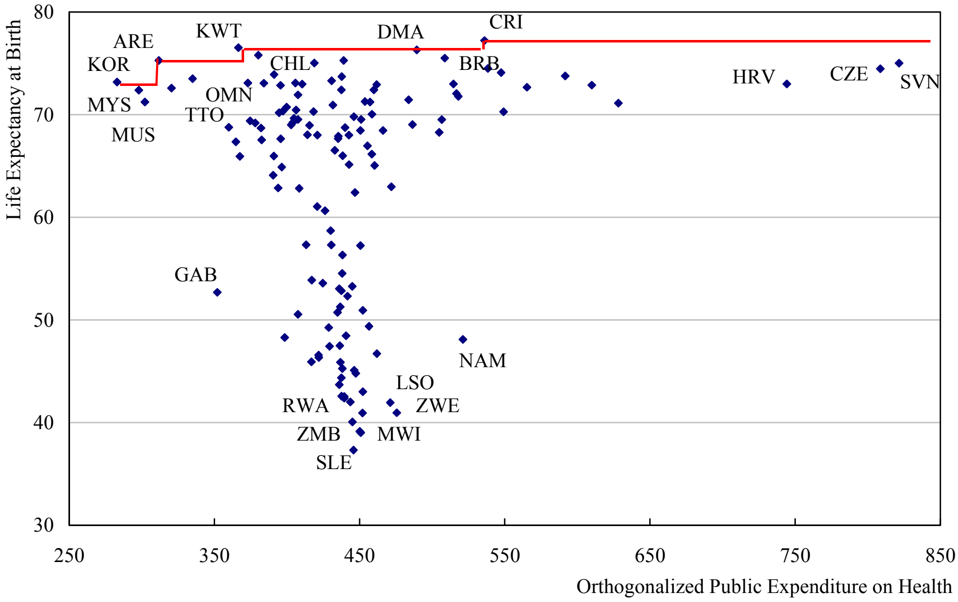
a) The input efficiency scores obtained for each of the output indicators are highly correlated. The Spearman rank-order correlation coefficient oscillates between .66 and .94, with a mean of 0.81. This indicates that the efficiency ranking is very similar regardless of the output indicator being used.

b) Despite the orthogonalization by GDP the relatively rich countries tend to be in the less efficient group. The group of inefficient producers tend to concentrate in

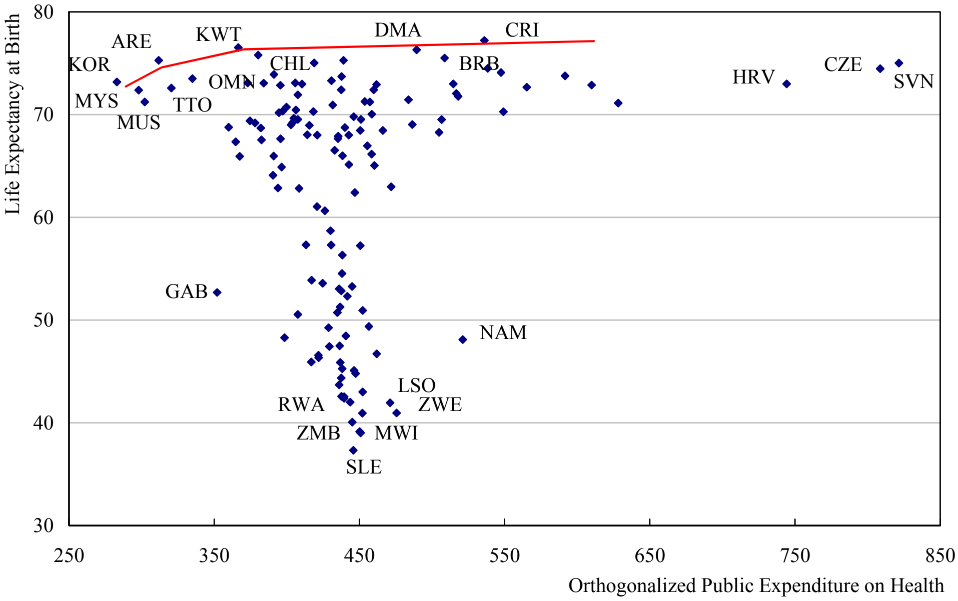
Figure 8

Health Efficiency Frontier – Single Input/Single Output

a.1: Life Expectancy vs. Health Expenditure (Free Disposable Hull, FDH)



a.2: Life Expectancy vs. Health Expenditure (Data Envelopment Analysis, DEA)

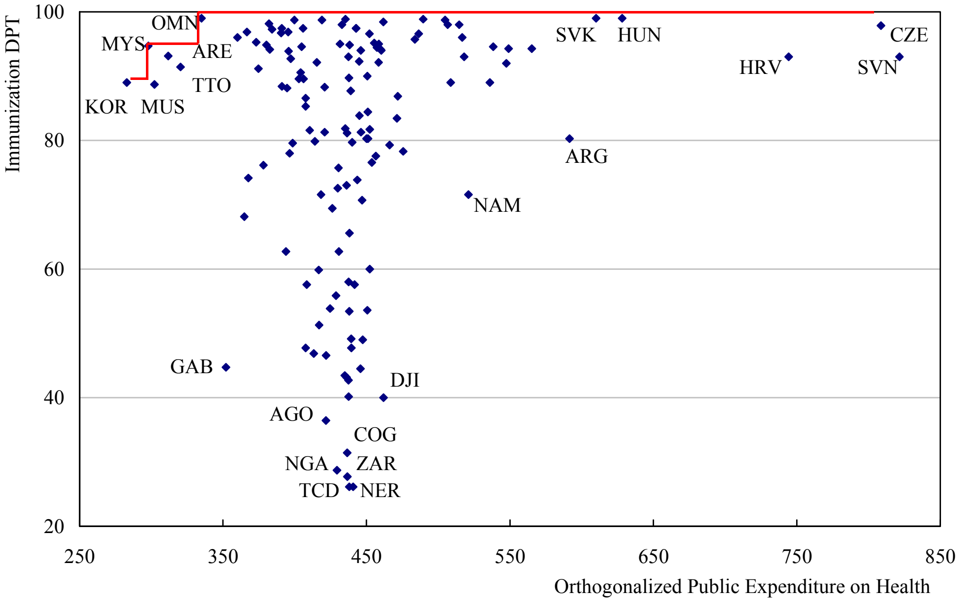


Source for the figures on this page: World Bank WDI.

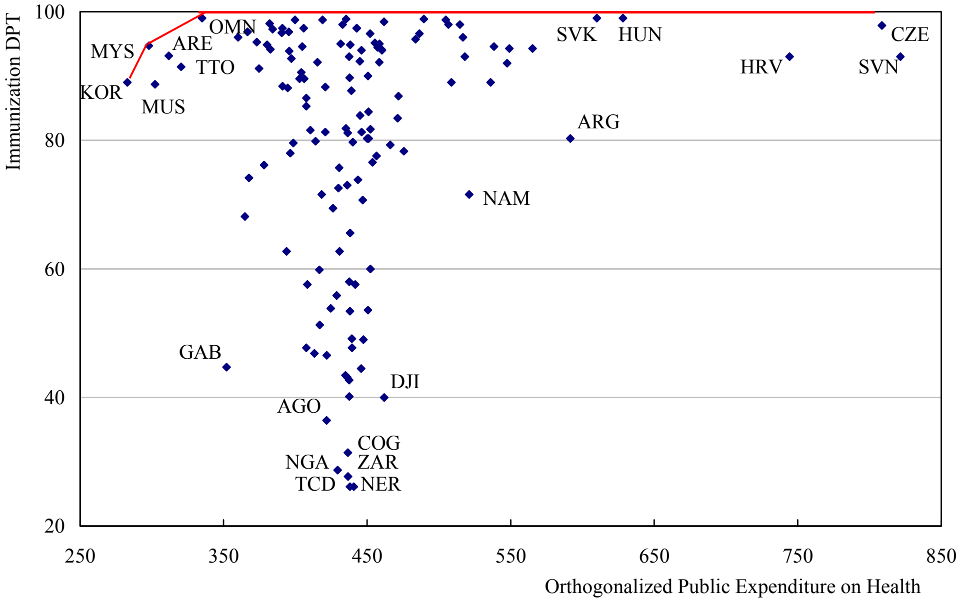
Figure 8 (continued)

Health Efficiency Frontier – Single Input/Single Output

b.1: Immunization DPT vs. Health Expenditure (Free Disposable Hull, FDH)



b.2: Immunization DPT vs. Health Expenditure (Data Envelopment Analysis, DEA)

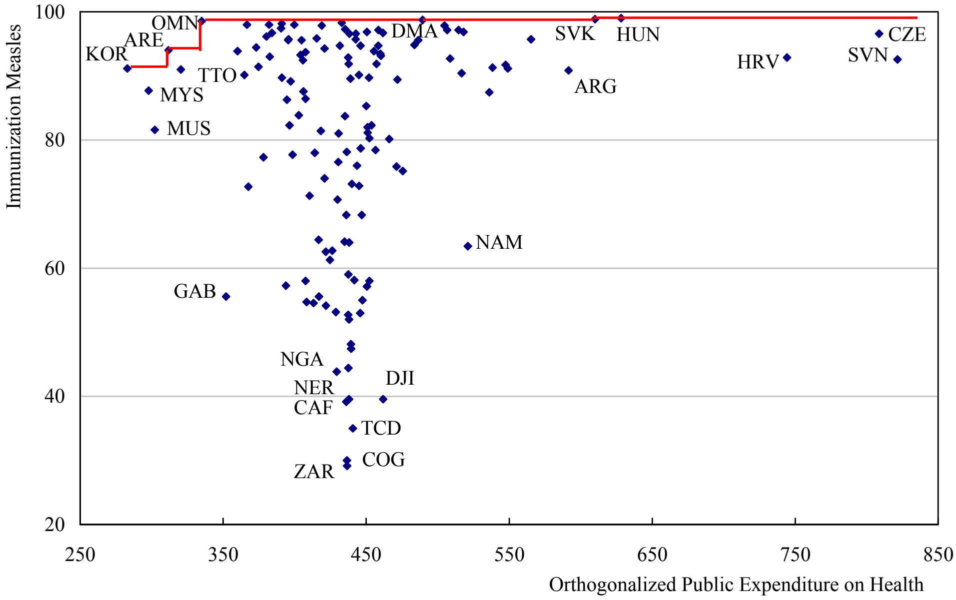


Source for the figures on this page: World Bank WDI.

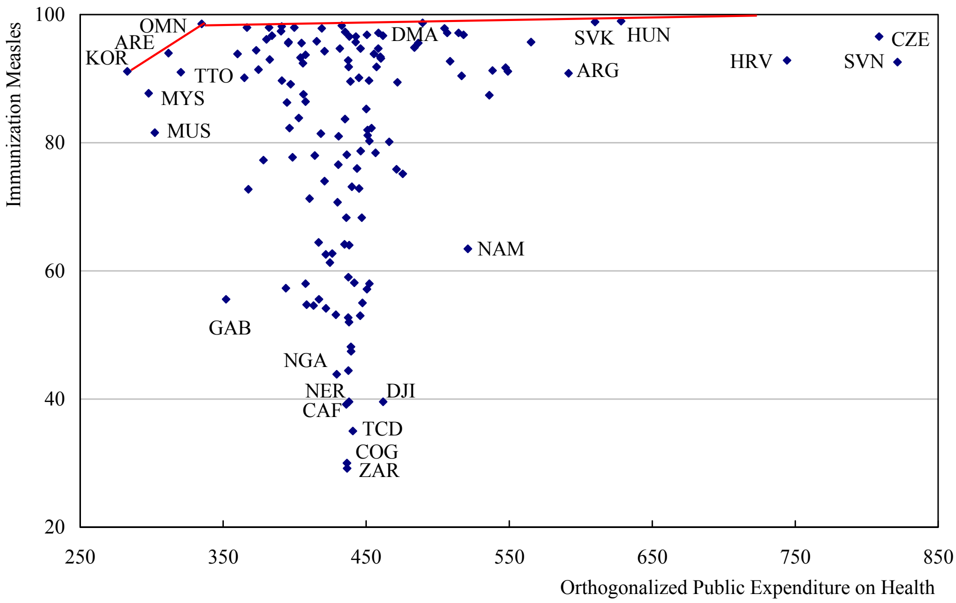
Figure 8 (continued)

Health Efficiency Frontier – Single Input/Single Output

c.1: Immunization Measles vs. Health Expenditure (Free Disposable Hull, FDH)



c.2: Immunization Measles vs. Health Expenditure (Data Envelopment Analysis, DEA)

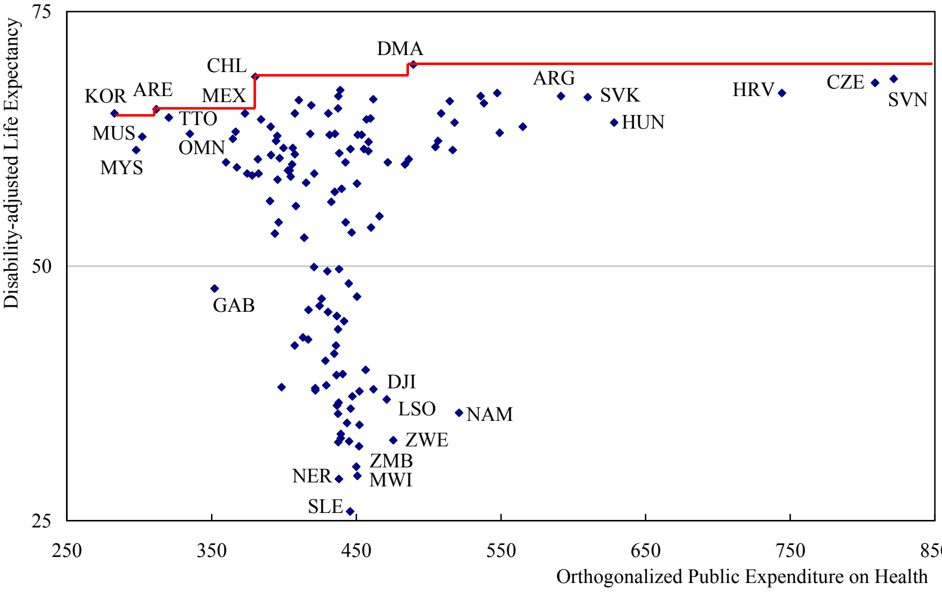


Source for the figures on this page: World Bank WDI.

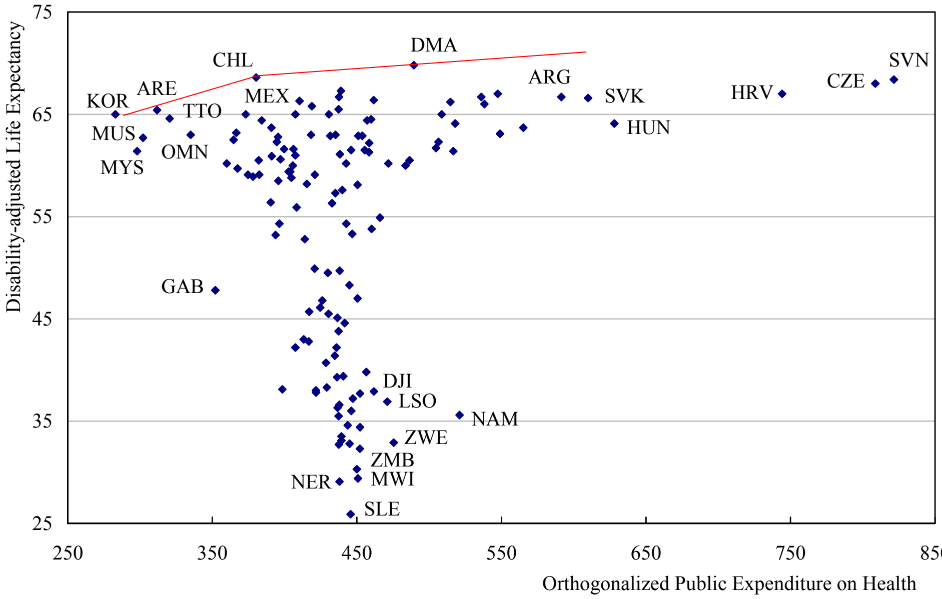
Figure 8 (continued)

Health Efficiency Frontier – Single Input/Single Output

d.1: DALE vs. Health Expenditure (Free Disposable Hull, FDH)



d.2: DALE vs. Health Expenditure (Data Envelopment Analysis, DEA)



Source for the figures on this page: World Bank WDI and Mathers *et al.* (2000).

two groups of countries: one group of relatively rich countries like the Czech Republic, Croatia, Slovenia, and Hungary that have big expenditure levels and not extremely high output (input inefficiency) and other group of countries that spend relatively little but their output indicators could be substantially larger, like Sierra Leone, Namibia, Zimbawe, and Lesotho.

c) To capture this difference, it is convenient to examine the output efficiency scoring. The rankings between input and output orientations are highly correlated.

d) With the four output indicators deciles, more efficient and least efficient countries are listed in Table 4. The group of least efficient countries could, on average, increase output significantly for a given expenditure level. For instance, the decile of most inefficient countries could almost double the disability-adjusted life expectancy (DALE) index to achieve the same efficiency as the benchmark. Similarly the DPT immunization would have to triple to achieve the same efficiency level than the benchmark developing countries.

e) The regional aggregation of the efficiency scores, by each individual output indicator shows that input efficiency scores (Table 5) are lower than output efficiency scores (Table 6). This is especially true in ECA, LAC and MNA, and to a lesser extent in EAP and SAS. In Africa, both scores are strikingly similar, indicating that, on average, the region spend about 35 per cent in excess of the benchmark cases to achieve the same output level. Alternatively, the output level is 35 per cent below comparable efficient countries that use the same input (expenditure) level.

3.3 Multiple inputs and multiple outputs

Both education and health attainment are not solely determined by public spending. Other inputs, such as private spending also affect the output indicators. For health, the World Bank WDI database reports a comparable statistic across countries. Unfortunately, a comprehensive database of this variable does not exist for education: for the education production technology we have multiple indicators of educational attainment, and three inputs (public spending, teachers per pupil, and adult literacy rate). In health, besides public spending, two other inputs were included: private spending and the education level of adults. The analysis was limited to include up to three outputs. Too many output indicators will complicate the analysis, biasing efficiency scores towards one, increasing the variance of the estimators, and reducing their speed of convergence to the true efficiency estimators (Simar and Wilson, 2000; Groskopff, 1996).

In education, the selected input-output combinations produce rankings that are somewhat similar: the average rank correlation coefficient is .53 The frequency distribution of the efficiency estimators is similar in all the models, and as the model shifts from a basic two-input two-output model to a more complex three-input/three-output model, the frequency distribution shifts to the right, that is, more concentrated around more efficient results.

Table 4

Health Attainment – Single Input/Single Output

	Input-Efficient	**Output-Efficient**
More efficient	Korea, Malaysia, Thailand, Trinidad and Tobago, Oman, United Arab Emirates, Mauritius, Kuwait, Chile	Korea, Dominica, Oman, United Arab Emirates, Anigua and Barbuda
Least efficient	Argentina, Estonia, Czech Republic, Slovenia, Macedonia, Croatia, Namibia, Tunisia, Latvia, Hungary, Barbados	Sierra Leone, Ethipia, Burkina Fasso, Central African Republic, Mali

Table 5

Health Attainment: Input Efficiency Scores by Regions across the World Single Input/Single Output

	AFR	**EAP**	**ECA**	**LAC**	**MNA**	**SAS**
Life Expectancy at Birth	.65	.72	.58	.69	.73	.69
Immunization DPT	.66	.73	.63	.68	.76	.71
Immunization Measles	.65	.73	.67	.69	.76	.71
DALE	.65	.72	.60	.70	.71	.69

Table 6

Health Attainment: Output Efficiency Scores by Regions across the World Single Input/Single Output

	AFR	**EAP**	**ECA**	**LAC**	**MNA**	**SAS**
Life Expectancy at Birth	.63	.87	.91	.92	.90	.83
Immunization DPT	.62	.83	.95	.87	.90	.75
Immunization Measles	.63	.83	.95	.91	.90	.71
DALE	.56	.83	.90	.90	.86	.79

Table 7

Educational Attainment – Multiple Inputs/Multiple Outputs

	Input-Efficient	**Output-Efficient**
More efficient	Bangladesh, Bahrain, Dominican Republic, Argentina, Estonia	Argentina, Bangladesh, Chile, Brazil, Bahrain, Dominican Republic, Congo
Least efficient	Zimbabwe, Lesotho, Botswana, Costa Rica, Swaziland, Saudi Arabia, Malaysia	Sudan, Ghana, Tanzania, Ethiopia, Kenya, Niger

The multi-input output model results (Table 7) in general confirm the results of Table 1. Some new countries that appear as efficient are Bangladesh, Congo and Argentina. In the case of Bangladesh and Congo, this is the result of considering literacy of adults as a factor of production, that in these countries is low, and hence, appearing as very efficient. Congo has also extremely low ratio of teachers per student, the other factor of production, reinforcing the bias towards the efficient score. Within the least efficient countries, the models point at Zimbabwe, Lesotho, Botswana, Malaysia, and Saudi Arabia as the single-input models. In addition, Costa Rica and Swaziland appear as input-inefficient.

The regional aggregation for input and output efficiency scores using the multiple input-output framework show (Tables 8 and 9) that as the model becomes more complex (adding inputs or outputs), scores tend to show more efficient regions. The input efficiency regional aggregation allows several interesting comparisons across the regions on the impact of an additional input on the efficiency scores. For instance, the first two rows of Table 8 allow examination of the impact of adding literacy of adults as an additional input. The biggest impact is in the MNA region, followed by ECA and LAC, while in the others the increase in efficiency scores is more marginal.[9] Output efficiency scores change substantially in MNA and Africa.

Rows 4 and 5 of Table 8 allow comparing the impact of adding the variable teachers per pupil as an additional input. In Africa the change is dramatic, while in ECA and MNA there is no significant change. Further analysis is required to explain this differential response to the inclusion of this input.

In health there are multiple combinations of inputs (public expenditure, private expenditure, and literacy of adults) and outputs (life expectancy at birth, immunization DPT, immunization measles, and disability-adjusted life expectancy (DALE)). The combinations we selected produce rankings that are more

[9] The statistical significance of these changes has yet to be determined. The tests developed by Banker , and used in previous sections do not apply to the multiple-output cases we are analyzing here (Simar and Wilson, 2000).

Table 8

Education Attainment: Input Efficiency Scores by Regions across the World Multiple Inputs/Multiple Outputs

	AFR	EAP	ECA	LAC	MNA	SAS
2 inputs (public expenditure, teachers per pupil) – 2 outputs (gross primary and secondary enrolment)	.88	.83	.72	.82	.73	.91
3 inputs (public expenditure, teachers per pupil, literacy of adult) – 2 outputs (gross primary and secondary enrolment)	.92	.89	.86	.89	.92	.96
3 inputs (public expenditure, teachers per pupil, literacy of adult) – 2 outputs (net primary and secondary enrolment)	.87	.94	.93	.93	.92	1.0
2 inputs (public expenditure, literacy of adult) – 3 outputs (first complete, second level complete, average years of school)	.78	.92	.95	.84	.80	.91
3 inputs (public expenditure, literacy of adult, teachers per pupil) – 3 outputs (first complete, second level complete, avg yrs of school)	.91	.97	.94	.89	.81	.95
3 inputs (public expenditure, teachers per pupil, literacy of adult) – 3 outputs (literacy of youth, first level complete, second level complete)	.91	.97	.94	.89	.80	.95

homogeneous. The rank correlation is in the range of .65 to .98. (Tables 10-12). In health, Bangladesh appears also as efficient, as well as Niger, this being the result of the low levels of literacy of adults that bias these countries to appear as efficient.

3.4 Efficiency change over time

To examine the evolution of input and output efficiency over time, we computed the efficiency scores in two different time periods: 1975-1980 and

Table 9

Education Attainment: Output Efficiency Scores by Regions across the World Multiple Inputs/Multiple Outputs

	AFR	EAP	ECA	LA	MNA	SAS
2 inputs (public expenditure, teachers per pupil) – 2 outputs (gross primary and secondary enrolment)	.68	.83	.80	.85	.71	.79
3 inputs (public expenditure, teachers per pupil, literacy of adult) – 2 outputs (gross primary and secondary enrolment)	.82	.88	.89	.89	.91	.90
3 inputs (public expenditure, teachers per pupil, literacy of adult) – 2 outputs (net primary and secondary enrolment)	.79	.97	.96	.96	.92	1.0
2 inputs (public expenditure, literacy of adult) – 3 outputs (first complete, second level complete, average years of school)	.64	.87	.94	.80	.79	.83
3 inputs (public expenditure, literacy of adult, teachers per pupil) – 3 outputs (first complete, second level complete, average years of school)	.86	.94	.93	.86	.80	.89
3 inputs (public expenditure, teachers per pupil, literacy of adult) – 3 outputs (literacy of youth, first level complete, second level complete)	.98	1.0	1.0	.98	.99	.99

Table 10

Health Attainment – Multiple Inputs/Multiple Outputs

	Input-Efficient	Output-Efficient
More efficient	Bangladesh, Malaysia, Costa Rica, Kuwait, Morocco, Oman, Mauritius, Niger	Bangladesh, Costa Rica, Kuwait, Malaysia, Morocco, Mauritius, Oman, Niger
Least efficient	Russia, Belarus, Namibia, Romania, Estonia, Croatia, Lituania, Hungary, Jordan	Namibia, Togo, Ethiopia, Mozambique, Cote d"Ivoire, Cameroon, Congo, Central African Republic, Nigeria, Uganda

Table 11

Health Attainment: Input Efficiency Scores by Regions across the World Multiple Inputs/Multiple Outputs

	AFR	EAP	ECA	LA	MNA	SAS
2 inputs (public expenditure, literacy of adult) – 2 outputs (life expectancy, immunization DPT.)	.85	.82	.72	.82	.91	.93
3 inputs (public expenditure, private spending, literacy of adult) – 2 outputs (life expectancy, immunization DPT.)	.86	.82	.74	.83	.91	.94
3 inputs (public expenditure, private spending, literacy of adult) – 2 outputs (life expectancy, immunization measles.)	.86	.82	.77	.83	.91	.94
3 inputs (public expenditure, private spending, literacy of adult) – 3 outputs (life expectancy, immunization DPT., DALE)	.86	.82	.80	.87	.93	.94

Table 12

Health Attainment: Output Efficiency Scores by Regions across the World Multiple Inputs/Multiple Outputs

	AFR	EAP	ECA	LA	MNA	SAS
2 inputs (public expenditure, literacy of adult) – 2 outputs (life expectancy, immunization DPT.)	.81	.91	.97	.93	.97	.96
3 inputs (public expenditure, private spending, literacy of adult) – 2 outputs (life expectancy, immunization DPT.)	.81	.91	.97	.94	.97	.96
3 inputs (public expenditure, private spending, literacy of adult) – 2 outputs (life expectancy, immunization measles.)	.80	.91	.96	.94	.98	.96
3 inputs (public expenditure, private spending, literacy of adult) – 3 outputs (life expectancy, immunization DPT., DALE)	.82	.91	.97	.95	.98	.97

1996-2002 for education study, and 1997-99 and 2000-02 for health study, the construction of which is driven by data availability.[10]

Comparison of different input-output bundles in different time periods has to be done carefully because the frontier can be shifting outward through time. In some cases the frontier displacement can be parallel (such as in the life expectancy case of Figure 9). In others, the frontier displacement can be very uneven (biased frontier shift in Figure 9) reflecting biased technological change.

The detailed comparison between observed input-output combinations in different time periods distinguishes whether variations in the levels of input utilization or output production levels are due to changes in efficiency or changes in technology. This testing is possible with observed levels of inputs and outputs, and are based on the concept of a Malmquist Index (Fare, Grosskpof, Norris and Zhang, 1994). This method has been used to study productivity change in the OECD economies, as well as productivity in agriculture across the world (Coelli and Rao, 2003; Nin, Arndt, and Preckel, 2003).

Results show that over the two decades output efficiency growth was faster in the most inefficient countries, showing that there is a "catching-up" phenomenon. However, when measuring input efficiency, the previous results do not hold: most regions increased expenditure levels without increasing output.[11]

4. Explaining inefficiency variation across countries

This chapter seeks to identify factors correlated with inefficiency scores variation across countries. This two-stage approach attempts to identify statistically significant regularities common to efficient or inefficient countries using the more basic statistical techniques. This exercise does not try to identify supply or demand factors that affect health and education outcomes, such as those described by Filmer (2003). The scope is limited to verifying statistical association between the efficiency scores and environmental variables.

4.1 Method, variables and data description

Given that the dependent variable, the efficiency scores, is continuous and distributed over a limited interval (between zero and one), it is appropriate to use a censored (Tobit) regression model to analyze the relationships with other variables. The panel consists of a large numebr of countries (varying from 70 to 140 depending on the output indicator) and only two time periods. The literature on panel estimation has shown that in panels with this configuration, that is, a large number of cross-section units (countries) and a relatively short time dimension (2 periods),

---

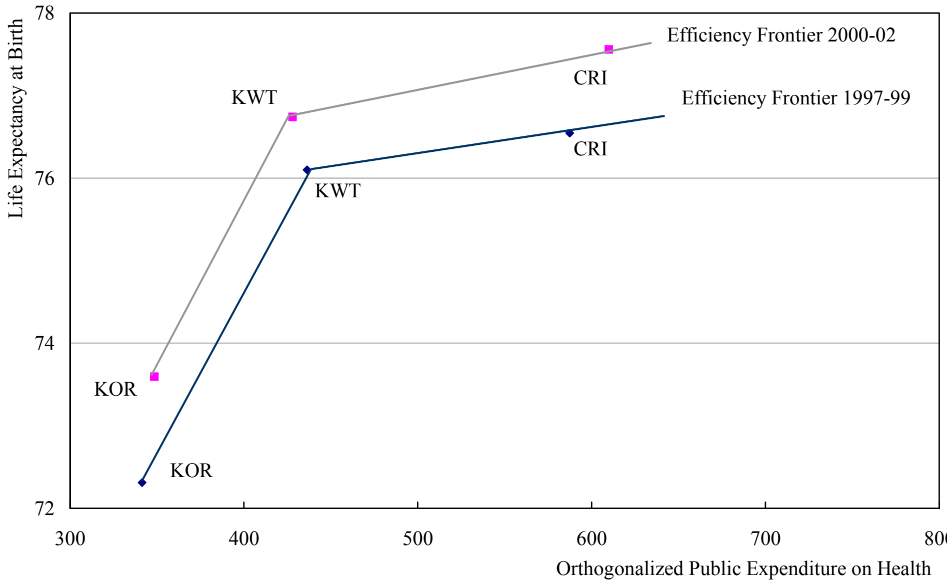
[10] Scores for individual countries are available at the PRMED website indicated in footnote 1.

[11] The results on country-by-country basis can be found at the PRMED website indicated in footnote 1.

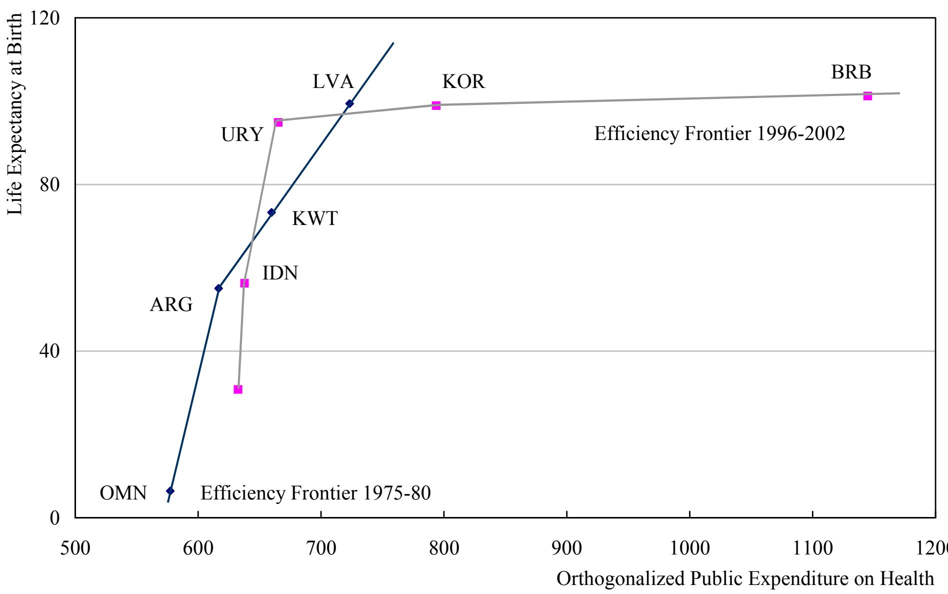
Figure 9

Efficiency Frontier Shift over Time

Parallel Shift



Biased Shift



Source for the figures on this page: World Bank WDI.

the fixed-effects estimators of the coefficients will be inconsistent (Maddala, 1987) and their variance will be biased downward (Greene, 2003b). Hence the random effects panel estimation method was preferred.

The dependent variable in the Tobit panel is the input efficiency score calculated by DEA method in the first stage. The input-oriented estimator reflects the consideration that input choices are more under the policymaker's control. The independent variables reflect environmental effects included in precursor papers, as well as suggested by others recently. We included the following independent variables:

a) *The size of government expenditure*. Most of the papers surveyed in the previous section explore the relationship between the size of the government (or expenditure as a percentage of GDP) and efficiency levels. The objective is to verify if additional pubic spending is associated with better education and health outcomes. While some papers have found a negative association between efficiency and expenditure levels (Gupta and Verhoeven, 2001, Jarasuriya and Woodon, 2003, and Afonso *et al.*, 2003), others have found a positive association (Evans *et al.*, 2003) and others have found no significant impact (Filmer and Pritchett, 1999).

b) *A government budget composition variable*. Given that both education and health are labor-intensive activities, the government's labor policies will determine the efficiency with which outputs are delivered. We choose a budget composition indicator to reflect this, in particular, the ratio of the wage bill to the total budget. A higher ratio is expected to be negatively correlated with efficiency.

c) *Per capita GDP*. We included the per capita GDP to control for the Balassa-Samuleson effect in comparing across countries. If richer countries tend to be more inefficient (given higher wages in these countries), a negative sign is expected. However, it must be recalled that to obtain the efficiency scores in the "fist stage" we constructed an auxiliary variable (the orthogonalized public expenditure). Hence the inclusion of this variable in the second stage is an attempt to control for any remaining Balassa-Samuleson effects.

d) *Urbanization*. The clustering of agents make it cheaper to provide services in urbanized areas rather than in rural. Higher degree of urbanization should reflect in higher efficiency, making positive as the expected sign of the coefficient on this variable.

e) *Prevalence of AIDS*. Based on WHO mappings of the disease, we included a dummy variable in the most severely affected countries to control for the role of this epidemic in the poor health outcomes. Evans *et al.* (2000) report that AIDS lowers the disability-adjusted life expectancy (DALE) by 15 years or more. Aids also affects education outcomes both directly and indirectly (Drake *et al.*, 2003), directly because school-age children are affected: UNAIDS estimates that almost 4 million children have been infected since the epidemic began, and two thirds have died. However, the indirect channel is relatively more important: AIDS leaves orphaned children that are more likely to drop out of school or repeat. All these factors reflect how AIDS affect the demand for education. But the supply is also affected by the decreasing teacher labor force due to illness or death, or the need to care for family (Pigozzi, 2004). Prevalence of HIV/AIDS should be negatively associated with education and health outcomes. Consequently, efficiency scores should be negatively associated with the dummy variable.

f) *Income distribution inequality*. Ravallion (2003) argues that, besides the mean income, its distribution affects social indicators because their attainment is mostly determined by the income of the poor. Hence, we controlled for the distribution of income by including the Gini coefficient as an explanatory variable. Higher inequality is expected to be associated with lower educational and health attainments, making negative the expected sign of this variable.

g) *Share of public sector in the provision of service*. Services can be provided by both the public and private sectors, and efficiency indicators will differ across countries depending on the relative productivities of both sectors. Previous studies have included this variable to explain differences in outcomes (Le Grand, 1987; Berger and Messer, 2002) or efficiency scores (Greene, 2003a). The specific variable we included was the ratio of publicly financed service over the total spending (sum of private and public spending).

h) *External aid*. To the extent that countries do not have to incur the burden of taxation, they may not have the incentive to use resources in the most cost-effective way. Another channel through which aid-financing may affect efficiency is through the volatility and unpredictability of its flows. Given that this financing source is more volatile than other types of fiscal revenue (Bulir and Hamann, 2000), it is difficult to undertake medium-term planning within activities funded with aid resources. If this is the case, we would expect a negative association between aid-dependence and efficiency in those activities funded with aid, mostly health services. To our knowledge there are no previous attempts to establish a relationship between efficiency and the degree to which activities are financed by external aid. There is, however, recent evidence of a negative association between donor financing and some health outcomes (Bokhari, Gottret and Gai, 2005).

i) *Institutional variables*. Countries with better institutions, more transparency, and less corruption are expected to have higher efficiency scores. Similarly, countries that have suffered wars or state failures are expected to register lower efficiency scores. To capture these effects we included different indicators: the ICRG International Country Risk Indicators, the Worldwide Governance Research Indicators, in particular the Control of Corruption component (Kaufmann *et al.*, 2002). We also included a dummy variable if there had been some type of state failure, such as internal wars, from the State Failure Task force database.

The data on educational and health indicators are not available on a continuous annual basis for many countries. Thus, averages of the variables were computed over sub-periods both in the first stage calculation of efficiency score and in the second stage of regression analysis. Specifically, educational indicators are averaged over two periods (1975-80 and 1996-2002) and health indicators over two periods (1996-99 and 2000-02). This discrepancy in the sub-period construction is due exclusively to the lack of data for earlier years. The averages are treated as separate observations. The advantages of this approach are threefold. First, the averages may serve as a better measure of the educational and health attainment, which can hardly be substantially improved on a yearly basis; second, the averaging maximizes the coverage of countries for each period, since one observation of a certain year is sufficient to help the country survive in the cross sectional comparison; Third, the time series thus constructed for each country, although short, facilitates the implementation of econometric techniques on panel data to explore the efficiency variations across countries and through time.

4.2 Results

The Tobit estimation on panel data is defined as follows:

$$VRSTE_{it} = f(WAGE_{it}, GOVEXP_{it}, PUBTOT_{it}, \\ GDPPC_{it}, URBAN_{it}, AIDS_{it}, GINI_{it}, EXTAID_{it}, INST_{it}, CONS)$$

where:

$VRSTE_{it}$ = Variable returns to scale DEA efficiency score for single output and multiple output cases

$WAGE_{it}$ = Wages and salaries (percent of total public expenditure)

$GOVEXP_{it}$ = Total government expenditure (percent of GDP)

$PUBTOT_{it}$ = Share of expenditures publicly financed (public/total)

$GDPPC_{it}$ = GDP per capita in constant 1995 US dollars

$URBAN_{it}$ = Urban population (percent of total)

$AIDS_{it}$ = Dummy variable for HIV/AIDS

$GINI_{it}$ = Gini Coefficient

$EXTAID_{it}$ = External aid (percent of fiscal revenue)

$INST_{it}$ = Institutional indicators including ICRG country risk, World Governance Research Indicators (Corruption Control), or a dummy for state failures from the State Failure Task Force database

$CONS$ = Constant

Tables 13 and 14 report the results for the single input/single output case and the multiple input/multiple output case, respectively. The more interesting findings are:

a) We find that countries with larger expenditure levels also register the more inefficient scores. This result is robust to changes in the output indicator selected, to considering health or education, and to adopting either the single output or multiple output frameworks. The trade-off between size of expenditure and efficiency is quite robust.

b) Countries in which the wage bill represents a higher fraction of total expenditure tend to be more inefficient. This result does not hold for health in the multiple output framework. This difference could be due partly to the relatively decreasing number of health care professionals in the world, especially in the poorer countries (Liese *et al*., 2003). Further investigation would be required to examine why this is not the case in education.

c) Countries in which public financing is a larger share of total expenditure on the service also register lower efficiency scores. This is probably due to differential productivity rates in the provision of services. Further research would be needed to explain why this is the case in health services. Recent case studies of water companies in Argentina show that private companies were more efficient than public ones and provided better service quality leading to lower child mortality rates (Galiani, Gertler and Schargrodsky, 2005). In education, there is some evidence that efficiency scores are lower in public schools (Alexander and Jaforullah, 2004), though the evidence regarding the impact of privatizing education on outcomes is mixed (World Bank, 2003).

d) Urbanization is positively associated with efficiency scores in both education and health. However, when life expectancy is included as an output, the relationship is non-significant (single output) or negative (multiple output). Possibly the urbanization variable is capturing other effects such as crime. There is ample literature studying the relationship between urbanization and crime (Glaeser and Sacerdote, 1999). Alternatively, as urbanization intensifies, communicable diseases are more difficult and costly to control, hence the negative association found between both variables in health.

e) The effect of the HIV/AIDS is clearly negative affecting health efficiency scores in the multiple-output models. However, its effect on education is less clear, as the expected negative sign is significant in few cases and has the opposite sign in equal number of cases. This confirms the difficulty of empirically verifying this relationship, reported in previous work (Wobst and Arndt, 2003).

f) Income distribution has the expected negative effect on the educational and health efficiency scores. The impact of inequality on health scores is less robust than in education, but confirms Greene's findings (2003). Other papers (Berger and Messer, 2002), have found a positive association between income inequality and health outcomes.

g) Results showed a negative relationship between some of the efficiency scores and the external aid dependency ratio. Only in one of the multiple output cases is the external aid associated with higher efficiency, but with borderline statistical significance. Though no causality relationship can be inferred from the exercise, this is one of the results that merit more detailed research. This result might be explained by the volatility of aid as a funding source that limits medium term planning and effective budgeting. Probably this is why the negative sign is more robust in health than in education, given that donor funding is mostly directed

Table 13

Explaining Cross-country Variation in Efficiency – Single Input/Single Output

Independent Variable	Gross Primary Enrolment	Net Primary Enrolment	Gross Secondary Enrolment	Net Secondary Enrolment	Literacy of Youth	Average Years of School	First Level Complete	Secondary Level Complete	Life Expect-ancy	Immun-ization DPT	Immun-ization Measles
WAGE	–.00117[c]	–.00357[a]	–.00172[b]	–.00680[a]	–.00189[b]	–.00570[a]	–.00470[b]	–.00546[a]	.00065	–.00052	–.00049
GOVEXP	–.00387[a]	–.00546[a]	–.00340[a]	–.00455[b]	–.00387[a]	–.00696[a]	–.00566[a]	–.00765[a]	–.00269[b]	–.00078	–.00227[c]
PUBTOT	-	-	-	-	-	-	-	-	–.00213[a]	–.00150[a]	–.00135[c]
GDPPC	–.00002[a]	–.00002[a]	–.00001[a]	.00002[b]	–.00002[a]	–1.5e–6	–.00001	–7.7e–6	7.6e–7	–.00001[a]	–.00001[a]
URBAN	.00167[a]	.00143[c]	.00168[a]	.00037	.00187[a]	.00532[a]	.00551[a]	.00555[a]	–.00018	.00099[b]	.00088
AIDS	–.04471[b]	–.08731[b]	–.02204	.01243	–.02974	.12717[c]	.1211[c]	.11041	–.05473	–.01108	–.02730
GINI	–.06688	.01507	–.19326[b]	–.42311	–.18484[c]	–.44658[b]	–.34402	–.45870[b]	.22118	.09510	.08692
EXTAID	–.00094	–.00196[b]	–.00021	–.00106	–.00054	.00089	–.00025	–.00006	–.00224[c]	–.00155	–.00324[b]
CONS	1.02996[a]	1.1282[a]	1.0472[a]	.84138[a]	1.0697[a]	.76791[a]	.70009[a]	.81705[a]	.79193[a]	.78734[a]	.84384[a]
# of Obs (# of Countrs)	79 (51)	44 (30)	79 (51)	34 (20)	72 (46)	71 (45)	71 (45)	71 (45)	118 (69)	118 (69)	118 (69)
Wald Chi2(6) (Prob > Chi2)	83.91 (.00)	66.09 (.00)	46.72 (.00)	55.31 (.00)	44.27 (.00)	64.13 (.00)	45.53 (.00)	61.94 (.00)	50.83 (.00)	123.97 (.00)	35.01 (.00)

Note: [a] 0.01 significance level, [b] 0.05 significance level, [c] 0.10 significance level, and insignificant otherwise.

Table 14

Explaining Cross-country Variation in Efficiency – Multiple Inputs/Multiple Outputs

Independent Variable	EDU2-2	EDU2-2n	EDU3-2	EDU3-2n	EDU3-3	EDU3-3bl	HEA2-2	HEA3-2	HEA3-2m	HEA3-3
WAGE	–.00212[b]	–.00767[a]	–.00219[b]	–.00425	–.001000	–.00340[c]	.00126[a]	.00205[a]	.00203[c]	.00203[c]
GOVEXP	–.00321[a]	–.00365	–.00203[c]	.00099	–.00123[c]	–.00316[c]	–.0012[c]	–.00273[a]	–.0009	–.00090
PUBTOT	-	-	-	-	-	-	–.00151[a]	–.00142[a]	–.00159[c]	–.00151[c]
GDPPC	–.00001[b]	–6.6e–7	–.00001[c]	–.00003	–4.2e–6	1.98e–6	–2.7e–6	4.2e–6[a]	–7.1e–7	–9.3e–7
URBAN	.00138[c]	–.00045	.00191[b]	.001997	.00127[a]	.00091	–.00095[a]	–.00148[a]	–.00106	–.00105
AIDS	–.03295	–.05843	–.00956	–.14763	.01797	.06022	–.04815[a]	–.033147[b]	–.07162	–.06999
GINI	–.06485	.43602	–.14717	.27058	–.17237[b]	–.15697	–.03997	–.07958[c]	–.01015	–.01387
EXTAID	.00010	–.00622	.00152	–.00274	–.00066	.00123	.00087	.00128[c]	–.00095	–.00106
CONS	1.0655[a]	1.0223	1.0642[a]	1.0124[a]	1.06570[a]	1.1218[a]	1.0098	1.0117[a]	.98891[a]	.98787[a]
# of Obs (# of Countrs)	76 (49)	34 (20)	69 (44)	32 (19)	69 (44)	63 (40)	97 (55)	98 (56)	98 (56)	98 (56)
Wald Chi2(6) (Prob > Chi2)	24.48 (.00)	11.69 (.11)	20.84 (.00)	7.44 (.38)	18.72 (.01)	9.18 (.24)	185.21 (.00)	229.98 (.00)	19.25 (.01)	18.62 (.02)

Notes:
[a] 0.01 significance level, [b] 0.05 significance level, [c] 0.10 significance level, and insignificant otherwise
EDU2-2: Inputs: orthogonalized public spending on education per capita, teachers per pupil
Outputs: gross primary and secondary enrolments
EDU2-2n: same inputs as EDU2-2, outputs: net primary and secondary enrolment
EDU3-2: literacy of adult is added to EDU2-2 as input
EDU3-2n: literacy of adult is added to EDU2-2n as input
EDU3-3: literacy of youth is added to EDU3-2 as output
EDU3-3bl: same inputs as in EDU3-2,
Outputs: average years of school, first level complete, and second level complete (Barro-Lee education indicators)
HEA2-2: Inputs: orthogonalized public spending on health per capita, literacy of adult
Outputs: life expectancy at birth, and immunization DPT
HEA3-2: orthogonalized private spending on health per capita is added to HEA2-2 as input
HEA3-2m: Immunization Measles is in place of DPT in HEA3-2 as output
HEA3-3: Immunization Measles is added to HEA3-2 as output.

towards the first. Recent research (Bokhari, Gottret and Gai, 2005) show a negative association between some health outcomes and the degree of donor funding, pointing in this same direction. This result also coincides with researchshowing that the quality of policies is not only unrelated to donor financing, but that highly indebted countries with "bad" policies received more net transfers as a share of GDP (Birdsall *et al.*, 2003).

h) None of the institutional variables proved to be statistically significant. We interpret this result as due to the data limitations, as some of the most crucial information, for instance the corruption index is only available since 1996 and the panel exercise was limited to a cross section. The state-failure dummy variable or the ICRG indicators did not prove to be significant either. Hence, these results are not reported in any of the tables.

To investigate the possibility of slope heterogeneity across countries, we followed the approach used in Haque, Pesaran, and Sharma (1999). Specifically, the slope coefficients in each country are assumed to be fixed over time, but varying across countries linearly with the individual sample mean of GDP per capita. The final results (Tables 15 and 16) only include the statistically significant interaction terms, in order to avoid co linearity arising from the correlation between original explanatory variables and the auxiliary variable capturing the interaction of these with the sample mean of GDP per capita. Hence the estimated model is:

$$VRSTE_{it} = f(WAGE_{it}, GOVEXP_{it}, GDPPC_{it}, URBAN_{it}, AIDS_{it}, \\ GINI_{it}, WAGEG_{it}, GOVG_{it}, GINIG_{it} CONS)$$

where:

$VRSTE_{it}$ = Variable returns to scale DEA efficiency score for single output and multiple output cases

$WAGE_{it}$ = Wages and salaries (percent of total public expenditure)

$GOVEXP_{it}$ = Total government expenditure (percent of GDP)

$PUBTOT_{it}$ = Share of expenditures publicly financed (public/total)

$GDPPC_{it}$ = GDP per capita in constant 1995 US dollars

$URBAN_{it}$ = Urban population (percent of total)

$AIDS_{it}$ = Dummy variable for HIV/AIDS

$GINI_{it}$ = Gini Coefficient

$CONS$ = Constant

$WAGEG_{it} = WAGE_{it} * \overline{GDPPC_i}$

$GOVG_{it} = GOVEXP_{it} * \overline{GDPPC_i}$

$$GINIG_{it} = GINI_{it} * \overline{GDPPC_i}$$

$$\overline{GDPPC_i} = T^{-1}\sum_{t=1}^{T} GDPPC_{it}$$

Results show that the interaction terms are significant, especially for the health regression, implying that there is a heterogeneous response of efficiency scores to the different explanatory variables. This confirms Greene's (2003) results on the WHO data. One of the key results of this section is that the negative association between the size of government expenditure and efficiency is stronger in countries with higher per capita GDP. Similarly, this happens with the wage variable. Results are somewhat similar to those of the homogeneous slopes, though statistical significance of many of the coefficients is lower. This is the result of colinearity between the auxiliary variables and the original set of explanatory variables. This problem deserves further work in the future.

Interpretation of these results requires caution due to several limitations. First, education and health outcomes are explained by multiple supply and demand factors (Filmer, 2003) that are not included here. This is not the object of the present paper. The omission of one of these factors in the health or education production functions in the previous stage could explain some of the cross-country covariation of the efficiency results (Ravallion, 2003). The goodness-of-fit analysis of the first stage indicated that no important factor seemed to be omitted. Of course, there can always be additional factors that could be included but the curse of dimensionality[12] is particularly pressing in non-parametric statistical methods (even if the data were available).

The second limitation derives from the intuitive question why the set of explanatory variables used in the second stage were not included in the first stage. The answer lies in that most of these variables are environmental and outside the control of the decision-making unit. The inclusion of these environmental variables would have had little justification from the production function perspective. Additionally, by maintaining the production function as simple as possible the dimensionality curse is avoided.

Finally, the third limitation arises from the fact that if the variables used in the first stage to obtain the efficiency estimator are correlated with the second stage explanatory variables, the coefficients will be inconsistent and biased (Simar and Wilson, 2004; Grosskopf, 1996; Ravallion, 2003). To examine the extent of this potential problem we calculated correlation coefficients between the first-stage inputs and the second-stage explanatory variables. The largest correlation coefficients were between GDP per capita and the teachers per pupil ratio and the literacy of the adult. To examine the sensitivity of the results to the inclusion of GDP per capita, all the estimations were performed without this variable and none of the results changed.

---

[12] As the number of outputs increase, the number of observations must increase exponentially to maintain a given mean-square error of the estimator. See Simar and Wilson (2000).

Table 15

Explaining Cross-country Variation in Efficiency – Single Input/Single Output – Heterogeneous Slopes

Independent Variable	Gross Primary Enrolment	Net Primary Enrolment	Gross Secondary Enrolment	Net Secondary Enrolment	Literacy of Youth	Average Years of School	First Level Complete	Secondary Level Complete	Life Expectancy	Immun-ization DPT	Immun-ization Measles
WAGE	–.00006	.00076	–.00035	–.00228	–.00056	–.00200	–.00120	–.00419	–.00306[c]	–.00079	–.00241
GOVEXP	–.00363[a]	–.00255[c]	–.00377[a]	–.00727[c]	–.00552[a]	–.00595[c]	–.00453	–.00611[c]	.00337[b]	.00168[c]	.00221
PUBTOT	-	-	-	-	-	-	-	-	-	–.00162[a]	–.00097
GDPPC	–.00002[a]	–.00002[a]	–5.4e–6	.00003[a]	–.00002[c]	.00004[a]	.00003[c]	.00003[c]	.00002[b]	–.00002[a]	–.00001
URBAN	.00179[a]	.00132[b]	.00193[a]	.00139	.00212[a]	.00566[a]	.00601[a]	.00593[a]	–.00080	–.00117[a]	.00021
AIDS	–.03866[c]	–.06603b	–.03153	.01010	–.02177	.05491	.06656	.06464	–.02321	–.04147[b]	–.00826
GINI	–.14230	–.42098a	–.14976	–.29395	–.13107	–.09995	–.15463	–.24762	–.12865	–.38851[a]	–.42162[b]
WAGG	–4.4e–6[c]	–1.2e–6a	–4.6e–7[c]	–9.4e–7	–4.5e–7	–8.1e–7	–8.8e–7	–2.4e–7	8.9e–7[b]	6.95e–8	5.1e–7
GOVG	–8.6e–8	–5.2e–7c	4.3e–8	3.6e–7	4.0e–7	–4.3e–7	–4.4e–7	–5.3e–7	–1.4e–6[a]	–5.4e–7[a]	–9.4e–7[a]
GINIG	.00003	.00011[a]	–2.4e–6	–.00003	2.0e–6	–.00006	–.00005	–.00006	.00001	.00009[a]	.00006[c]
CONS	1.0156[a]	1.1036[a]	1.0098[a]	.74603[a]	1.0365[a]	.60371[a]	.53977[a]	.68648[a]	.82665[a]	1.0119[a]	.93820[a]
# of Obs (# of Countrs)	82 (52)	47 (31)	82 (52)	36 (21)	75 (47)	74 (46)	74 (46)	74 (46)	120 (70)	121 (71)	121 (71)
Wald Chi2(6) (Prob > Chi2)	87.32 (.00)	93.98 (.00)	62.74 (.00)	105.34 (.00)	58.40 (.00)	94.00 (.00)	69.32 (.00)	82.38 (.00)	74.33 (.00)	450.54 (.00)	52.71 (.00)

Note: [a] 0.01 significance level, [b] 0.05 significance level, [c] 0.10 significance level, and insignificant otherwise.

Table 16

Explaining Cross-country Variation in Efficiency – Multiple Inputs/Multiple Outputs – Heterogeneous Slopes

Independent Variable	EDU2-2	EDU2-2n	EDU3-2	EDU3-2n	EDU3-3	EDU3-3bl	HEA2-2	HEA3-2	HEA3-2m	HEA3-3
WAGE	.00051	–.00140	.00005	.00494	–.00018	–.00045	–.00063	–.00065	–.00093	–.00092
GOVEXP	–.00323[b]	.00501	–.00385[b]	.00520	–.00256[b]	–.00459	.00122[c]	.00063	–.00070	–.00064
PUBTOT	–	–	–	–	–	–	–.00180[a]	–.00145[b]	–.00149[c]	–.00141[c]
GDPPC	–8.6e–6	.00002	1.7e–6	.00003	–1.8e–6	–2.1e–6	–.00001[b]	–.00001	–.00003[b]	–.00003[b]
URBAN	.00137[b]	.00079	.00166[b]	.00096	.00134[a]	.00064	–.00246[a]	–.00167[c]	–.00160	–.00159
AIDS	–.04139	–.06211	–.04744	–.20362[a]	.00646	.04633	–.06289[a]	–.04001	–.07217	–.07025
GINI	–.14418	–.18676	.07096	–.02601	–.07474	–.20029	–.32844[a]	–.45695[b]	–.29885	–.30857
WAGG	–8.3e–7[b]	–1.2e–6	–6.4e–7[c]	–1.9e–6	–2.0e–7	–7.9e–7	7.8e–7[a]	7.2e–7	6.0e–7	6.0e–7
GOVG	–6.3e–8	–2.6e–6[c]	3.5e–7	–1.2e–6	3.0e–7	3.5e–7	–5.98e–7[a]	–4.9e–7	2.7e–8	1.4e–8
GINIG	.00003	.00012	–.00003	.00005	–.00002	.00003	.00005[a]	.00005[c]	.00006	.00006[c]
CONS	1.0515[a]	.89986[a]	1.0021[a]	.84756[a]	1.0464	1.1257[a]	1.1494[a]	1.1457[a]	1.1512[a]	1.1495[a]
# of Obs (# of Countrs)	79 (50)	36 (21)	72 (45)	34 (20)	72 (45)	66 (41)	101 (58)	101 (58)	101 (58)	101 (58)
Wald Chi2(6) (Prob > Chi2)	41.93 (.00)	18.57 (.03)	31.15 (.00)	18.71 (.22)	23.89 (.00)	13.22 (.15)	600.70 (.00)	37.22 (.00)	25.33 (.00)	24.74 (.01)

Notes:
[a] 0.01 significance level, [b] 0.05 significance level, [c] 0.10 significance level, and insignificant otherwise
EDU2-2: Inputs: orthogonalized public spending on education per capita, teachers per pupil
Outputs: gross primary and secondary enrolments

EDU2-2n: same inputs as EDU2-2, outputs: net primary and secondary enrolment
EDU3-2: literacy of adult is added to EDU2-2 as input
EDU3-2n: literacy of adult is added to EDU2-2n as input
EDU3-3: literacy of youth is added to EDU3-2 as output
EDU3-3bl: same inputs as in EDU3-2,

Outputs: average years of school, first level complete, and second level complete (Barro-Lee education indicators)

HEA2-2: Inputs: orthogonalized public spending on health per capita, literacy of adult
Outputs: life expectancy at birth, and immunization DPT
HEA3-2: orthogonalized private spending on health per capita is added to HEA2-2 as input
HEA3-2m: Immunization Measles is in place of DPT in HEA3-2 as output
HEA3-3: Immunization Measles is added to HEA3-2 as output.

5. Concluding remarks and directions for future work

The paper presented an application of non-parametric methods to analyze the efficiency of public spending. Based on a sample of more than 140 countries, the paper estimated efficiency scores for nine education output indicators and fourth health output indicators. Our results indicate that, in general, the least efficient countries could achieve substantially higher education and health output levels. Alternatively they could produce the same output level consuming approximately 50 per cent less of the inputs implicit in the efficiency frontier. It is crucial to identify what are the institutional or economic factors that cause some countries to be more efficient than others in the service delivery.

In terms of policy implications, it is crucial to differentiate between the technically efficient level and the optimal or desired spending level. Even if a country is identified as an "efficient" benchmark country, it may very well still need to expand its public spending levels to achieve a target level of educational or health attainment indicators. Such is the case of countries with low spending levels and low attainment indicators, close to the origin of the efficient frontier. The important thing is that countries expand their scale of operation along the efficient frontier.

The methods used in the paper can be interpreted as tools to identify extreme cases of efficient units and inefficient cases. Once the cases have been identified, more in-depth analysis is required to explain departures from the benchmark, as proposed and done by Sen (1981). Given that the methods are based on estimating the frontier directly from observed input-output combinations they are subject to sampling variability and are sensitive to the presence of outliers. Recent advances allow dealing with these problems such as in Wilson (2004). Additionally, it would be useful to contrast these results with those obtained with the use of parametric stochastic frontier estimation.

In a "second stage" the paper verified statistical association between the efficiency scores and environmental variables that are not under the control of the decision-making units. The panel Tobit regressions showed that the variables, which are negatively associated with efficiency scores, include the size of public expenditure, the share of the wage bill in the total public budget, the proportion of the service that is publicly financed, the prevalence of HIV/AIDS epidemic on health efficiency scores, income inequality on education efficiency scores, and external aid-financing on some of the efficiency scores. This last impact is probably due to the volatility of aid that impedes effective medium term planning and budgeting, and probably explains why the result is more robust in health than in education where most of the donor-funding is directed. This result points in the same direction of previous research showing that donor financing is unrelated to the quality of domestic policies and that, in the case of highly indebted counties, those with worse policies received more transfers. A positive association between urbanization and efficiency outcomes is also identified in education but some of the health efficiency scores are negatively associated. This last result probably is due to higher crime rates in the cities or the effect of communicable diseases that spread with agglomeration. These are topics for further research in case studies.

APPENDIX
DATA ENVELOPMENT ANALYSIS (DEA) MODEL

A measure of production efficiency, perhaps the simplest one, is defined as the ratio of output to input. It is, however, inadequate to deal with the existence of multiple inputs and outputs. The relative efficiency for all decision-making units (DMU), *j=1,...,n*, is then modified as the ratio of weighted outputs to weighted inputs, more precisely:

$$\textit{Relative efficiency} = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \tag{A.1}$$

where x and y are inputs and outputs, respectively, and u and v are the common weights assigned to outputs and inputs, respectively. A challenge of this measure immediately follows: it is difficult to justify the common weights given that DMUs may value inputs and outputs differently.

The seminal paper by Charnes, Cooper and Rhodes (1978) proposed the following ratio form to allow for difference in weights across DMUs, which establishes the foundation of data envelopment analysis (DEA).

$$\begin{aligned}
&Max \quad h_0 = \frac{\sum_{r=1}^{s} \mu_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \\
&\textit{subject to}: \\
&\qquad \frac{\sum_{r=1}^{s} \mu_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad j = 1, \cdots, n \\
&\qquad \mu_r \ge \varepsilon, \quad r = 1, \cdots, s \\
&\qquad v_i \ge \varepsilon, \quad i = 1, \cdots, m \\
&\qquad \varepsilon > 0
\end{aligned} \tag{A.2}$$

In the model, there are $j = 1, \ldots, n$ observed DMUs which employ $i = 1, \ldots, m$ inputs to produce $r = 1, \ldots, s$ outputs. One DMU is singled out each time, designated as DMU_0, to be evaluated against the observed performance of all DMUs. The objective of model (A.2) is to find the most favorable weights, μ_r and v_i, for DMU_0 to maximize the relative efficiency. The constraints are that the same weights will make ratio for every DMU be less than or equal to unity. The optimal value of the ratio must be $0 \le h_0^* \le 1$ and DMU_0 is efficient if and only $h_0^* = 1$, otherwise it is considered as relatively inefficient. One problem with the ratio formulation is that there are an infinite number of solutions: if μ_r and v_i are solutions to (A.2), so are $\alpha\mu_r$ and αv_i, $\forall \alpha > 0$.

It is worth observing one important feature of model (A.2). In maximizing the objective function it is the relative magnitude of the numerator and the denominator that really matters and not their individual values. It is thus equivalent to setting the denominator to a constant, say 1, and maximizing the numerator. This transformation will not only lead to the uniqueness of solution but also convert the fractional formulation of model (A.2) into a linear programming problem in model (A.3).

$$
\begin{aligned}
&Max \quad \sum_{r=1}^{s} \mu_r y_{r0} \\
&subject\ \ to: \\
&\qquad \sum_{i=1}^{m} v_i x_{i0} = 1 \\
&\qquad \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1, \cdots, n \\
&\qquad -\mu_r \leq -\varepsilon, \quad r = 1, \cdots, s \\
&\qquad -v_i \leq -\varepsilon, \quad i = 1, \cdots, m
\end{aligned}
\tag{A.3}
$$

Model (A.3) facilitates straightforward interpretation in terms economics. The objective is now to maximize the weighted output per unit weighted input under various conditions, the most critical one of which is that the virtual output does not exceed the virtual input for any DMU.

Since model (A.3) is a linear programming, we can convert the maximization problem into a minimization problem, e.g. a *dual* problem, by assigning a dual variable to each constraint in the *primal* (A.3). Specifically, dual variables $\theta, \lambda_j, s_r^+, s_i^-$ are assigned as follows.

$$
\begin{array}{lll}
Max \quad \sum_{r=1}^{s} \mu_r y_{r0} & \text{Dual Variable} & \\
subject\ \ to: & & \\
\quad \sum_{i=1}^{m} v_i x_{i0} = 1 & \theta & \text{(A.3')} \\
\quad \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1, \cdots, n & \lambda_{\mathrm{j}} & \\
\quad -\mu_r \leq -\varepsilon, \quad r = 1, \cdots, s & s_r^+ & \\
\quad -v_i \leq -\varepsilon, \quad i = 1, \cdots, m & s_i^- &
\end{array}
$$

A *dual* minimization problem is thus derived as model (A.4). It is clear that model (A.4) has *m+s* constraints while model (A.3) has *n+m+s+1* constraints. Since *n* is usually considerably larger than *m+s*, the dual DEA significantly reduces the computational burden and is easier to solve than the primal.

$$
\begin{aligned}
&Min \qquad \theta - \varepsilon\left[\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right] \\
&subject\ \ to: \\
&\qquad \theta x_{i0} - \sum_{j=1}^{n} x_{ij}\lambda_{\mathrm{j}} - s_i^- = 0 \\
&\qquad y_{r0} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ \\
&\qquad \lambda_j \ge 0,\ \ s_r^+ \ge 0,\ \ s_i^- \ge 0 \\
&\qquad i = 1,\cdots,m,\ \ r = 1,\cdots,s,\quad j = 1,\cdots,n
\end{aligned}
\tag{A.4}
$$

More importantly, the duality theorem of linear programming states that the solution value to the objective function in (A.4) is exactly equal to that in (A.3). And, the dual variables, $(\lambda_1, \lambda_2, \cdots, \lambda_n)$, have the interpretation of Lagrange multipliers. That is, the value of a dual variable is equal to the shadow price of Lagrange Multiplier. It is also known that, from constrained optimization problem, $\lambda_j > 0$ normally when the constraint in (A.3') is binding and $\lambda_j = 0$ if not. Note that the binding constraint in (A.3) implies that the corresponding DMU is efficient. In another word, efficient units are identified by positive $\lambda's$ while inefficient units are given $\lambda's$ of zero. The DMU in question in model (A.4) is thus compared with the efficient DMUs only, named as comparison *peers* in the literature. The solution values of $\lambda's$ reflect the exact weights assigned to each peer in the evaluation of DMU_0.

Since only efficient DMUs exert effective constraints in model (A.4), as argued above, the input/output bundle, $(\sum_{j=1}^{n} x_{ij}\lambda_j, \sum_{j=}^{n} y_{rj}\lambda_j)$, is the most efficient combination for $i = 1,\cdots,m$ and $r = 1,\cdots,s$. To achieve an output level y_{r0}, which is as close as possible to $\sum_{j=}^{n} y_{rj}\lambda_j$, DMU_0 has to use an input bundle to meet the minimum requirement, $\sum_{j=1}^{n} x_{ij}\lambda_j$. This further implies that the solution θ^* is the lowest proportion of the current input bundle, x_{i0} used by DMU_0, that is actually required to meet the minimum input requirement and produce target output y_{r0}. The solution θ^* is defined as the efficiency score for DMU_0. For instance, $\theta^* = 0.60$ implies that 40 per cent of current input is a waste of resources.

Model (A.4) also offers the explanation why the data envelopment analysis is so named. The first constraint in (A.4) defines a lower limit of inputs and the second constraint an upper limit of outputs for DMU_0, and within the limits θ is minimized. The set of solutions to all DMUs forms an upper bound that envelops all observations.

REFERENCES

Afonso, A. and M. St. Aubyn (2004), "Non-parametric Approaches to Education and Health: Expenditure Efficiency in OECD Countries", Technical University of Lisbon, mimeo.

Afonso, A., L. Schuknecht and V. Tanzi (2003), "Public Sector Efficiency: An International Comparison", European Central Bank, Working Paper, No. 242.

Aigner, D.J., C.A.K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics*, Vol. 6, pp. 21-37.

Alexander, R. and M. Jaforullah (2004), "Explaining Efficiency Differences in New Zealand Secondary Schools", paper presented at the "International Conference on Policy Modeling", June, Paris.

Banker, R. (1993), "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation", *Management Science,* Vol. 39, No. 10.

Berger, M. and J. Messer (2002), "Public Financing of Health Expenditures, Insurance, and Health Outcomes", *Applied Economics*, Vol. 34, pp. 2105-13.

Birdsall, N., S. Claessens and I. Diwan (2003), "Policy Selectivity Forgone: Debt and Donor Behavior in Africa", *The World Bank Economic Review*, Vol. 17, No. 3.

Bishop, P. and S. Brand (2003), "The Efficiency of Museums: A Stochastic Frontier Production Function Approach", *Applied Economics*, Vol. 35, pp. 1853-58.

Bokhari, F., P. Gottret and Y. Gai (2005), "Government Health Expenditures and Health Outcomes", The World Bank, mimeo.

Bulir, A. and A. Hamann (2000), "Aid Volatility: An Empirical Assessment", IMF, Staff Papers, Vol. 50, No. 1, pp. 64-89.

Burgess, J. and P. Wilson (1998), "Variation in Inefficiency in U.S. Hospitals", *Canadian Journal of Operational Research and Information Processing*, Vol. 6, pp. 84-102.

Bowlin, W.F. (1998), "Measuring Performance: An Introduction to Data Envelopment Analysis (DEA)", *Journal of Cost Analysis*, pp. 3-27.

Cazals, C., J. Florens and L. Simar (2002), "Non-parametric Frontier Estimation: A Robust Approach", *Journal of Econometrics*, Vol. 106, No. 1, pp. 1-25.

Chang, T. (2002), "An Econometric Test of Wagner's Law for Six Countries Based on Cointegration and Error-correction Modeling Techniques", *Applied Econometrics*, No. 34, pp. 1157-69.

Charnes, A., W. Cooper and E. Rhodes (1978), "Measuring Efficiency of Decision-making Units", *European Journal of Operational Research*, No. 3, pp. 429-44.

Cherchye, L. and T. Post (2001), "Methodological Advances in DEA: A Survey and an Application for the Dutch Electricity Sector", ERIM Report Series Research in Management, ERS-2001-53-F&A.

Crouch, L. and T. Fasih (2004), "Patterns in Educational Development: Implications for Further Efficiency Analysis", The World Bank, mimeo.

Cullinane, K. and D. Song (2003), "A Stochastic Frontier Model of the Productive Efficiency of Korean Container Terminals", *Applied Economics*, Vol. 35, pp. 251-67.

Drake, L., C. Maier, M. Jukes, A. Patrikios, D. Bundy, A. Gardner and C. Dolan (2003), "School-age Children: Their Nutrition and Health".

Evans, D., A. Tandon, C.J.L. Murray and J.A. Lauer (2000), "The Comparative Efficiency of National Health Systems in Producing Health: An Analysis of 191 Countries", World Health Organization, GPE Discussion Paper Series, No. 29.

Färe, R. and S. Grosskopf (1995), "Non-parametric Tests of Regularity, Farrell Efficiency, and Goodness of Fit", *Journal of Econometrics*, Vol. 69, No. 2, pp. 415-25.

Farrell, M. (1957), "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society*, Series A, Vol. 120, No. 3, pp. 253-90.

Filmer, D. (2003), "Determinants of Health and Education Outcomes", background note for *World Development Report 2004*, The World Bank, mimeo, May.

Filmer, D. and L. Pritchett (1999), "The Impact of Public Spending on Health: Does Money Matter?", *Social Science and Medicine*, No. 49, pp. 1309-23.

Galiani, S., P. Gertler and E. Schargrodsky (2005), "Water for Life: The Impact of the Privatization of Water Services on Child Mortality", *Journal of Political Economy*, Vol. 113, No. 1.

Glaeser, E. and B. Sacerdote (1999), "Why Is There More Crime in Cities?", *Journal of Political Economy*, No. 107, pp. 225-58.

Greene, W. (2003a), "Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Cares Systems", mimeo, New York University, www.stern.nyu.edu/~wgreene

―――――― (2003b), "Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model", New York University, mimeo, www.stern.nyu.edu/~wgreene

Grosskopf, S. (1996), "Statistical Inference and Non-parametric Efficiency: A Selective Survey", *The Journal of Productivity Analysis*, Vol. 7, pp. 161-76.

Gupta, S. and M. Verhoeven (2001), "The Efficiency of Government Expenditure, Experiences from Africa", *Journal of Policy Modeling*, Vol. 23, pp. 433-67.

Haque, N., H. Pesaram and S. Sharma (1999), "Neglected Heterogeneity and Dynamics in Cross-country Savings Regressions", IMF, Working Paper.

Hanushek, E.A. (2002), "The Failure of Input-based Schooling Policies", National Bureau of Economic Research.

Hanushek, E.A. and J.A. Luque (2002), "Efficiency and Equity in Schools Around the World", National Bureau of Economic Research.

ICRG International Country Risk Indicators (2004), http://www.icrgonline.com

Jayasuriya, R. and Q. Wodon (2002), "Measuring and Explaining Country Efficiency in Improving Health and Education Indicators", mimeo, The World Bank, April.

—————— (2003), "Explaining Country Efficiency in Improving Health and Education Indicators: The Role of Urbanization", background paper for the *World Development Report 2003*, The World Bank.

Jounard, I., M. Konsgrud, Y. Nam and R. Price (2004), "Enhancing the Cost Effectiveness of Public Spending: Experience in OECD Countries", OECD, Economics Department, Working Paper, No. 380, February.

Kaufmann, D., A. Kraay and P. Zoido-Lobaton (1999a), "Governance Matters", The World Bank, Policy Research, Working Paper, No. WPS2196.

—————— (1999b), "Aggregating Governance Indicators", The World Bank, Policy Research, Working Paper, No. WPS2195.

—————— (2002), "Governance Matters II – Updated Indicators for 2000-1", The World Bank, Policy Research, Working Paper, No. WPS2772.

Le Grand, J. (1987), "Inequalities in Health: Some International Comparisons", *European Economic Review*, No. 31, pp. 182-91.

Liese, B., N. Blanche and G. Dussault (2003), "The Human Resource Crisis in Health Services in Sub-Saharan Africa", background paper for the WDR 2004.

Maddala, G.S. (1987), "Limited Dependent Variable Models Using Panel Data", *The Journal of Human Resources*, Vol. XXII, No. 3.

Mathers, C.D., R. Sadana, J.A. Salomon, C.J.L. Murray and A.D. Lopez (2000), "Estimates of DALE for 191 Countries: Methods and Results", World Health Organization, GPE, Working Paper, No. 16.

Murillo-Zamorano, L. (2004), "Economic Efficency and Frontier Techniques", *Journal of Economic Surveys,* Vol. 18, No. 1, pp. 33-77.

Murray, C.J.L., J.A. Salomon and C. Mathers (1999), "A Critical Examination of Summary Measures of Population Health", World Health Organization, GPE, Discussion Paper, No. 2.

Nin, A., C. Arndt and P. Preckel (2003), "Is Agricultural Productivity in Developing Countries Really Shrinking? New Evidence Using a Modified Non-parametric Approach", *Journal of Development Economics*, No. 71, pp. 395-415.

Pigozzi, M. (2004), "Quality Education and HIV/AIDS", UNESCO, mimeo.

Ravallion, M. (2003), "On Measuring Aggregate Social Efficiency", The World Bank, Policy Research, Working Paper, No. 3166, November.

Sen, A. (1981), "Public Action and the Quality of Life in Developing Countries", *Oxford Bulletin of Economics and Statistics*, Vol. 43, No. 4, November.

Simar, L. and P. Wilson (2000), "Statistical Inference in Non-parametric Frontier Models: The State of the Art", *Journal of Productivity Analysis*, No. 13, pp. 49-78.

The State Failure Task Force (2004), "Political Instability (aka State Failure) Task Force Data", http://www.cidcm.umd.edu/inscr/stfail/sfdata.htm

Tanzi, V. (2004), "Measuring Efficiency in Public Expenditure", paper presented at the *Conference on Public Expenditure Evaluation and Growth*, The World Bank, October.

Varian, H. (1990), "Goodness-of-fit in Optimizing Models", *Journal of Econometrics,* No. 46, pp. 125-40.

Wheelock, D. and P. Wilson (2003), "Robust Non-parametric Estimation of Efficiency and Technical Change in U.S. Commercial Banking", Federal Reserve Bank of St. Louis, Working Paper, November.

Wilson, P. (2004), "A Preliminary Non-parametric Analysis of Public Education and Health Expenditures in Developing Countries", The World Bank, mimeo.

Wobst, P. and C. Arndt (2003), "HIV/AIDS and Primary School Performance in Tanzania", presented at the *25th International Conference of The International Association of Agricultural Economists*, August 16-22, Durban, South Africa.

World Bank (2004), *Making Services Work for the Poor. World Development Report 2004.*

Worthington, A. (2001), "An Empirical Survey of Frontier Efficiency Measurement Techniques in Education", *Education Economics*, Vol. 9, No. 3.