An international comparison of educational systems: with an application of the Malmquist-Luenberger global index*

Víctor Giménez
Universitat Autònoma de Barcelona

Claudio Thieme
Universidad Diego Portales

Diego Prior
Universitat Autònoma de Barcelona

Emili Tortosa-Ausina
Universitat Jaume I and Ivie

November 7, 2013

Abstract

This study uses the Global Malmquist-Luenberger productivity index to measure performance change of educational systems in 28 countries participating in TIMSS 2007 and 2011 for eighth grade of basic education in the discipline of mathematics. This methodology is particularly appropriate both for its desirable properties as well as its suitability for the educational context. Results indicate that the different countries participating in the study not only chose several paths to improve their educational performance but, in addition, results varied remarkably among them. They also suggest that there has been, on average, a deterioration of educational performance between 2007 and 2011, although we also observe a (successful) interest in several countries in order to improve equality.

Key words and phrases: education, efficiency, Malmquist-Luenberger, TIMSS.

JEL Classifications: C61, H52, I21.

Communications to: Víctor Giménez, Departament of Business Economics, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain. Tel.: +(34) 935812264, fax: +(34) 935812555, e-mail: `victor.gimenez@uab.es`.

*Claudio Thieme and Emili Tortosa-Ausina thank FONDECYT (National Fund of Scientific and Technological Development, grant #1121164) for generous financial support. Diego Prior and Emili Tortosa-Ausina acknowledge the financial support of Ministerio de Ciencia e Innovación (ECO2010-18967/ECON and ECO2011-27227). Emili Tortosa-Ausina also acknowledges the financial support of Generalitat Valenciana (PROMETEO/2009/066). The usual disclaimer applies.

1. Introduction

In a world characterized by rapid technological change and the importance of innovation processes, the level of academic attainment that students of a given country may achieve is essential for improving its citizens' levels of wealth and welfare. For this reason, it is not surprising that, in the field of public policy in education, there is a growing concern in the assessment of student learning (Denvir and Brown, 1986; Ercikan, 2006). Understanding educational outcomes is critical for an effective planning of educational policies, as well as the assessment of its reforms.

In this sense, and most recently, the International Association for the Evaluation of Educational Achievement (IEA) has released the results of the fifth version of its tests for analyzing Trends in International Mathematics and Science Study (TIMSS). TIMSS 2011 evaluates and describes the learning of students in participating countries, for these two disciplines, providing also additionally vital information on other relevant factors (curricular, instructional, or related to the availability of resources) which can affect the process of teaching and learning.

While the results obtained by a given country in a standardized test (such as TIMSS, or PISA) are a good reflection of the students' academic level, they cannot be regarded by themselves as an indicator for the performance of their educational systems and, therefore, their school system authorities. Indeed, the main limitations associated with these standardized international test are: (i) the assessment of an organization's performance (in this particular case, a country) does not depend exclusively on outcome variables; instead, we may consider efficiency indicators which measure different aspects involved in the educational process; the results achieved (output) during this process are a consequence of the resources used, the process itself as well as environmental variables beyond educational authorities' control (Teddlie and Reynolds, 2000); (ii) for a given country, the measure of the results of the educational process should not be constrained to the knowledge acquired during the school by its students, but should also include other outcomes such as the percentage of students failing to meet minimum learning standards (which would correspond to an *undesirable* outcome of the educational process, in terms of educational inequality); and (iii) when measuring students' educational achievements in a given point in time, it is difficult to disentangle how much of it is attributable to the student herself, her family, or the strategies started by previous educational authorities.

To our knowledge, the number of studies comparing the performance of educational systems in different countries are relatively few, and there are no previous studies analyzing explicitly how performance has changed over time, as well as its components. Among the few studies that have partially addressed these issues we may find Giménez et al. (2007), who considered a cross-country analysis using Data Envelopment Analysis (DEA) to analyze the efficiency and maximum potential output of educational systems for 31 countries with data from TIMSS 1999. Thieme et al. (2011) carried out a similar comparison for the 54 countries participating in PISA 2006, addressing the first two limitations stated in the preceding paragraph; specifically, they use directional distance functions (DDF) for evaluating efficiency indicators which relate outcome variables with resource variables used in the educational process. The authors evaluate jointly *good* (or desirable) outputs of academic achievement with *bad* (or undesirable) outputs due to educational inequality. Their results show that it is feasible for a higher education system to combine high students' learning levels and, simultaneously, obtaining low inequality levels; however, the authors found that in most instances both dimensions required significant improvements.

However, in order to obtain a fuller evaluation of the performance of educational systems it would be desirable to evaluate the *change* in performance over time—which, as suggested above, could be referred to as a third limitation of previous research initiatives. It is important to measure this, since there is a general consensus about how important it is not only to measure students' achievement but also how much they have progressed, and which share is attributable to the educational system itself or to external factors. This particular field of research in education economics refers to them as *growth* studies, requiring at least two evaluations at different points in time.

Therefore, and according to the rationale presented above, some desirable properties of a good education system would relate not only to its ability to to obtain high average students' academic achievement, but also to be able to ensures that all students make progress. For this, it is also necessary to develop strategies that enable relatively disadvantaged students to make progress as well, and achieve basic standards. Therefore, an educational system that evolves satisfactorily will be the one which improves the *average* student's academic achievement while simultaneously minimizing the percentage of students not achieving the most basic learning standards. Similarly, the change in the resources' endowments used by the system will indicate if the changes in the level of educational achievement (either positive or negative) are due to *technical change*, which might be attributable to an improvement of the educational resources available, or to an enhanced *efficiency* when utilizing these resources.

In order to deal with these issues, some studies have been proposing a variety of measures for evaluating performance change over time (either due to efficiency change or technical change), most of them in the spirit of Färe et al. (1994), although related proposals (closer to the ones we will consider here) have also been developed, including Chung et al. (1997), Pastor and Lovell (2005), or Luenberger (1992), among others. In our case, for measuring educational systems' change in performance (for achieving educational objectives), we will model both good and bad outputs, for which we will use the Global Malmquist-Luenberger productivity index (hereafter, GML index), developed by Oh (2010). This index, based on contributions by Luenberger (1992) and Pastor and Lovell (2005), has the ability to correct some of the weaknesses of the Malmquist-Luenberger index (ML index), by solving the problem of linear programming problems' infeasibility when measuring various cross-sections periods using directional distance functions (DDFs).

The GML index is used to measure performance change of educational systems in 28 countries participating in TIMSS 2007 and 2011 for eighth grade of basic education in the discipline of mathematics. Results offer a multiplicity of angles. They can be exploited from an orientation point of view (good and bad outputs, good outputs, or bad outputs) or evaluating the decomposition of the Global Malmquist-Luenberger productivity index into its two components—best practice gap change and efficiency change. In general (on average), results indicate that there has been a deterioration of educational performance between 2007 and 2011, which was mainly driven by an average best practice gap decline, for both good and bad outputs orientation, as well as good output orientation. In the case of the bad output orientation, educational performance has actually improved, but only slightly, and also due to an average best practice gap improvement. We labelled this the *bipartite* decomposition of educational performance, and the ensuing analysis of how the underlying distributions evolved indicated there were remarkable disparities at country level. Therefore, the different countries participating in the study not only chose several paths to improve their educational performance, which we also described but, in addition, results varied remarkably among them.

The paper is organized as follows. After this introduction, Section 2 describes the methodological aspects of Global Malmquist-Luenberger Index (GML) and its decomposition to evaluate the performance of education systems over time. The data used for the analysis of educational systems is presented in Section 3. The main results are presented in Section 4, and Section 5 outlines the main conclusions.

2. Methodology

2.1. Modeling educational performance dynamics

Dynamic efficiency studies often employ the Malmquist index (Caves et al., 1982). This index is used to explain the change in total factor productivity as a result of the change in efficiency or catch-up and technological change. Chung et al. (1997) modified the Malmquist index to apply to the case of directional distance functions (DDF). These have been widely used in studies measuring efficiency incorporating the environmental impact of the units analyzed by considering the bad outputs of the production process (Sueyoshi and Goto, 2011; Watanabe and Tanaka, 2007; Färe et al., 2005). The new index was named Mamlquist-Luenberger.

However, both indices suffer from two problems (Pastor and Lovell, 2005; Oh, 2010). First, circularity is not assured. This property refers to the fact that the change in productivity over a period can be explained from the product of changes in productivity in the different sub-periods within. Secondly, there is the possibility on infeasibilities in the calculation of the cross-distance functions necessary for their calculation. Although it is necessary and sufficient condition that technical change is Hicks-neutral to ensure circularity (Balk, 2001) or that a particular data structure ensures the absence of feasibility problems (Xue and Harker, 2002), compliance of this conditions in empirical applications is often difficult. To remedy both deficiencies, Pastor and Lovell (2005) proposed a modification of the Malmquist index that was called global Malmquist index. Oh (2010) similarly adapted Mamlquist-Luenberger index to achieve the same properties leading to global Mamlquist-Luenberger index (hereafter GML).

This paper uses the global Malmquist-Luenberger index proposed by Oh (2010) for the dynamic analysis of the results obtained by the countries' educational systems. The reasons for the choice of this index would be, apart from their desirable properites, the existence of bad outputs that would make sense to be minimized by educational systems while maximizing the outputs (good outputs). Therefore, in the particular context of education using this index is particularly appropriate.

Let be K countries for which information is available about their educational systems for the years $t = 1 \ldots T$ on M *good* outputs produced, the H *bad* outputs generated from the consumption of N inputs. The production possibility set, is defined by:

$$P(x) = \{(y, b) | x \text{ can produce } (y, b)\} \tag{1}$$

Axioms that must meet the technology described in Equation (1) are the classical proposed by the production theory. See, for instance, Färe et al. (2007) for more details.

The efficiency for a given unit belonging to $P(x)$ can be measured by the following directional distance function (Luenberger, 1992; Sueyoshi and Goto, 2011; Oh, 2010):

$$D\,(x,y,b) \;=\; \max\left(\beta \;\mid\; \left(y+\beta g_y,\; b-\beta g_b\right) \;\in\; P(x)\right) \tag{2}$$

The DDF in Equation (2) above determines the maximum attainable simultaneously increase and decrease (β) in the good and bad of the output over the vector $g = \left(g_y, g_b\right)$, which defines the improvement desirable directions for both types of outputs. In this paper the vector of $M+H$ components $g = (y,\; b)$ is used as suggested by Chung et al. (1997) and Oh (2010).

The GML index for years t and $t+1$ is defined as follows:

$$GML^{t,t+1}\left(x^t, y^t, b^t, x^{t+1}, y^{t+1}, b^{t+1}\right) = \frac{1+D^G\left(x^t, y^t, b^t\right)}{1+D^G\left(x^{t+1}, y^{t+1}, b^{t+1}\right)} \tag{3}$$

where $D^G\,(x,y,b) \;=\; \max\left(\beta \;\mid\; \left(y+\beta g_y,\; b-\beta g_b\right) \;\in\; P^G(x)\right)$ is the DDF defined on the global set of production possibilities $P^G(x)$, that is, the set generated by considering all the observations for t and $t+1$. A value greater than one for $GML^{t,t+1}$ means that there has been improvement in productivity between t and $t+1$, since the distance to the global frontier was greater in t than $t+1$. A value less than the unit for is interpreted contrary.

Expression (3) can be decomposed as follows (Oh, 2010):

$$\begin{aligned} GML^{t,\,t+1}\left(x^t,\; y^t,\; b^t,\; x^{t+1},\; y^{t+1},\; b^{t+1}\right) &= \tfrac{1+D^G\left(x^t, y^t,\, b^t\right)}{1+D^G\left(x^{t+1},\, y^{t+1},\, b^{t+1}\right)} = \\ &= \tfrac{1+D^t\left(x^t, y^t,\, b^t\right)}{1+D^{t+1}\left(x^{t+1},\, y^{t+1},\, b^{t+1}\right)}\; x \left[\frac{1+D^G\left(x^t,\; y^t,\; b^t\right)\Big/1+D^t\left(x^t,\; y^t,\; b^t\right)}{1+D^G\left(x^{t+1},\; y^{t+1},\; b^{t+1}\right)\Big/1+D^{t+1}\left(x^{t+1},\; y^{t+1},\; b^{t+1}\right)}\right] = \\ &= \tfrac{TE^{t+1}}{TE^t}\; x \left[\tfrac{BPG_{t+1}^{t,t+1}}{BPG_t^{t,t+1}}\right] \;=\; EC^{t,\,t+1}\; x\; BPC^{t,\,t+1} \end{aligned} \tag{4}$$

where $EC^{t,\,t+1}$ reflects the change in technical efficiency or *catching-up* between year t and year $t+1$. If $EC^{t,\,t+1} > 1$ it means that there has been improvement in technical efficiency in the period. In other words, the unit is closer to its contemporary frontier in year $t+1$ than in t. A value less than unity is interpreted inversely. The term $BPC^{t,\,t+1}$ is a measure of technological change in the period, that is, of how contemporary frontiers have shifted in the period. Expression (4) shows that the expression for the $BPC^{t,t+1}$ calculation is:

$$BPC^{t,t+1} = \frac{BPG_{t+1}^{t,t+1}}{BPG_{t}^{t,t+1}} = \frac{1 + D^G\left(x^t,\, y^t,\, b^t\right) / 1 + D^t\left(x^t,\, y^t,\, b^t\right)}{1 + D^G\left(x^{t+1},\, y^{t+1},\, b^{t+1}\right) / 1 + D^{t+1}\left(x^{t+1},\, y^{t+1},\, b^{t+1}\right)} \tag{5}$$

where:

$$BPG_{t+1}^{t,t+1} = \frac{1}{1 + D^G\left(x^{t+1},\, y^{t+1},\, b^{t+1}\right) / 1 + D^{t+1}\left(x^{t+1},\, y^{t+1},\, b^{t+1}\right)} \tag{6}$$

Expression (6) shows that $BPG_{t+1}^{t,t+1}$ is the inverse of the ratio between the distance to the global frontier defined by $P^G\,(x)$ and the contemporary frontier defined by $P\,(x)$ at time $t+1$. If $BPC^{t,t+1} > 1$ is because the contemporary frontier in $t+1$ is closer to the global than in t, therefore, there have been technological progress. The case $BPC^{t,t+1} < 1$ represents the opposite situation.

The calculation of $D^G\left(x^t, y^t, b^t\right)$ and $D^t\left(x^t, y^t, b^t\right)$ can be performed using various methods. In this paper we use Data Envelopment Analysis models (Charnes et al., 1978), which have been widely used in efficiency studies (see for an extensive review of its use in the literature Emrouznejad et al., 2008). For the calculation of $D^t\left(x^t, y^t, b^t\right)$ we consider the following linear program (under the assumption $g = (y, b)$) for each country analyzed (Mandal and Madheswaran, 2010):

$$\begin{aligned}
&Max\ \beta \\
&s.t \\
&\textstyle\sum_{k=1}^{K} \lambda_k\, y_{km}^{t} \geq y_m^{ot}\,(1+\beta) \quad m = 1 \ldots M \\
&\textstyle\sum_{k=1}^{K} \lambda_k\, b_{kh}^{t} \leq b_h^{ot}\,(1-\beta) \quad h = 1 \ldots H \\
&\textstyle\sum_{k=1}^{K} \lambda_k\, x_{kn}^{t} \leq x_n^{ot} \quad n = 1 \ldots N \\
&\textstyle\sum_{k=1}^{K} \lambda_k = 1 \\
&\beta \geq 0\ ;\ \lambda_k \geq 0 \quad k = 1 \ldots K
\end{aligned} \tag{7}$$

where β is the maximum achievable increase and decrease simultaneously in both good and bad outputs, respectively, y_{km}^t represents the output m of the unit k in year t, b_{kh}^t the bad output h of the unit or country k in year t, and x_{kn}^t the input n used by the country's education system k in year t. The observed levels of good, bad outputs and inputs for the evaluated country in year t are represented by y_m^{ot}, b_h^{ot} and x_{kn}^{ot}, respectively.

Analogously, for the calculation of $D^G\left(x^t,\, y^t,\, b^t\right)$ the following linear program has to be solved:

$$
\begin{aligned}
& \text{Max } \beta \\
& s.t \\
& \textstyle\sum_{k=1}^{K} \sum_{T=t}^{t+1} \lambda_k^t \, y_{km}^T \;\geq\; y_m^{ot} \; (1+\beta) \quad m = 1 \ldots M \\
& \textstyle\sum_{k=1}^{K} \sum_{T=t}^{t+1} \lambda_k^t \, b_{kh}^T \;\leq\; b_h^{ot} \; (1-\beta) \quad h = 1 \ldots H \\
& \textstyle\sum_{k=1}^{K} \sum_{T=t}^{t+1} \lambda_k^t \, x_{kn}^T \;\leq\; x_n^{ot} \quad n = 1 \ldots N \\
& \textstyle\sum_{k=1}^{K} \sum_{T=t}^{t+1} \lambda_k^T \;=\; 1 \\
& \beta \;\geq\; 0 \;;\; \lambda_k^T \;\geq\; 0 \quad k = 1 \ldots K
\end{aligned}
\tag{8}
$$

In this study, apart from calculating the GML assuming the directional vector $g = (y, b)$, two alternative directional vectors are explored: $g = (y, 0)$ and $g = (0, b)$. With these additional calculations are quantified productivity changes in two mew directions: one prioritizing increases only in the *good* outputs and other prioritizing good decreasing only in the *bad* outputs. This information is interesting to test whether countries have moved in the period analyzed in the direction in which higher *potential* earnings could be achieved. Obviously, other directional vectors could have also been explored, as none of explored vectors have to represent the optimal movement towards the frontier. But certainly, it is relevant to ascertaining whether the objectives of an educational system should focus on improving *both* types of outputs simultaneously or only one of them. This should be an interesting question for the definition of any educational policy.

In the existent literature, one can find different approaches to integrate the undesirable outputs in the efficiency estimations. Probably, the most popular is to consider the bad outputs as *weakly disposable* (basically modifying the restrictions in order to accept proportional reductions in the bad as well as in the good outputs). For more details on this option see Färe et al. (1989) and Färe and Grosskopf (2004). However, the debate pointing out the problems and the solutions of this option is far to be ove. See, for instance, Kuosmanen (2005), Kuosmanen and Podinovski (2009), Färe and Grosskopf (2009), or Picazo-Tadeo and Prior (2009), among others. Another possibility is to convert the undesirable bad outputs into *desirable* (i.e. strongly disposable) good outputs, as Golany and Roll (1989) and Seiford and Zhu (2002), but this conversion may influence significant changes in the level of efficiency found. Finally, following Reinhard et al. (2002) and Hailu and Veeman (2001), perhaps the most intuitive option is to consider the bad outputs as strongly disposable inputs. Because of its simplicity, this has been the selected option in our proposal.

2.2. Bipartite decomposition of the relative contributions to educational performance

According to the expressions above, the Global Malmquist-Luenberger (GML) index is decomposed into technical change (EC) and best practice gap change (BPC). Apart from analyzing how the different components contribute to the overall change of GML *on average*, we can also consider a distribution dynamics approach for analyzing which the largest contributors to the variation in performance are, as measured by GML between periods t (2007) and $t+1$ (2011). For this, we will employ nonparametric density estimation, based on kernel smoothing.

For this, we rewrite expression (3) above as follows:

$$gml^{EC\times BPC} = EC^{t,t+1} \times BPC^{t,t+1} \tag{9}$$

according to which we will use expression $gml^{EC\times BPC}$ to indicate that the change in educational achievement is obtained by successively multiplying its three components. This in turn, enables the possibility for constructing counterfactual distributions by sequential introduction of each of the factors. Specifically, the counterfactual educational achievement change attributable to changes in efficiency would be:

$$gml^{EC} = EC^{t,t+1} \tag{10}$$

which isolates the effect on the distribution of changes due to efficiency only, assuming BPC does not contribute to the change in educational achievement (gml).

Analogously, and if we extend this sequential decomposition, we would proceed as follows:

$$\begin{aligned} gml^{EC\times BPC} &= EC^{t,t+1} \times BPC^{t,t+1} \\ &= gml^{EC} \times BPC^{t,t+1} \end{aligned} \tag{11}$$

We will refer to the decomposition in (11) as the <u>bipartite decomposition</u> of the <u>relative contributions to the changes in the distribution of educational performance</u>.

3. Data, inputs and outputs

This study uses information from the education systems of 28 countries participating in TIMSS 2007 and 2011. TIMSS is named after *Trends in International Mathematics and Science Study* developed by the International Association for the Evaluation of Educational Achievement (IEA). Its purpose is to measure learning achievement of students at the end of 4^{th} and 8^{th} grade in basic mathematics and science. It takes place every four years, and the 2011 edition corresponds to the fifth version of the study. Its design allows for comparisons *over time* and *across countries* participating in the study. In this study, we consider information corresponding to 8th grade in mathematics. For this particular grade, the number of countries participating in the 2011 study involved 42 countries, and 50 in 2007.

In all countries, the sample of schools and students is selected by the IEA at country level, representing each of the grades under analysis. In general, for every country in the sample we have approximately 4,000 students from 150 to 200 schools, for each of the assessed grades. Additionally, each country can apply for a larger sample size should it be interested in a particular type of segmentation (by type of administration, location, etc.). TIMSS also collects information on principals, teachers and students, which enables setting a framework to analyze the results corresponding to the learning process. This is done from an evaluation framework agreed between the participating countries in the study, so that it does not necessarily compile the same information from these actors in each version of the study.

As noted previously, the methodology described in the preceding section is used to evaluate the *change* in the performance of educational systems for achieving their goals. For this, following the extant literature (Carlson, 2001) we consider that a good education system is not only the one which allows obtaining a high degree of *average* academic achievement for their students, but also the one ensuring that all students make *progress*. For this, it is also necessary to develop strategies enabling more disadvantaged students to make progress and achieve basic educational levels. Therefore, an educational system which makes a satisfying progress will be the one enabling not only an *average* improvement in the academic achievement of their students but also the one *minimizing* the percentage of students which cannot achieve basic learning standards. Analogously, changes in the resource endowments used will indicate if the changes in the achievement of the educational goals (either positive or negative) are due to *technical change* (which might be originated by the higher educational resource endowments) or to an improved *efficiency* in their usage.

In this regard, the TIMSS reports enable us to have standardized information both on educational outcomes and resources used for this purpose in at least two points in time. For this particular study we used the reports corresponding to years 2007 and 2011. In the case of learning outcomes in both mathematics and science, they are reported by TIMSS in two ways for each participating country: (i) on a scale which ranges from 0 to 1,000, with an international average standardized of 500 points, and a standard deviation of 100 points. This average corresponds to the set of countries participating in the first edition of TIMSS in 1995, and was set as the benchmark for comparability between years; (ii) for four performance levels describing different learning levels achieved by students. The *advanced* level corresponds to students with more than 625 points, the *high* level corresponds to students with more than 550 points and less than 625 points, the *intermediate* level corresponds to students with more than 475 and less than 550 points, and the *low* or basic level to students with more than 400 and less than 475 points. Analogously, we report the percentage of students failing to achieve this basic standard and, therefore, is out of the range.

Accordingly, we have defined as good output (y_1) the average achievement of each country for its 8$^{\text{th}}$ year students of basic education for the mathematics discipline, and as bad output (y_2) the percentage of students from each country that achieve the basic standards, corresponding to the same grade and discipline. For both indicators we have data for 2007 (time t) and 2011 (time $t+1$). The average values, corresponding to both outputs for each country for years 2007 and 2011, are reported in the first four columns of Table 1.

As a reference, in 8$^{\text{th}}$ year of basic (primary) education of TIMSS 2011, only 14 countries (out of 42) obtained scores higher than 500 in the TIMSS scale, whereas in 2007 only 12 countries (out of 50) had scores higher than such a value. In this subject (mathematics) and primary education degree, in 2011 the highest average score corresponded to the Republic of South Korea (613 points), followed by Singapore (611 points), Chinese Taipei (609 points) and Hong Kong (586 points). In 2007, the highest academic achievements were obtained by Chinese Taipei (598 points), followed by the Republic of Korea (597 points), Singapore (593 points), Hong Kong (572 points) and Japan (570 points). Out of the 28 participating countries in both years, the highest progress corresponded to the Palestinian National Authority (37 points increase), followed by Italy (35 points increase). In contrast, the largest decline corresponded to Malaysia (34 points decrease between both years), followed by Jordan (21 points decline).

Meanwhile, for year 2011, in 8*th* year mathematics, the average distribution of participating countries indicates that out of the 18% of students who did not reach the minimum standards, 21% corresponded to the lowest level, 30% to students of intermediate level, 22% to students of high level, and only 9% to students of advanced level. The percentage of students not reaching the minimum standards is very heterogeneous across countries. Among the 10 countries with the highest scores, this percentage is 2% (on average), whereas for those 10 countries with the lowest scores the percentage is 55%.

In our sample, the highest declines in the percentage of backward students corresponds to Italy, Georgia and the Palestinian National Authority—these countries reduced this percentage by 10% between 2007 and 2011. In contrast, Malaysia was the country with a highest increase in the percentage of students below the minimum standards (from 50% to 64%).

The inputs of the model correspond to two variables for which there was available information for both years, namely, learning hours in mathematics during the academic year (x_1), and teachers' quality, measured as the percentage of students whose teachers feel "very well" trained for teaching mathematics (x_2). The average values corresponding to both inputs, for each country in our sample, and for years 2007 and 2011 are reported in the last four columns of Table 1.

The hours corresponding to of 8th grade mathematics range from 76 hours (Syria) and 158 hours (Chinese Taipei) for the year 2007, and 97 hours (Sweden) and 173 hours (Indonesia) for 2011. The sample average is 135 hours for 2011 and 120 hours for 2007. The country that increased the number of teaching hours the most is Bahrain (46 hours), followed by Syria (42 hours). Only two countries decreased teaching time: Jordan (11 hours) and Hong Kong (10 hours).

The data indicate that the best teaching quality corresponds to teachers in the United States, England, Georgia and Romania, with 94% of students whose teachers feel "very well" prepared for teaching mathematics. The lowest value for this resource corresponds to Indonesia (54%), followed by Thailand (55%). The country with the greatest improvement in this indicator is Lithuania (23% increase), followed by Japan (16% increas). In contrast, the country experiencing a sharper decline in its teaching quality was Indonesia (27% decrease), followed by Ukraine (18% decrease).

The average values corresponding to the inputs and outputs of the model for each country in 8th grade of mathematics for years 2007 and 2011 used in the analysis of global index Malmquist-Luenberger are reported in Table 1.

4. Results

4.1. Efficiency change, best practice gap change, and performance change: analysis based on summary statistics

Tables 2, 3 and 4 show the results of the global Malmquist-Luenberger index and its decomposition for the countries in our sample and for the different directions selected, i.e. good and bad output orientation (Table 2), good output orientation (Table 3), and bad output orientation (Table 4).

Results vary remarkably in two dimensions, namely, for the two components of the Malmquist-Luenbger global index (efficiency change and best practice gap), and for the different orientations chosen—good and bad output orientation, good output orientation, and bad output orientation.

Comparing the results in Tables 2, 3, and 4, from the three bottom rows it is quite apparent that the differences are remarkable considering both dimensions of variability. On average (for both the arithmetic and geometric means), the GML index shows a *mean* deterioration in educational performance when considering either the *good and bad* output orientation, and the *good* output orientation (Tables 2 and 3) for the countries in our sample. However, when focusing on the *bad* output orientation (Table 4), *on average*, there is virtually no change in educational performance—the means are 1.0030 for the arithmetic mean and 1.0008 for the geometric mean, respectively.

Analyzing the components of the GML index, i.e. the efficiency change (EC) and the best practice gap (BPC), there are also different results when considering, on the one hand, good and bad output orientation and good output orientation and, on the other hand, bad output orientation. Whereas in the former case efficiency change (EC) contributes *positively* to overall performance change, in the latter case the contribution is *negative*. In contrast, the best practice gap (BPC) shows an opposite result—the contribution is negative, in the case of good output and good and bad output orientation, and positive in the case of the bad output orientation.

However, these are *average* results, concealing very heterogeneous findings at the country level. This is partly shown by the standard deviation values, which are particularly high for the efficiency change in the case of the good and bad and good output orientations (Tables 2 and 3), with values of 0.1706 and 0.1851 (see the first columns in Tables 2 and 3, respectively). In contrast, for the same orientations, the dispersion for the best practice gap is much lower, with values of 0.0678 and 0.0613 (see the second columns in Tables 2 and 3, respectively). Combining both components of performance change, the standard deviation for each orientation is 0.1145 and 0.1272—see the third column in Tables 2 and 3, respectively.

In the case of the bad output orientation (Table 4), there is a greater balance, in terms of dispersion across countries, between both components of performance change (0.0610 and 0.0625 for EC and BPC, respectively), resulting into a 0.0674 value corresponding to the standard deviation for the global Malmquist-Luenberger index.

However, despite the more moderate dispersion values found for both EC and GML in the case of the bad output orientation (compared to the other two orientations), results differ remarkably for some specific countries. For instance, as indicated in the third column of Table 4, (positive) change in performance has been substantial for Chinese Taipei and Singapore; however, the reasons were not coincidental—for the former it was mostly due to efficiency change (whose value was $EC^{t,t+1}_{\text{Chinese Taipei}} = 1.2845$), whereas for the latter best practice gap change played a major role ($BPC^{t,t+1}_{\text{Singapore}} = 1.1591$). In contrast, Japan has experienced a remarkable deterioration in educational performance ($GML^{t,t+1}_{\text{Japan}} = 0.7671$), due to best practice practice gap decline ($BPC^{t,t+1}_{\text{Japan}} = 0.7671$), whereas efficiency stagnated ($EC^{t,t+1}_{\text{Japan}} = 1.0000$).

This multiplicity of different cases is even higher when focusing the analysis in the case of the good and bad output orientation (Table 2) and the case of good output orientation (Table 3), although in the particular case of bad output orientation virtually *all* countries showed best practice decline ($BPC^{t,t+} < 1$), with the exception of Korea (for which $BPC^{t,t+}_{\text{Korea}} < 1$).

These variety of paths through which countries' educational performance evolves is difficult to summarize in two statistics only (standard deviation and mean, either arithmetic or geometric), making it difficult to explore in detail the results achieved. An informative complement consists of applying the bipartite decomposition of performance change based on kernel density estimation proposed in Section 2.2, as we shall see below.

4.2. Classifying educational systems: shall we stress excellence or reduce inequalities

Table 5 shows a classification of the countries evaluated considering both the results of the GML index and the assessment orientation. The first group of countries ($G1$), composed of Italy and Singapore, are those that have improved in the period under any orientation. We refer to this group as "overall improvement". The second group ($G2$) is composed by countries with a positive evolution in the period when considering the bad-outputs orientation only. Therefore, they are countries that appear to have concentrated their efforts on reducing the number of students below acceptable levels and, consequently, they have focused on reducing inequality in their educational systems. This group has been labeled "inequality improvement" and is composed of Australia, Bahrain, Chinese Taipei, Georgia, Hong Kong, Iran, Norway, Oman, Palestinian, Slovenia, Tunisia and United States.

The next group ($G3$) is formed by Jordan and Indonesia. These countries are characterized because their results suggest that they have mainly focused on improving the average performance of their students over the 2007–2011 period; we can draw this conclusion due to the favourable evaluation they obtain only when the evaluation is under an orientation that prioritizes good-outputs. This group has been called "average achievement". The fourth group ($G4$)

named "simultaneous improvement" is composed of those countries whose results suggest that their efforts have been directed towards achieving improvements in both average achievement and inequality simultaneously. Countries with these characteristics are Ukraine and Ghana.

The *G*5 group clusters those countries that have maintained high levels of efficiency in both periods. This group has been named "stable", and its only member is South Korea. Finally, there is a group (*G*6) of countries which, regardless of their orientation, have gone through performance decline during the analysed period; this group is constituted by England, Hungary, Lithuania, Malaysia, Romania, Sweden, Syrian and Thailand, and we have labelled them as "Decline".

4.3. Bipartite decomposition of performance change

The results of the analysis proposed in Section 2.2 are reported in Figures 1, 2, 3 and 4. Each figure is divided in three panels, and each panel contains two sub-figures. The upper panels correspond to the analysis of performance change considering a good and bad output orientation, the central panel reports results for the good output orientation, and the lower panel refers to bad output orientation. Each panel is also divided in two sub-figures, in order to provide a sequential analysis of the contribution of each component of performance change. The sequential order is shown in both directions (Figure 1 vs. 2, and Figure 3 vs. 4). The vertical line in each figure corresponds to the (arithmetic) mean of the underlying density.

Given some of the particularities of the data used, the densities have also been estimated for different values of the smoothing parameter, or bandwidth (h), which tunes the amount of *bumps* under each curve—higher values of h tend to smooth more, revealing less particularities of the data, low values of h tend to smooth less, providing more detail but generating (in some cases) fuzzy graphics. Specifically, Figures 1 and 2 report results for a *global* bandwidth (the amount of smoothing is the same at all data points), for which we followed the proposals by Sheather and Jones (1991). In the case of Figures 3 and 4, the amount of smoothing varies *locally*, depending on the structure of the data at a given point, for which we followed Loader (1996).

The analysis in the upper panel of Figure 1 shows that, when considering a good and bad output orientation, the contribution of efficiency to the change of the global Malmquist-Luenberger index is very heterogeneous, as indicated by the several bumps shown by density corresponding to gml^{EC} (Figure 1.a). However, the contribution of the best practice change, shown in Figure 1.b, offsets the heterogeneity of gml^{EC}, leading to a much smoother density when both effects are combined ($gml^{EC \times BPC}$). Actually, *on average*, as indicated by the vertical lines in Figures 1.a. and 1.b, although the effect of efficiency change (gml^{EC}) is positive (the solid vertical line is above 1), the contribution of the best practice gap leads to a *negative* combined effect (the dashed vertical line is below 1). The smoother lines depicted when choosing local bandwidths, as shown in Figures 3.a and 3.b point in the same direction, excepting for the bumps corresponding to gml^{EC}, which are smoothed out in Figure 3.a.

These discrepancies are also present when reversing the sequential order, as shown in the upper panel of Figure 2 (Figures 2.a and 2.b), for the global bandwidth, and the upper panel of Figure 4, for the local bandwidth (Figures 4.a and 4.b). Conducting the analysis in the reverse order indicates that the discrepancies for gml^{BPC} are even higher than those for gml^{EC}; this is particularly apparent when choosing a global bandwidth. Therefore, countries follow very different paths to obtain their productivity change index.

Results change when considering either a good output orientation (Figures 1.c and 1.d for the global bandwidth, and Figures 3.c and 3.d for the local bandwidth), as well as a bad output orientation (Figures 1.e and 1.f for the global bandwidth, and Figures 3.e and 3.f for the local bandwidth). Computations have also been performed reversing the direction of causality (see lower panels of Figures 2 and 4.

Regarding the good output orientation, when considering a global bandwidth there are remarkable disparities across countries for gml^{BPC} (see Figures 1.c and 1.d) to the point that, on average, the contribution of the best practice gap change is *negative*. In contrast, the probability mass corresponding to the contribution of efficiency change is positive—much of the density is above 1 (see Figures 1.c and 3.c). However, when considering the bad output orientation, discrepancies are remarkable for both components of of the educational performance index. Although, according to the local bandwidth figures (Figures 3.e, 3.f, 4.e and 4.f), probability concentrates tightly around unity, this effect is partly derived from the choice of bandwidths, since choosing global bandwidths results in much fuzzier graphics (Figures 1.e, 1.f, 2.e and 2.f).

5. Conclusions

In this paper, we have considered some relatively recent proposals for analyzing educational performance and how it changes over time. Specifically, we have taken advantage of the latest release of the tests for analyzing Trends in International Mathematics and Science Study (TIMSS). This initiative provides a framework for evaluating and describing the learning processes of students in participating countries, for the two disciplines analyzed, providing relevant information on additional factors which also these processes.

For this we have considered the global Malmquist-Luenberger index (GML), which is particularly interesting in the context of education. Specifically, it is not only appropriate for its very desirable properites. In also suits our context due to the existence of bad outputs which, ideally, should be minimized by educational systems, while simultaneously maximizing the outputs—or, more properaly, good/desirable outputs.

The results of the different evaluations of the GML index show, on average, a deterioration on educational systems' performance when considering both a simultaneous good and bad outputs orientation, as well as when considering a good output orientation only. However, assessing performance when adopting a bad outputs orientations only shows, on average, a slight improvement in performance for the countries in the sample.

Similarly, the average results reveal the variety of emphases for the different of national education systems during the analyzed period (2007–2011). In general (on average) we observe there is a clear preference for technological changes aimed at obtaining higher levels of educational equality, in contrast with emphasizing efficiency improvements when pursuing academic achievement (excellence) and equality simultaneously (i.e. good and bad outputs orientation), or academic achievement (good outputs orientation) exclusively.

This is consistent with: (i) an increasing concern of the different countries in the sample in order to enhance the quality of human capital among the population as equally as possible; (ii) the budgetary constraints that countries faced during most of the analyzed period; and (iii) the different levels in the learning curve of the educational systems with respect to the different objectives pursued (or orientations adopted). Therefore, after years of public policies aimed at achieving improvements in overall academic achievement and equality, or only average academic achievement, it is not entirely surprising the tendency to adopt policies aimed towards achieving efficiency improvements. This contrasts with the necessary commitment to stress technological improvements when there is not a knowledge base facilitating performance improvements via enhanced efficiency.

Apart from the *average* results, there are remarkable discrepancies in the results pointing in two directions. First, there is a remarkable heterogeneity in the results corresponding to the different components of the GML index with respect to the different emphases (orientations) evaluated. The dispersion of results is especially high among countries for the efficiency change (EC) when adopting good and bad outputs as well as good outputs orientations (0.1706 and

0.1851, respectively); however, it is smaller when emphasizing bad outputs (0.0610). Moreover, for the three different orientations the dispersion observed for countries' performance is low regarding the technological change component (BPC). This would have a twofold implication: (i) the differences in overall performance (GML) among countries is mainly driven by changes in efficiency instead of technological change; and (ii) changes in equality are more difficult to achieve and, probably, they require a longer time horizon.

Second, this heterogeneity is also reflected in the ranking of countries. In this classification we may highlight the fact that only two countries (Italy and Singapore) improve their performance considering the three possible orientations, or emphases. However, only Singapore adopts this strategy considering both efficiency improvements *and* technological change. Italy carries it out via improvements in efficiency exclusively. It is also interesting noting that the most populated group corresponds to that constituted only by those countries which only improve in the area of equality (bad output orientation). This is consistent with the new challenges facing educational systems. This group is composed of both high and low academic achievement countries; in the former group we can find Australia, Chinese Taipei, Hong Kong, Norway, United States and Slovenia, whereas in the latter we would include Bahrain, Georgia, Iran, Oman, Palestinian, Tunisia. This constitute evidence for the transversality of this priority. Finally, it is also noteworthy the presence of high academic achievement countries such as England, Hungary and Sweden in the group of countries with performance decline. However, this outcome is coincidental with TIMSS results indicating that, during the evaluation period, all these countries experienced deteriorations in their academic achievements, while increasing the percentage of students who failed the minimum standards.

References

Balk, B. M. (2001). Scale efficiency and productivity change. *Journal of Productivity Analysis*, 15(3):159–183.

Carlson, D. (2001). Focusing state educational accountability systems: four methods for judging school quality and progress. Technical report, Center for Assessment (NCIEA), Dover, NH.

Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6):1393–1414.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.

Chung, Y. H., Färe, R., and Grosskopf, S. (1997). Productivity and undesirable outputs: a directional distance function approach. *Journal of Environmental Management*, 51:229–240.

Denvir, B. and Brown, M. (1986). Understanding of number concepts in low attaining 7–9 year olds: Part i. development of descriptive framework and diagnostic instrument. *Educational Studies in Mathematics*, 17(1):15–36.

Emrouznejad, A., Parker, B. R., and Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, 42(3):151–157.

Ercikan, K. (2006). Examining guidelines for developing accurate proficiency level scores. *Canadian Journal of Education*, 29(3):823–838.

Färe, R. and Grosskopf, S. (2004). Modeling undesirable factors in efficiency evaluation: Comment. *European Journal of Operational Research*, 157(1):242–245.

Färe, R. and Grosskopf, S. (2009). A comment on weak disposability in nonparametric production analysis. *American Journal of Agricultural Economics*, 91(2):535–538.

Färe, R., Grosskopf, S., Lovell, C. A. K., and Pasurka, C. (1989). Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. *The Review of Economics and Statistics*, 71(1):90–98.

Färe, R., Grosskopf, S., Noh, D.-W., and Weber, W. W. (2005). Characteristics of a polluting technology: theory and practice. *Journal of Econometrics*, 126(2):469–492.

Färe, R., Grosskopf, S., and Pasurka Jr, C. A. (2007). Environmental production functions and environmental directional distance functions. *Energy*, 32(7):1055–1066.

Färe, R., Grosskopf, S., Norris, M., and Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1):66–83.

Giménez, V., Prior, D., and Thieme, C. (2007). Technical efficiency, managerial efficiency and objective-setting in the educational system: An international comparison. *Journal of the Operational Research Society*, 58(8):996–1007.

Golany, B. and Roll, Y. (1989). An application procedure for DEA. *Omega*, 17(3):237–250.

Kuosmanen, T. (2005). Weak disposability in nonparametric production analysis with undesirable outputs. *American Journal of Agricultural Economics*, 87:1077–1082.

Kuosmanen, T. and Podinovski, V. V. (2009). Weak disposability in nonparametric production analysis: Reply to Färe and Grosskopf. *American Journal of Agricultural Economics*, 91:539–545.

Loader, C. R. (1996). Local likelihood density estimation. *The Annals of Statistics*, 24(4):1602–1618.

Luenberger, D. (1992). New optimality principles for economic efficiency and equilibrium. *Journal of Optimization Theory and Applications*, 75(2):221–264.

Mandal, S. K. and Madheswaran, S. (2010). Environmental efficiency of the Indian cement industry: an interstate analysis. *Energy Policy*, 38(2):1108–1118.

Oh, D. (2010). A global Malmquist-Luenberger productivity index. *Journal of Productivity Analysis*, 34(3):183–197.

Pastor, J. T. and Lovell, C. A. K. (2005). A global Malmquist productivity index. *Economics Letters*, 88(2):266–271.

Picazo-Tadeo, A. J. and Prior, D. (2009). Environmental externalities and efficiency measurement. *Journal of Environmental Management*, 90(11):3332–3339.

Reinhard, S., Lovell, C. A. K., and Thijssen, G. J. (2002). Analysis of environmental efficiency variation. *American Journal of Agricultural Economics*, 84:1054–1065.

Seiford, L. M. and Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, 142(1):16–20.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, Ser.B,53(3):683–690.

Sueyoshi, T. and Goto, M. (2011). Measurement of returns to scale and damages to scale for dea-based operational and environmental assessment: How to manage desirable (good) and undesirable (bad) outputs? *European Journal of Operational Research*, 211(1):76–89.

Teddlie, C. and Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. Routledge, London.

Thieme, C., Giménez, V., and Prior, D. (2011). A comparative analysis of the efficiency of national educational systems. *Asia Pacific Education Review*, En prensa.

Watanabe, M. and Tanaka, K. (2007). Efficiency analysis of Chinese industry: A directional distance function approach. *Energy Policy*, 35(12):6323–6331.

Xue, M. and Harker, P. T. (2002). Note: ranking DMUs with infeasible super-efficiency DEA models. *Management Science*, 48(5):705–710.

Table 1: Descriptive statistics on inputs and outputs based on TIMMS (2007 and 2011)

Country	Good output (y_1)		Bad output (y_2)		Input 1 (x_1)		Input 2 (x_2)	
	Academic achievement, 2007	Academic achievement, 2011	2007	2011	Mathematics (learning hours) 2007	Mathematics (learning hours) 2011	Teaching quality, 2007	Teaching quality, 2011
Australia	496	505	39	37	131	143	91	91
Bahrain	398	409	81	74	96	142	88	88
Chinese Taipei	598	609	14	12	158	166	74	72
England	513	507	31	35	113	116	95	94
Georgia	410	431	74	64	110	123	86	94
Ghana	309	331	96	95	146	165	85	87
Hong Kong (S.A.R.)	572	586	15	11	148	138	67	82
Hungary	517	505	31	35	99	119	89	86
Indonesia	397	386	81	85	136	173	81	54
Iran	403	415	80	74	99	124	78	82
Italy	463	498	46	36	136	155	65	64
Japan	570	570	13	13	105	108	51	67
Jordan	427	406	65	74	141	130	89	84
Korea, Rep. of	597	613	10	7	104	137	70	79
Lithuania	506	502	35	36	116	132	70	93
Malaysia	474	440	50	64	123	123	79	83
Norway	469	475	52	49	113	125	87	85
Oman	372	366	86	84	150	161	84	87
Palestinian Nat'l Auth.	367	404	85	75	100	134	86	86
Romania	461	458	54	56	122	145	87	94
Singapore	593	611	12	8	124	138	82	86
Slovenia	501	505	35	33	113	121	79	88
Sweden	491	484	40	43	93	97	79	87
Syrian Arab Rep.	395	380	83	83	76	118	74	79
Thailand	441	427	66	72	124	129	47	55
Tunisia	420	425	79	75	126	131	87	78
Ukraine	462	479	54	47	130	132	90	72
United States	508	509	33	32	148	157	93	94

Table 2: Educational improvement, good and bad output orientation (2007–2011)

Country	Efficiency change $EC^{t,t+1}$	Best practice gap change $BPC^{t,t+1}$	Global Malmquist-Luenberger index $GML^{t,t+1}$
Australia	1.0145	0.9296	0.9430
Bahrain	0.7738	0.9739	0.7536
Chinese Taipei	1.1229	0.9482	1.0647
England	1.0471	0.9194	0.9627
Georgia	1.0225	0.9194	0.9401
Ghana	0.9866	1.0166	1.0029
Hong Kong (S.A.R.)	1.0267	0.9791	1.0052
Hungary	0.8839	0.9194	0.8126
Indonesia	1.5435	0.7627	1.1773
Iran	0.8942	0.9226	0.8250
Italy	1.4351	0.7612	1.0924
Japan	1.0000	0.9337	0.9337
Jordan	1.0643	0.9287	0.9884
Korea, Rep. of	1.0000	1.0000	1.0000
Lithuania	0.9291	0.9194	0.8542
Malaysia	1.0006	0.9194	0.9200
Norway	0.9958	0.9194	0.9156
Oman	0.9884	1.0054	0.9937
Palestinian Nat'l Auth.	0.8935	0.9540	0.8524
Romania	0.9092	0.9490	0.8628
Singapore	1.0677	1.0180	1.0870
Slovenia	1.0239	0.9194	0.9413
Sweden	1.0279	0.9194	0.9451
Syrian Arab Rep.	0.6739	0.9204	0.6203
Thailand	1.0870	0.7612	0.8274
Tunisia	1.1030	0.9028	0.9957
Ukraine	1.2631	0.8394	1.0603
United States	1.0360	0.9338	0.9675
Arithmetic mean	1.0291	0.9213	0.9409
Geometric mean	1.0163	0.9187	0.9337
Standard deviation	0.1706	0.0678	0.1145

Table 3: Educational improvement, good output orientation (2007–2011)

Country	Efficiency change $EC^{t,t+1}$	Best practice gap change $BPC^{t,t+1}$	Global Malmquist-Luenberger index $GML^{t,t+1}$
Australia	1.0145	0.9293	0.9427
Bahrain	0.7564	0.9329	0.7057
Chinese Taipei	1.1909	0.9224	1.0984
England	1.0471	0.9194	0.9627
Georgia	1.0225	0.9194	0.9401
Ghana	1.1810	0.8159	0.9636
Hong Kong (S.A.R.)	1.0815	0.9531	1.0308
Hungary	0.8839	0.9194	0.8126
Indonesia	1.6152	0.7612	1.2294
Iran	0.8942	0.9226	0.8250
Italy	1.4351	0.7612	1.0924
Japan	1.0000	0.9337	0.9337
Jordan	1.1081	0.9266	1.0268
Korea, Rep. of	1.0000	1.0000	1.0000
Lithuania	0.9291	0.9194	0.8542
Malaysia	1.0006	0.9194	0.9200
Norway	0.9958	0.9194	0.9156
Oman	1.1071	0.8329	0.9221
Palestinian Nat'l Auth.	0.8935	0.9278	0.8290
Romania	0.9092	0.9261	0.8420
Singapore	1.1576	0.9767	1.1306
Slovenia	1.0239	0.9194	0.9413
Sweden	1.0279	0.9194	0.9451
Syrian Arab Rep.	0.6739	0.9204	0.6203
Thailand	1.0870	0.7612	0.8274
Tunisia	1.1030	0.9028	0.9957
Ukraine	1.2631	0.8394	1.0603
United States	1.0507	0.9069	0.9528
Arithmetic mean	1.0519	0.9003	0.9400
Geometric mean	1.0369	0.8982	0.9313
Standard deviation	0.1851	0.0613	0.1272

Table 4: Educational improvement, bad output orientation (2007–2011)

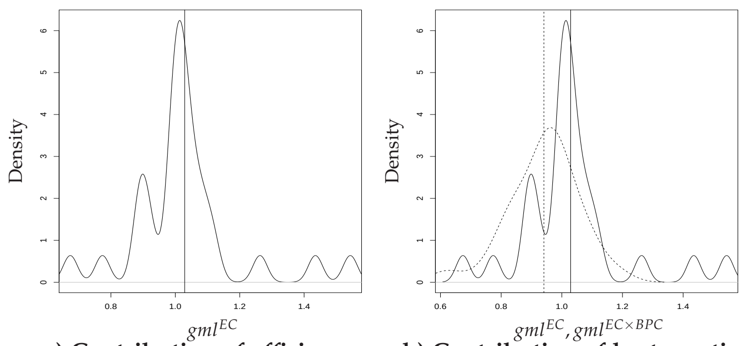
Country	Efficiency change $EC^{t,t+1}$	Best practice gap change $BPC^{t,t+1}$	Global Malmquist-Luenberger index $GML^{t,t+1}$
Australia	0.9690	1.0379	1.0058
Bahrain	0.9901	1.0137	1.0036
Chinese Taipei	1.2845	0.8922	1.1460
England	0.9391	1.0479	0.9840
Georgia	0.9917	1.0155	1.0071
Ghana	0.9928	1.0088	1.0015
Hong Kong (S.A.R.)	0.9774	1.0551	1.0312
Hungary	0.9376	1.0181	0.9546
Indonesia	0.9845	1.0136	0.9979
Iran	0.9895	1.0140	1.0034
Italy	0.9963	1.0279	1.0242
Japan	1.0000	0.7671	0.7671
Jordan	0.9755	1.0185	0.9936
Korea, Rep. of	1.0000	1.0000	1.0000
Lithuania	0.9549	1.0438	0.9968
Malaysia	0.9582	1.0274	0.9845
Norway	0.9786	1.0260	1.0041
Oman	0.9884	1.0120	1.0002
Palestinian Nat'l Auth.	0.9944	1.0119	1.0063
Romania	0.9740	1.0245	0.9979
Singapore	1.0394	1.1591	1.2048
Slovenia	0.9644	1.0434	1.0062
Sweden	1.0055	0.9790	0.9844
Syrian Arab Rep.	0.9859	1.0049	0.9907
Thailand	0.9692	1.0042	0.9733
Tunisia	0.9874	1.0148	1.0021
Ukraine	0.9857	1.0246	1.0099
United States	0.9581	1.0471	1.0032
Arithmetic mean	0.9919	1.0126	1.0030
Geometric mean	0.9903	1.0106	1.0008
Standard deviation	0.0610	0.0625	0.0674

Table 5: Classification of countries according to their educational achievements

Group ID	Strategy/achievement	Countries in the group
$G1$	Overall improvement	Italy, Singapore
$G2$	Inequality improvement	Australia, Bahrain, Chinese Taipei, Georgia, Hong Kong, Iran, Norway, Oman, Palestinian, Slovenia, Tunisia, United States
$G3$	Average achievement	Jordan, Indonesia
$G4$	Simultaneous improvement	Ukraine, Ghana
$G5$	Stable	Korea
$G6$	Decline	England, Hungary, Lithuania, Malaysia, Romania, Sweden, Syrian, Thailand

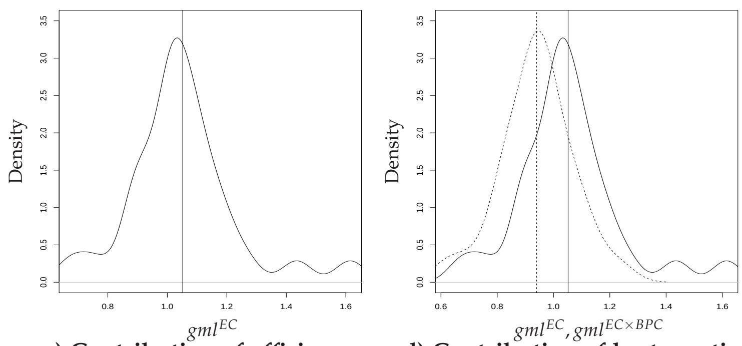
Figure 1: Kernel density plots of the bipartite decomposition of educational improvement, good and bad output orientation

Good and bad output orientation



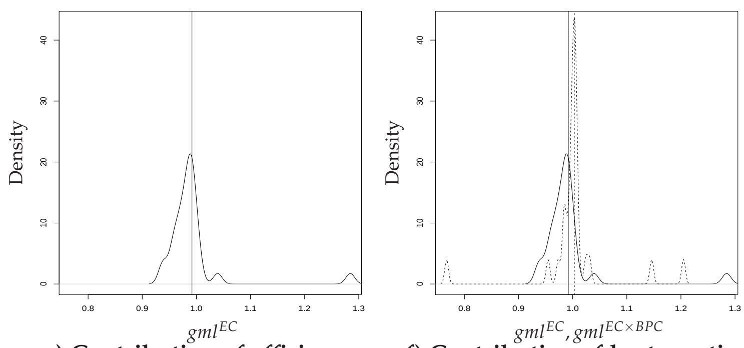
a) Contribution of efficiency change **b) Contribution of best practice gap change**

Good output orientation



c) Contribution of efficiency change **d) Contribution of best practice gap change**

Bad output orientation



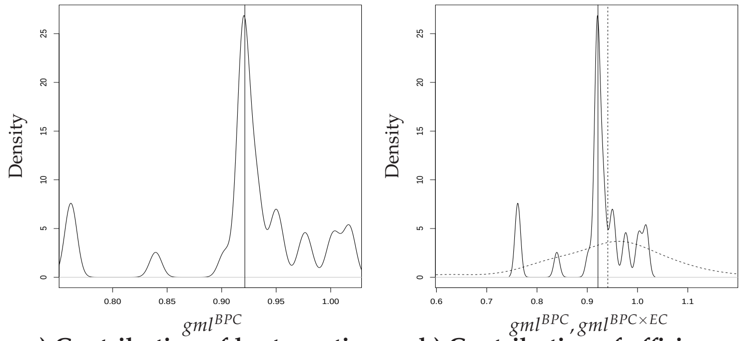
e) Contribution of efficiency change **f) Contribution of best practice gap change**

gml^{EC} ——— $gml^{EC\times BPC}$ _ _ _ _ _ _

Notes: All figures contain densities estimated using kernel density estimation for the different components of the biipartite decomposition in expression (11), considering the good and bad output orientation. The vertical lines in each plot represents the average for each component of the decomposition. Densities were estimated using a Gaussian kernel and the Sheather and Jones (1991) plug-in bandwidth.

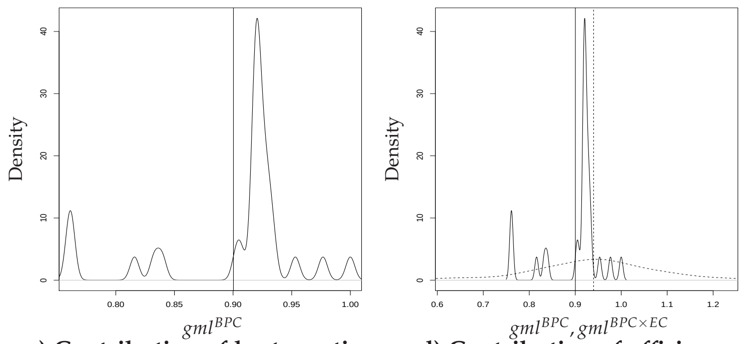
Figure 2: Kernel density plots of the bipartite decomposition of educational improvement, good and bad output orientation

Good and bad output orientation



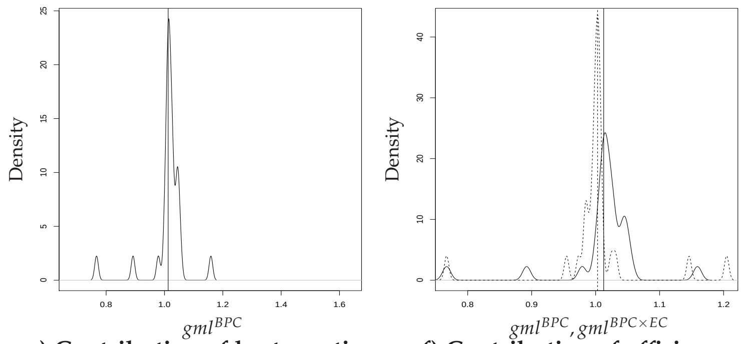
a) Contribution of best practice gap change **b) Contribution of efficiency change**

Good output orientation



c) Contribution of best practice gap change **d) Contribution of efficiency change**

Bad output orientation



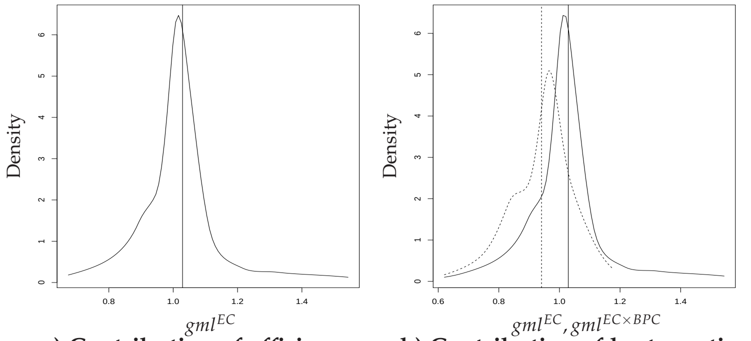
e) Contribution of best practice gap change **f) Contribution of efficiency change**

gml^{BPC} ——— $gml^{BPC \times EC}$ - - - - - -

Notes: All figures contain densities estimated using kernel density estimation for the different components of the biipartite decomposition in expression (11), considering the good and bad output orientation. The vertical lines in each plot represents the average for each component of the decomposition. Densities were estimated using a Gaussian kernel and the Sheather and Jones (1991) plug-in bandwidth.

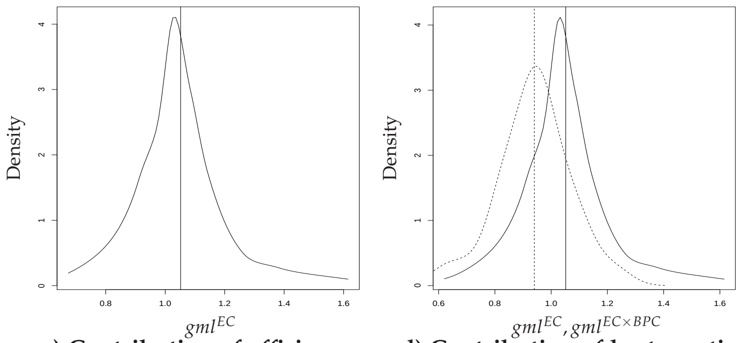
Figure 3: Kernel density plots of the bipartite decomposition of educational improvement, good and bad output orientation (local bandwidth)

Good and bad output orientation



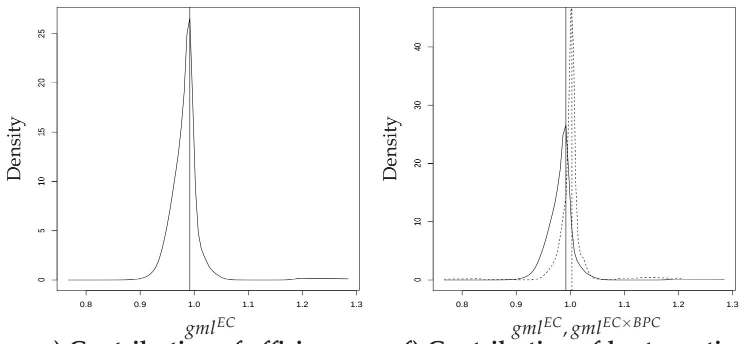
a) Contribution of efficiency change **b) Contribution of best practice gap change**

Good output orientation



c) Contribution of efficiency change **d) Contribution of best practice gap change**

Bad output orientation



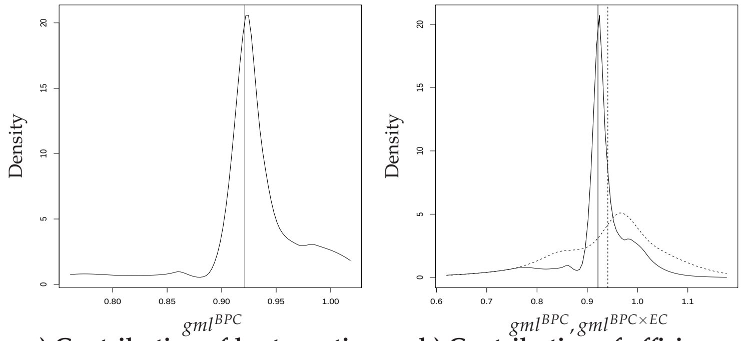
e) Contribution of efficiency change **f) Contribution of best practice gap change**

gml^{EC} ——— $gml^{EC\times BPC}$ _ _ _ _ _ _

Notes: All figures contain densities estimated using kernel density estimation for the different components of the biipartite decomposition in expression (11), considering the good and bad output orientation. The vertical lines in each plot represents the average for each component of the decomposition. Densities were estimated using local likelihood methods (Loader, 1996), and a Gaussian kernel was chosen.

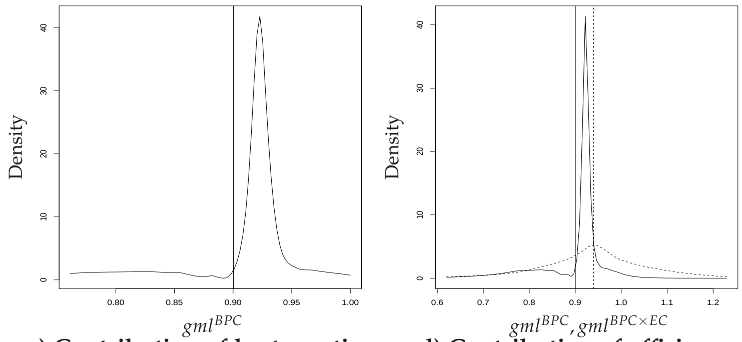
Figure 4: Kernel density plots of the bipartite decomposition of educational improvement, good and bad output orientation (local bandwidth)

Good and bad output orientation



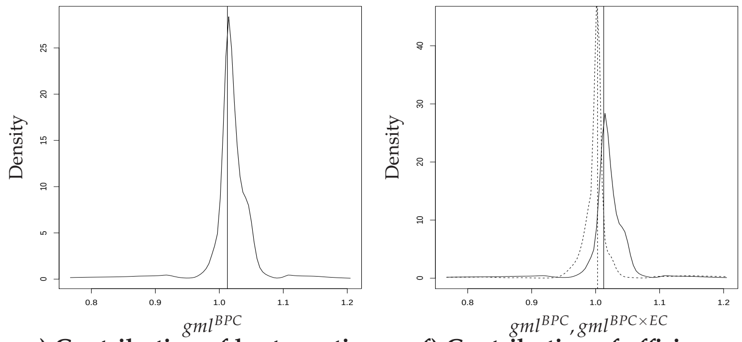
a) Contribution of best practice gap change **b) Contribution of efficiency change**

Good output orientation



c) Contribution of best practice gap change **d) Contribution of efficiency change**

Bad output orientation



e) Contribution of best practice gap change **f) Contribution of efficiency change**

gml^{BPC} ———— $gml^{BPC \times EC}$ - - - - - -

Notes: All figures contain densities estimated using kernel density estimation for the different components of the biipartite decomposition in expression (11), considering the good and bad output orientation. The vertical lines in each plot represents the average for each component of the decomposition. Densities were estimated using local likelihood methods (Loader, 1996), and a Gaussian kernel was chosen.