J Prod Anal (2017) 47:83–101
DOI 10.1007/s11123-017-0491-9

# An international comparison of educational systems: a temporal analysis in presence of bad outputs

Víctor Giménez[1] · Claudio Thieme[2] · Diego Prior[1] · Emili Tortosa-Ausina[3]

Published online: 25 January 2017
© Springer Science+Business Media New York 2017

**Abstract** This study uses the global non-radial Malmquist index to measure performance change in the educational systems of 29 countries/economies participating in PISA 2003 and 2012 for students at age 15 in the disciplines of mathematics and reading. This methodology is particularly appropriate both for its desirable properties as well as its suitability for the educational context. Results indicate a positive evolution in educational systems' performance during this period. This improvement is mainly due a positive efficiency change, which offsets the negative technological change observed. Nevertheless, a deeper scrutiny at the country level shows that results varied remarkably among them.

**Keywords** Education · Efficiency · Global non-radial Malmquist index · PISA

**JEL Classifications** C61 · H52 · I21

✉ Emili Tortosa-Ausina
tortosa@uji.es

[1] Universitat Autònoma de Barcelona, Bellaterra (Barcelona) 08193, Spain

[2] Universidad Diego Portales, Avenida Santa Clara 797, Huechuraba, Santiago, Chile

[3] Universitat Jaume I, Avenida de Vicent Sos Baynat, s/n, Castellón 12071 Castelló, Spain

## 1 Introduction

In a world characterized by rapid technological change and the importance of innovation processes, the level of academic attainment that students can achieve is essential to improving the levels of wealth and welfare of the citizens in their countries. In the field of public policy in education it is therefore unsurprising to see a growing concern about the assessment of student learning (Denvir and Brown 1986; Ercikan 2006). Understanding educational outcomes is critical to effective planning of educational policies, and the assessment of educational reforms.

In this vein, the OECD Programme for International Student Assessment (PISA) recently published the results from its 2012 assessment.[1] The PISA international survey is carried out every three years and evaluates education systems across the world with tests of 15-year-old students' skills and knowledge; it also provides vital information on other relevant factors (related to students background, school system and the learning environment) that can affect the learning process.

While the results obtained by a given country in a standardized test (such as PISA, or the Trends in International Mathematics and Science Study, TIMSS) are a good reflection of students' academic levels, by themselves they cannot be regarded as a performance indicator for their educational systems and, therefore, their school authorities. The main limitations of these standardized international tests are as follows: (i) the assessment of an organization performance (in this particular case, a country) does not

[1] Around 510,000 students in 65 economies took part in Pisa 2012 assessment of reading, mathematics and science representing about 28 million 15-year-olds globally. Complete information about PISA and databases can be found at https://www.oecd.org/pisa/.

depend exclusively on outcome variables; instead, we can consider efficiency indicators that measure different aspects of the educational process; the results achieved (output) during this process are a consequence of the resources used, the process itself, and 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 students acquire at school, but should also include other outcomes such as the standard deviation of test scores (an *undesirable* outcome of the educational process, in terms of educational inequality); and (iii) when measuring students educational achievements at a given point in time, it is difficult to disentangle how much achievement is attributable to the student herself, to her family, or to the strategies applied by previous educational authorities.

Consistent with this, over the last few years there has been a growing interest in assessing and comparing the performance of educational systems in different countries and regions using cross-sectional data.[2] A first approach to this issue considered aggregate data for different samples of countries participating in international tests. These include studies by Afonso and St. Aubyn (2006), Giambona et al. (2011), Thieme et al. (2012), Aristovnik and Obadić (2014) or Giménez et al. (2007), among others. For instance, Giménez et al. (2007) performed 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. (2012) carried out a similar comparison for the 54 countries participating in PISA 2006, addressing the first two limitations stated in the preceding paragraph; specifically, these authors apply directional distance functions (DDF) to evaluate efficiency indicators that relate outcome variables to resource variables used in the educational process.[3] The authors jointly evaluate good (or desirable) outputs of academic achievement and bad (or undesirable) outputs arising from educational inequality. Their results show that it is feasible for a higher education system to combine high levels of student learning and, simultaneously, obtain low inequality levels; however, they found that in most instances both dimensions required significant improvements.

A second approach includes studies that compare the performance of educational systems in different countries using either school- or student-level data. The study by Cordero et al. (2017), based on school-level data, evaluates school performance using the metafrontier framework to compare and decompose the technical efficiency of primary schools of 16 European countries participating in PIRLS 2011. They also consider an extension of the conditional nonparametric robust approach to test for the potential influence of environmental school factors and cultural values of each country. Some of the most prominent findings are that rankings of countries based on academic results are only modified to a certain extent when controlling for data on school inputs involved in the educational process, and that heterogeneity across countries is more relevant than among schools. In addition to this, the study by Deutsch et al. (2013) uses student-level data to measure efficiency from results of the 2006 PISA survey for five Latin American countries by means of corrected ordinary least squares, whereas De Jorge and Santín (2010) apply Data Envelopment Analysis to estimate efficiency for 18 European countries participating in PISA 2003.

However, to obtain a fuller evaluation of educational systems' performance it would be desirable to evaluate the *change* in performance over time—which, as suggested above, could constitute a third limitation of previous research initiatives. Measuring this change is critical, since there is a general consensus that not only students' achievement needs to be measured, but also their progress, and how much of this progress is attributable to the educational system itself or to external factors. This particular research area in education economics refers to these measures as *growth* studies, which require at least two evaluations at different points in time.

To our knowledge, only two studies have analyzed how performance has changed over time, as well as the components of performance (Agasisti 2014; Aparicio et al. 2016a). Agasisti (2014) uses data from PISA to compare spending efficiency on education in 20 European countries during the 2006–2009 period. The average mathematics test score is considered as the output of the education processes and the efficiency scores are calculated using a bootstrapped Data Envelopment Analysis (DEA) approach. In a second stage, the efficiency scores are regressed against contextual variables. Finally, Malmquist indexes are calculated to measure the change in efficiency in the analyzed period. Results show that the average efficiency remained fairly stable because of the action of two contrary forces: a slight improvement in terms of (average) pure efficiency, with a simultaneous deterioration of the efficiency frontier.

[2] Recent literature reviews on efficiency in education include De Witte and López-Torres (2017), Johnes (2015), Grosskopf et al. (2014), Emrouznejad et al. (2010) and, a bit more distant in time, Johnes (2004) and Worthington (2001). In several of these studies, among other issues, the authors review thoroughly the studies that have dealt with the issue of efficiency in education, listing the inputs, outputs and environmental/contextual variables, considering the different levels of analysis (university, school/high school, district/county/city, or country), as well as the different methodological approaches. In addition, some authors (De Witte and López-Torres 2017) have an explicit attempt to link the standard economics of education literature and the (nonparametric) efficiency literature.

[3] See also the recent contribution by Aparicio et al. (2016a), in which the Malmquist index is applied to different samples of PISA data (2006, 2009 and 2012).

Aparicio et al. (2016a) also uses data from PISA 2006, 2009 and 2012 to compare the performance of public and private government-dependent secondary schools in the Basque Country (an Autonomous Region of Spain). These authors propose a new pseudo-panel Malmquist index to analyze evolution the schools' evolution during the period. Their results suggest that performance was persistently and significantly lower for public schools than private government-dependent schools. See Table 1 for a comparative literature review on international performance of education systems.

Therefore, in accordance with the rationale presented above, desirable properties of a good education system would include not only its ability to obtain high average academic achievement among its students, but also to ensure that all its students make progress. To achieve this, strategies must be developed that enable relatively disadvantaged students to also make progress and reach basic standards. Hence, an educational system that evolves satisfactorily will be one that improves the *average* student academic achievement while simultaneously minimizing the standard deviation of test scores (the educational inequality). Similarly, changes in the endowment of resources used by the system will indicate whether the changes in the level of educational achievement (either positive or negative) are due to *technical change*, which might be attributable to an improvement in the educational resources available, or to enhanced *efficiency* when utilizing these resources.

To explore these issues, researchers have proposed a variety of measures to evaluate performance change over time (either due to efficiency change or technical change). Most of these studies follow Färe et al. (1994b), 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), or Oh (2010), among others. To measure changes in performance (to achieve educational objectives), in this study we model both good and bad outputs, using the global non-radial Malmquist index (hereafter *GNRMI*), similar to the proposed by Zhang and Choi (2013). This index is not only appropriate for its highly desirable properties; it also suits our context because it incorporates bad outputs which, ideally, educational systems should minimize while simultaneously maximizing the good/desirable outputs.

The global non-radial Malmquist index is used to measure performance change in the educational systems of 29 countries (21 OECD countries and 8 OECD partner countries) participating in PISA 2003 and 2012 for students at age 15 in the disciplines of mathematics and reading. The results can be interpreted globally or by evaluating the decomposition of the global non-radial Malmquist index into its two components—best practice gap change (*BPC*) and efficiency change (*EC*). On average, results indicate a positive evolution in educational performance between 2003 and 2012, mainly driven by a positive efficiency change, which offsets the negative technological change observed. Nevertheless, results also varied remarkably among countries.

The paper is organized as follows. After this introduction, Section 2 describes the methodological aspects of the global non-radial Malmquist index 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 principal conclusions.

## 2 Methodology

### 2.1 Modeling the educational performance over time

Studies analyzing the evolution of efficiency over time often apply the Malmquist index (Caves et al. 1982). This index is used to explain the change in 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 it to the case of directional distance functions (DDF). These have been widely used in studies measuring efficiency that incorporate the environmental impact of the units analyzed by considering the bad or undesirable outputs of the production process (Sueyoshi and Goto 2011; Watanabe and Tanaka 2007; Färe et al. 2005). The new index was named the Malmquist-Luenberger Index (ML). The application of the ML has often been related to radial expansions of good and bad outputs (Weber and Domazlicky 2001; Yörük and Zaim 2005; Kumar 2006; Nakano and Managi 2008), probably to avoid the problems of translation invariance recently highlighted by Aparicio et al. (2016b). However, an important contribution by Zhou et al. (2012) introduced the concept of non-radial directional distance function (NDDF) where potential improvements are determined individually for each good and bad output as well as for each input. Dynamic analysis has also been applied to NDDF in Zhang et al. (2013) a metafrontier non-radial Malmquist index.

Most of the former temporal 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 by the product of changes in productivity in the different sub-periods within it. Secondly, there is a possibility of infeasibilities in the calculation of the cross-distance functions necessary to calculate these indices. Although it is a necessary and sufficient condition that technical change be Hicks-neutral to ensure circularity (Balk 2001) and a particular data structure must ensure the absence of feasibility problems (Xue and Harker 2002), it is often difficult to comply with these conditions in empirical applications. To remedy these two deficiencies, Pastor and Lovell (2005) proposed a modification of the Malmquist index known as the global Malmquist index.

Table 1 Studies comparing the performance of educational systems in different countries

| Authors | Database | Sample | Methodology | Output(s) | Input(s) | Other variables |
|---|---|---|---|---|---|---|
| Aparicio et al. (2016a) | PISA 2006, 2009, 2012 | Public and private government-dependent secondary schools in the Basque Country (Spain) | DEA and pseudo-panel Malmquist index | (i) Mean values for the student results in mathematics, reading and science | (i) Index of the highest level of parental education; (ii) index of the highest parental occupational status; (iii) index of the quality of the school resources; (iv) ratio between the total number of teachers (weighted by their commitment) and the total number of students | |
| Cordero et al. (2017) | PIRLS 2011, World Bank | 16 European countries (2398 schools) | Metafrontier, robuts and conditional FDH. Output orientation | (i) Plausible values in reading, average (school level) | (i) Number of teachers per 100 students (school level); (ii) instruction hours per week (school level); (iii) socioeconomic status of student index (school level) | (i) Different environmental variables (school and country levels) |
| Agasisti and Zoido (2015) | PISA 2012 | 30 countries (8600 schools) | DEA-bootstrap, bias-corrected, Tobit regression in the second stage | (i) Average score in mathematics (school level); (ii) average score in reading (school level) | (i) Inverse of the student/teacher ratio (school level); (ii) number of computers per student at school (school level); (iii) ESCS, the average of socioeconomic status of students in the school (school level) | (i) Different variables in the second stage |
| Aristovnik and Obadić (2014) | PISA 2006, Eurostat, UNESCO, World Bank | 31 EU and OECD countries | DEA-VRS, output orientation | (i) School enrolment, secondary education (% gross, country level); (ii) PISA average (country level); (iii) teacher/pupil ratio, secondary education (country level) | (i) Expenditure per student, secondary education (% of GDP pc, country level); (ii) teacher/pupil ratio, secondary education (country level); (iii) school enrolment, secondary education (% gross, country level) | – |
| Agasisti (2014) | PISA 2009 and PISA 2006 | 20 European countries | DEA-bootstrap, regression in the second stage, Malmquits index | (i) Average score in mathematics (country level); (ii) average score in reading (country level) | (i) Expenditure per student (country level; (ii) studens/teacher ratio (country level) | Different contextual and structural variables (country level) |
| Deutsch et al. (2013) | PISA 2006 | 5 Latin-American countries (7138 students) | Corrected ordinary least squares (COLS); regression in the second stage; Shapley decomposition to evaluate the relative importance of the determinants of efficiency | Cognitive ability of the student (latent variable considering student score in reading, science and mathematics test, student level) | (i) Educational means at home index (student level); (ii) pedagogical characteristics of the school index (student level); (iii) physical and human capital at school index (student level); (iv) time devoted to informal | – |

**Table 1** continued

| Authors | Database | Sample | Methodology | Output(s) | Input(s) | Other variables |
|---|---|---|---|---|---|---|
| | | | | | learning index (student level); (v) time devoted to formal learning index (student level) | |
| Thieme et al. (2012) | PISA 2006 | 54 countries | Directional Distance Function, output orientation | (i) Educational achievement in reading (country level); (ii) average educational achievement in science and maths (country level); (iii) average inequality in science, reading and mathematics (bad output, country level) | (i) Index of availability and quality of educational resources (country level); (ii) index of availability and quality of human resources (country level) | Environmental variables (socioeconomic and cultural index of PISA, country level) |
| Giambona et al. (2011) | PISA 2006 | 24 European countries | DEA-boostrap, output orientation | (i) Mathematics score (country level); (ii) reading score (country level); (iii) sciences score (country level) | (i) Educational resources available for the student at home index (country level); (ii) family background index (country level) | – |
| De Jorge and Santín (2010) | PISA 2003 | 18 European countries | DEA-VRS and DEA-CRS, truncated regression in second stage, analysis of variance among and within schools | (i) Plausible scores in mathematics (student level); (ii) plausible scores in reading (student level); (iii) plausible scores in science (student level) | (i) Peer effect; (ii) quality of educational resources; (iii) quality of school physical infrastructure | Different contextual variables. |
| Sutherland et al. (2009) | PISA 2023 | 30 OECD countries | Stochastic Frontier Analysis (SFA) | Average school score in mathematics, sciences and reading (school level) | (i) Ratio of teaching staff per 100 students (school level); (ii) school-average of socioeconomic status (school level) | – |
| Giménez et al. (2007) | TIMSS 1999 | 31 countries | DEA and DEA with contextual variables (one stage) | (i) Academic performance in mathematics (country level); (ii) academic performance in science (country level) | (i) Intensity of teaching resources; (ii) index of facility availability; (iii) index of material consumptions; (iv) quality of teaching staff | Four factors identified from contextual variables (contextual variables) |
| Afonso and St. Aubyn (2006) | PISA 2003 | 25 countries, mostly OECD | DEA-VRS, output oriented, Tobit regression in the second stage | Average educational achievement in science, reading and mathematics (country level) | (i) Instructional hours per year in school (country level); (ii) teachers per 100 students (country level) | (i) GDP per capita (environmental variable, country level); (ii) parental educational attainment (environmental variable, country level) |
| Afonso and St. Aubyn (2005) | PISA 2000, OECD 2002 | 17 OECD countries | FDH and DEA, input and output orientation | Average educational achievement in science, | (i) Instructional hours per year in school (country level); (ii) | – |

**Table 1** continued

| Authors | Database | Sample | Methodology | Output(s) | Input(s) | Other variables |
|---|---|---|---|---|---|---|
| | | | | reading and mathematics (country level) | teachers per 100 students (country level) | |
| Clements (2002) | TIMSS | EU countries | FDH | Attainment TIMSS scores (country level) | (i) Expenditure per student and teachers (country level); (ii) students' ratio (country level) | – |
| Gupta and Verhoeven (2001) | Development Assistance Comitte (DAC) of the OECD, World Bank, United Nations | 85 countries (Africa, Asia and Western Hemisphere) | FDH, input orientation | (i) Educational attainment by primary school enrolment (country level); (ii) secondary school enrollment (country level); (iii) adult literacy (country level) | Per capita education (PPP) (country level) | – |

Similarly, Oh (2010) adapted the Malmquist-Luenberger index to achieve the same properties, leading to the global Malmquist-Luenberger index. This paper proposes a global non-radial Malmquist index similar to that proposed by Zhang and Choi (2013) and Zhang et al. (2013) for the temporal analysis of education system performance. The reason for the choice of this index is that, apart from its desirable properties, it incorporates bad outputs, which educational systems should minimize while maximizing the outputs (good outputs). This index is therefore particularly appropriate in the specific context of education.

Let $k$ be the countries with available information on their educational systems for $t$ years, where $m$ good outputs were produced, and $h$ bad outputs generated from the consumption of $n$ inputs. These are denoted by $(X_j, Y_j, B_j)$, $j = 1, \ldots, k$. It is assumed that $X_j = (x_{1j}, \ldots, x_{nj}) \geq 0$, $X_j \neq 0$, $Y_j = (y_{1j}, \ldots, y_{mj}) \geq 0$, $Y_j \neq 0$, and $B_j = (b_{1j}, \ldots, b_{hj}) \geq 0$, $B_j \neq 0$.

The production technology is defined by:

$$T = \left\{ (X, Y, B) : \sum_{j=1}^{k} \lambda_j Y_j \geq Y, \sum_{j=1}^{k} \lambda_j X_j \leq X, \sum_{j=1}^{k} \lambda_j B_j \leq B, \lambda \geq 0 \right\} \quad (1)$$

Various approaches to integrate the undesirable outputs in the efficiency estimations can be found in the literature. The most popular approach is probably 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 on the problems and the solutions of this option is far from over; see, for instance, Kuosmanen (2005), Kuosmanen and Podinovski (2009), Färe and Grosskopf (2009), 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 suggested by Golany and Roll (1989) and Seiford and Zhu (2002), but this conversion may bring about significant changes in the level of efficiency found. Finally, Reinhard et al. (2002) and Hailu and Veeman (2001), argue that perhaps the most intuitive option is to consider the bad outputs as strongly disposable inputs. Because of its simplicity, this option was selected in our proposal.

The efficiency for a given unit belonging to $T$ can be measured by the following non-radial directional distance function based on Zhou et al. (2012):

$$\vec{D}(X, Y, B : g) = sup\left\{ w^T \beta : ((X, Y, B) + g \cdot diag(\beta)) \in T \right\} \quad (2)$$

The *NDDF* in Eq. (2) above determines the maximum attainable increases in the good outputs as well as the maximum decreases for both bad outputs and inputs over the vector $g = (-g_{x1}, \ldots, -g_{xn}, g_{y1}, \ldots, g_{ym}, -g_{b1}, \ldots, -g_{bh})$. $\beta$ denotes a vector in $\mathbb{R}^{m+n+h}$ of the scaling factors representing inefficiency measures for inputs and outputs. This approach will lead to an evaluation where each educational system will be assessed in the direction that is more favorable to it, without assuming *ex-ante* any desirable approaching direction towards the frontier. Consequently, the model will be able to adapt to any specific educational policy or strategy. In the case of analyzing the performance of educational systems, an output-oriented approach seems to be justified, given that any country should be interested in maximizing its educational level with the available resources. Our main focus is then on the outputs side. For this reason the vector $g = (0, g_y, -g_b)$ is selected, which defines the desirable directions for improvement for both types of outputs. Finally, $w^T$ denotes a normalized weights vector in order to allow the introduction of some value judgments on the importance of the outputs. The weights are usually assumed to be equal for each input and/or output as we assume in our case. Nevertheless, there are other alternatives, such as the one proposed by Zhang and Choi (2013) who suggest assigning the same weight to each category of inputs and outputs and then distributing them equally among the number of variables included in each category. For this reason, the maximum degree of generality for the formulation of weights has been adopted. If $\vec{D}(X, Y, B : g) = 0$ there is no margin for improving either good or bad outputs in the $g$ direction and consequently the educational system is located on the frontier.

Since $w^T\beta$ is not lying between zero and unity, and based on Zhang et al. (2013) and Zhang and Choi (2013), and in order to facilitate comparisons with a conventional distance function, we define a factor performance index (FPI) for each country as follows:

$$FPI = \frac{1 - \sum_{v=1}^{h} \beta_v^{b^*} w_v^b}{1 + \sum_{r=1}^{m} \beta_r^{y^*} w_r^y} \tag{3}$$

where $\beta_r^{y^*}$ and $\beta_v^{b^*}$ are the optimal potential improvements for the good and bad outputs, respectively. Clearly, $FPI \in (0, 1]$. If a country is efficient, then the FPI will be equal to one, while the smaller the values, the greater the distance to the frontier. This index will also allow us to propose a global non-radial Malmquist index based on the one proposed by Zhang and Choi (2013) for metafrontier analysis.

In our empirical application the inputs and outputs, described in the following section, are fixed ratios. A ratio is fixed when it can be assumed that it remains constant to scale their underlying volume variables by a non-negative constant (Olesen et al. 2015). The use of ratio variables in DEA has important implications, especially in modeling the returns to scale exhibited by the technology (Golany and Thore 1997; Dyson et al. 2001; Hollingsworth and Smith 2003; Cooper et al. 2007). For example, Hollingsworth and Smith (2003) recommend a technology that assumes variable returns to scale (VRS) in the cases where ratio variables are present. The rationale is that assuming constant returns to scale (CRS) proportionality in the variation of inputs and outputs when increasing or decreasing, the size of a decision-making unit is also assumed, something that does not occur when a ratio is scaled by a constant.

Instead, by assuming VRS this problem is mitigated since the need for scaling is lower. However, only rencently have Olesen et al. (2015) explored the implications of using ratio variables in DEA. They proposed several solutions to properly model CRS and VRS when ratio variables are present. In the latter case (VRS), the proposal converges with the FDH technology when all the model variables are ratios, as in our case. For this reason, we have considered this technology by defining the $\lambda$ variables as binary (Deprins et al. 1984). FDH models are especially appropriate when the convexity assumption may be difficult to justify (Grifell-Tatjé and Kerstens 2008). In fact, an important argument against convexity of production correspondences in economic theory is related to indivisibilities. (Scarf 1994) highlighted the importance of indivisibilities in selecting the technology. This argument has often been applied against using convex technologies (Tone and Sahoo 2003).

Apart from the technical reasons arising from the presence of ratio variables, when comparing countries the convexity assumption is probably more debatable from a conceptual point of view. Although it is well known that the discriminant capacity of FDH models is generally reduced for small samples, we consider it to be the most appropriate methodological alternative given the variables used and the nature of the units analyzed. However, in order to check the robustness of the results, we also made the calculations under the DEA-VRS technology. A similar approach has also been adopted by other studies on education efficiency for OECD countries (Afonso and St. Aubyn 2005).

Let $T^u$ be the technology production for year $u$. Then $FPI^u(p)$ denotes the factor performance index for year $p$ with respect to $T^u$. $FPI^u(p)$ is calculated from the optimal values for $\beta$ obtained by solving the following FDH-type model:

$$\begin{aligned}
&\vec{D}^u(X^p, Y^p, B^p : g) \\
&\quad = \max \sum_{r=1}^{m} w_r^y \beta_r^y + \sum_{v=1}^{h} w_v^b \beta_v^b + \sum_{s=1}^{n} w_s^x \beta_s^x \\
&\quad \text{s.t.} \\
&\sum_{j=1}^{k} \lambda_j y_{rj}^u \geq y_{ro}^p + \beta_r^y g_{yr},\ r = 1, \ldots, m \\
&\sum_{j=1}^{k} \lambda_j b_{vj}^u \leq b_{vo}^p - \beta_v^b g_{bv},\ v = 1, \ldots, h \\
&\sum_{j=1}^{k} \lambda_j x_{sj}^u \leq x_{so}^p - \beta_s^x g_{xs},\ s = 1, \ldots, n \\
&\sum_{j=1}^{k} \lambda_j = 1 \\
&\qquad \beta_r^y, \beta_v^b, \beta_s^x \geq 0\ \lambda_j \in \{0, 1\}
\end{aligned} \tag{4}$$

where $\lambda_j$ is the intensity vector and $y^u_{rj}$, $b^u_{vj}$ and $x^u_{sj}$ the output $r$, bad output $v$ and input $s$, respectively, for unit $j$ in year $u$.

We also define the global $FPI$ for the year $p$ with respect to the global technology $T^g = T^1 \cup T^2 \cup \ldots T^t$ as $FPI^g(p)$, which can be obtained from the $\beta$ optimal values yielded by the following linear programming problem:

$$\vec{D}^g(X^p, Y^p, B^p : g) = \max \sum_{r=1}^{m} w^y_r \beta^y_r + \sum_{v=1}^{h} w^b_v \beta^b_v + \sum_{s=1}^{n} w^x_s \beta^x_s$$

s.t.

$$\begin{aligned}
&\sum_{j=1}^{k}\sum_{u=1}^{t} \lambda_{uj} y^u_{rj} \geq y^p_{ro} + \beta^y_r g_{yr},\ r = 1, \ldots, m \\
&\sum_{j=1}^{k}\sum_{u=1}^{t} \lambda_{uj} b^u_{vj} \leq b^p_{vo} - \beta^b_v g_{bv},\ v = 1, \ldots, h \\
&\sum_{j=1}^{k}\sum_{u=1}^{t} \lambda_{uj} x^u_{sj} \leq x^p_{so} - \beta^x_s g_{xs},\ s = 1, \ldots, n \\
&\sum_{j=1}^{k}\sum_{u=1}^{t} \lambda_{uj} = 1 \\
&\beta^y_r, \beta^b_v, \beta^x_s \geq 0\ \lambda_{uj} \in \{0, 1\}
\end{aligned} \tag{5}$$

For the DEA-type formulation, the constraint type $\lambda \in \{0, 1\}$ should be removed from linear programs (4) and (5).

Similarly to Zhang et al. (2013) and Zhang and Choi (2013), we propose the global non-radial Malmquist index ($GNRMI$) as:

$$\begin{aligned}
GNRMI &= \frac{FPI^g(t+1)}{FPI^g(t)} = \left[\frac{FPI^{t+1}(t+1)}{FPI^t(t)}\right] \times \left[\frac{\frac{FPI^g(t+1)}{FPI^{t+1}(t+1)}}{\frac{FPI^g(t)}{FPI^{t+1}(t)}}\right] \\
&= EC \times BPC
\end{aligned} \tag{6}$$

If $GNRMI = 1$, then there have been no changes changes in productivity during the period $t$ and $t+1$. A value greater than one means an increase in productivity, while a value less than unity shows a decline in productivity during the period. $EC$ (efficiency change) reflects the change in technical efficiency or catching-up between year $t$ and year $t+1$. If $EC > 1$, technical efficiency improved in the period. In other words, the unit is closer to its contemporary frontier in year $t+1$ than in $t$. A value lower than unity is interpreted inversely. The term $BPC$ (best-practice gap change) is a measure of technological change in the period, that is, of how contemporary frontiers have shifted in the period with respect the global frontier.

## 2.2 Bipartite decomposition of the relative contributions to educational performance

In accordance with the expressions detailed in the previous section, the global non-radial Malmquist index ($GNRMI$) 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 *GNRMI on average*, we can also consider a distribution dynamics approach to analyze what the largest contributors to the variation in performance are, as measured by $GNRMI$ between periods $t$ (2003) and $t+1$ (2012). To this end, we use nonparametric density estimation, based on kernel smoothing.

We rewrite expression (6) above as follows:

$$gnrmi^{EC \times BPC} = EC^{t,t+1} \times BPC^{t,t+1} \tag{7}$$

according to which we use expression $gnmri^{EC \times BPC}$ to indicate that the change in educational achievement is obtained by successively multiplying its three components. This in turn, allows us to construct counterfactual distributions by sequentially introducing each of the factors. Specifically, the counterfactual educational achievement change attributable to changes in efficiency would be:

$$gnmri^{EC} = EC^{t,t+1} \tag{8}$$

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

Analogously, for extending this sequential decomposition, we would proceed as follows:

$$\begin{aligned}
gnmri^{EC \times BPC} &= EC^{t,t+1} \times BPC^{t,t+1} \\
&= gnmri^{EC} \times BPC^{t,t+1}
\end{aligned} \tag{9}$$

We can consider this sequential decomposition in a different order. In such a case, the counterfactual educational achievement change attributable to best practice gap change would be:

$$gnmri^{BPC} = BPC^{t,t+1} \tag{10}$$

which, in this case, isolates the effect on the distribution of best practice gap changes only, assuming $EC$ does not contribute to the change in educational achievement ($gnmri$). Then expression (9) would become:

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

We refer to the decomposition in both expressions (9) and (11) as the bipartite decomposition of the relative contributions to the changes in the distribution of educational performance.

Although the use of these counterfactual distributions is not very popular in the field of economics of education, or education in general, their use is more frequent in other contexts such as impact evaluation. In our case, we have followed the proposals made by Kumar and Russell (2002) in the field of macroeconomic convergence, whose proposal was based on combining the distribution dynamics approach to convergence analysis (Quah 1993a, b) and the (deterministic) frontier production function literature (Färe et al. 1994a). This pioneering contribution by Kumar and Russell (2002) was soon followed by some extensions of the model, in order to account for relevant issues in economics such as the contributions of human capital (Henderson and Russell 2005) or financial development (Badunenko and Romero-Ávila 2013) to productivity growth and convergence.

As indicated in the first paragraph of this subsection, the densities can be estimated via kernel smoothing, which entails two unequally important decisions, the choice of kernel and the choice of bandwidth ($h$), which tunes the amount of *bumps* under each curve—higher values of $h$ tend to smooth more, revealing fewer data particularities, low values of $h$ tend to smooth less, providing more detail but generating (in some cases) fuzzy graphics.[4] Regarding the kernel, we chose a popular alternative, the Gaussian kernel. Although other choices are also possible (e.g., Epanechnikov, triangular, etc.), its impact on the final outcome is much lower than that of the bandwidth. In this case, the available literature is lengthy and, and we have attempted to reflect this relatively larger literature, considering both a *global* bandwidth (the amount of smoothing is the same at all data points) and a *local* bandwidth (the amount of smoothing varies *locally*, depending on the structure of the data at a given point). For the former, we followed the proposals by Sheather and Jones (1991), whereas for the local bandwidth estimator we followed Loader (1996).

## 3 Data, inputs and outputs

This study considers information from the educational systems of the 29 countries (21 OECD member countries and 8 OECD partner countries) participating in the OECD's Programme for International Student Assessment (PISA) for years 2003 and 2012. PISA has been operating since 2000 and assesses average results for 15-year-old students between 7th and 12th grade. PISA seeks to determine the extent to which students have acquired the competencies that will enable them to face the challenges of today's knowledge society. To do so, every three years, the PISA survey tests general reading, mathematical and scientific literacy in terms of general competencies, that is, how well students can apply the knowledge and skills they have learned at school to real-life challenges. 2012 was the fifth edition version of this study. Around 510,000 students participated and 65 countries/economies took part. In PISA 2003 (second version) 41 countries participated.[5]

As noted above, the methodology described in the preceding section is used to evaluate the change in the performance of educational systems in achieving educational goals. To do this we consider, in line with the previous literature (Carlson 2001), that a good educational system is not only one that obtains high results (on average) in terms of students' academic achievement, but also one that can ensure all its students make progress. To do this, it must also develop strategies that enable its most disadvantaged students to advance and reach standards. Therefore, an educational system that evolves satisfactorily will be one that can improve the average academic achievement of its students while at the same time minimizing the differences among them. Similarly, the change in the allocation of resources used will reveal whether if these changes in achieving educational objectives (either positive or negative) are due to a technical change (due to an improvement in the provision of the resources allocated for educational purposes), or by improving how efficiently they are used.

In this line, therefore, our selection of variables considers as good outputs the average academic achievement of students in each country in the mathematics test and in the reading test. Including these two subject areas eliminates any potential specialization bias by the participating countries in either of these subjects, and follows the logic of previous work in the area (see Table 1). A relevant question is the comparability of the data used, since it comes from different years (Brown et al. 2007). The OECD (2014, p. 159) states that *for PISA 2012 the decision was made to report the reading, mathematics and science scores on these previously developed scales. That is the reading*

[4] Some excellent monographs on this issue are those by Silverman (1986), Scott (1992), Li and Racine (2007) and, more recently, Henderson and Parmeter (2015).

[5] The international contractor in each country randomly selects schools for participation in PISA. At these schools, the test is given to students between the ages of 15 years 3 months and 16 years 2 months at the time of the test, rather than to students in a specific year of school; this age represents the end of compulsory education in most participating countries. In general, each version of PISA considers a minimum of 150 schools per participant country/economy (or all the schools if there are fewer than 150 schools in that country/economy). Within each participating school, a sample of students, usually numbering 35, is selected with equal probability (all students take that test if there are fewer than 35 in the school and with a minimum of 20 students so as to guarantee the validity of the test within and among schools). In total, in each country a minimum size of 4500 students are tested.

*scales used for PISA 2000, PISA 2003. PISA 2006, PISA 2009 and PISA 2012 are directly comparable. PISA 2012 mathematics reporting scale is directly comparable to PISA 2003, PISA 2006 and PISA 2009 and the science reporting scale is directly comparable to PISA 2006 and PISA 2009 scale*. Therefore, the scores from the mathematics and reading scales used in this study are directly comparable for years 2003 and 2012. The same argument is also used in Aparicio et al. (2016a) to justify the comparability of the PISA data they use.

As educational inequality (bad outputs) variables, the study considers the average standard deviation of the results the students in each country obtain in these two disciplines, and also considers them separately. The concept of educational inequality has been widely discussed in the field of education (Jacob and Holsinger 2008; Wenglinsky 1998). However, despite public policies that prioritize the importance of quality and equity in the provision of education, only Thieme et al. (2012) have previously considered this question in a study of international comparisons of efficiency in education systems.

An analysis of the data shows that for the sample countries both disciplines evolved positively, on average, between 2003 and 2012. This improvement is greater in reading than in mathematics (7-point improvement vs. 3 points, respectively), and occurs not only in terms of students' academic achievement but also in terms of inequality, with a 3-point standard deviation decrease in reading and, and a 2-point standard deviation decrease in mathematics.

At the country level, the data show a high correlation in the results for academic achievement between the two disciplines for each of the two assessments (2003 and 2012), with a higher ratio in 2012 ($R^2 = 0.935$) than in 2003 ($R^2 = 0.894$). In mathematics, the countries with the highest improvements in their students' results were Brazil (35 points), Tunisia (29 points) and Mexico (28 points). By contrast, the countries where students' academic achievement decreased were Sweden (−31 points), Finland (−26 points) and New Zealand (−24 points). As for reading, the countries with the greatest improvements were Japan (40 points), Hong Kong-China (35 points) and the Russian Federation (33 points), whereas those with the sharpest declines were Sweden (31 points), Uruguay (23 points) and Finland (19 points).

Despite the high correlation in the results for both disciplines, each country's performance in the two disciplines is not necessarily coincidental. For example, in 2012 there is a significantly better result in mathematics than in reading for Macao-China, Slovak-Republic, Austria, South Korea, Hong Kong, the Russian Federation and Czech Republic. In contrast, some countries (Ireland, Greece, USA, New Zealand and Hungary) show a specialization in reading in 2012.

Similarly, it appears that high levels of academic achievement can be obtained with low inequality levels. For example, in the case of mathematics in 2012, countries where this occurred include Finland, Canada, Ireland, Latvia and Spain, whereas in the case of reading, the countries were South Korea, Hong Kong, Ireland, Poland, Macao-China, USA, and the Czech Republic.

For the selection of input variables, we mainly chose those which have been most frequently used in empirical assessments of efficiency in education, especially in international comparisons of educational systems performance, for which information was available for both years. In general, these studies consider:

(i) A measure of the intensity of the human resources involved in the teaching-learning process (e.g. Agasisti and Zoido 2015; Aristovnik and Obadić 2014; Agasisti 2014; Sutherland et al 2009; Afonso and St. Aubyn 2005, 2006). In our case we use as a proxy the ratio of teachers per 100 students in secondary education.

(ii) A measure of the availability and quality of the main elements of the school's physical infrastructure that affect students' learning process such as school buildings and grounds; heating/cooling and lighting systems; and instructional space (De Jorge and Santín 2010; Giménez et al. 2007). In our case we use the index of quality of physical Infrastructure provided by PISA.

(iii) A measure of the socioeconomic characteristics of students and their families (e.g. Agasisti and Zoido 2015; Thieme et al. 2012; Giambona et al. 2011; Sutherland et al. 2009; Giménez et al. 2007). To measure this, PISA created the index of economic, social and cultural status (ESCS). As reported on the PISA website, this index "was created on the basis of the following variables: the International Socio-Economic Index of Occupational Status (ISEI); the highest level of education of the student's parents, converted into years of schooling; the PISA index of family wealth; the PISA index of home educational resources; and the PISA index of possessions related to *classical* culture in the family home."

The inputs were chosen at school level rather than at macroeconomic level (e.g. expenditure in secondary education as a % of GDP) to more accurately reflect the real impact of expenditures on the achievement of the students participating in the PISA sample.

The evolution between the two assessments (2003 and 2012) for these input variables for the sample average shows a disparate behavior. On the one hand, an improvement is seen in both the ratio of teachers per 100 students (average value increases from 6.48 to 7.61 teachers per

**Table 2** List of inputs and outputs

| | |
|---|---|
| $y_1$ | Mathematics score |
| $y_2$ | Reading score |
| $y_3$ | Mathematics score, standard deviation |
| $y_4$ | Reading score, standard deviation |
| $x_1$ | Teachers per 100 students ratio in primary education |
| $x_2$ | Index of quality of physical infrastructures |
| $x_3$ | ESCS |

100 students). In contrast, we observe a slight deterioration in the index of quality of physical infrastructure (from 9.97 in 2003 to 9.91 in 2012) and the index of economic, social and cultural status (ESCS) of the countries in the sample (from 9.78 to 9.73). All the index variables were transformed in order to avoid the original negative values.[6]

For all these variables we have information for 2003 (time $t$) and 2012 (time $t+1$). The list of inputs and outputs is reported in Table 2, and the corresponding values used in the empirical analysis are reported in Table 3.

**Table 3** Values for the inputs and outputs based on PISA (2003 and 2012)

| Country | Good output | | | | Bad output | | | | Inputs | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $y_1$ | | $y_2$ | | $y_3$ | | $y_4$ | | $x_1$ | | $x_2$ | | $x_3$ | |
| | 2003 | 2012 | 2003 | 2012 | 2003 | 2012 | 2003 | 2012 | 2003 | 2012 | 2003 | 2012 | 2003 | 2012 |
| Austria | 506 | 506 | 491 | 490 | 93 | 92 | 103 | 92 | 7.59 | 9.24 | 10.13 | 9.84 | 10.06 | 10.08 |
| Belgium | 529 | 515 | 507 | 509 | 110 | 102 | 110 | 102 | 8.42 | 8.93 | 10.08 | 9.85 | 10.15 | 10.15 |
| Canada | 532 | 518 | 528 | 523 | 87 | 89 | 89 | 92 | 6.33 | 7.50 | 10.19 | 10.32 | 10.45 | 10.41 |
| Czech Republic | 516 | 499 | 489 | 493 | 96 | 95 | 96 | 89 | 5.96 | 5.27 | 10.57 | 10.45 | 10.16 | 9.93 |
| Finland | 544 | 519 | 543 | 524 | 84 | 85 | 81 | 95 | 6.12 | 7.37 | 9.76 | 9.68 | 10.25 | 10.36 |
| Germany | 503 | 514 | 491 | 508 | 103 | 96 | 109 | 91 | 7.12 | 8.54 | 10.14 | 9.97 | 10.16 | 10.19 |
| Greece | 445 | 453 | 472 | 477 | 94 | 88 | 105 | 99 | 8.38 | 10.87 | 9.58 | 9.81 | 9.85 | 9.94 |
| Hungary | 490 | 477 | 482 | 488 | 94 | 94 | 92 | 92 | 10.43 | 9.54 | 9.82 | 10.21 | 9.93 | 9.75 |
| Ireland | 503 | 501 | 515 | 523 | 85 | 85 | 87 | 86 | 5.36 | 6.23 | 9.72 | 9.97 | 9.92 | 10.13 |
| Italy | 466 | 485 | 476 | 490 | 96 | 93 | 101 | 97 | 9.24 | 10.12 | 9.97 | 9.67 | 9.89 | 9.95 |
| Japan | 534 | 536 | 498 | 538 | 101 | 94 | 106 | 99 | 5.11 | 5.85 | 9.91 | 9.87 | 9.92 | 9.93 |
| Luxembourg | 493 | 490 | 479 | 488 | 92 | 95 | 100 | 105 | 8.70 | 11.95 | 9.85 | 9.51 | 10.18 | 10.07 |
| Mexico | 385 | 413 | 400 | 424 | 85 | 74 | 95 | 80 | 3.75 | 3.57 | 9.90 | 9.60 | 8.87 | 8.89 |
| New Zealand | 523 | 500 | 522 | 512 | 98 | 100 | 105 | 106 | 5.62 | 6.86 | 10.25 | 10.03 | 10.21 | 10.04 |
| Poland | 490 | 518 | 497 | 518 | 90 | 90 | 96 | 87 | 7.91 | 9.81 | 10.29 | 10.50 | 9.80 | 9.79 |
| Portugal | 466 | 487 | 478 | 488 | 88 | 94 | 93 | 94 | 9.00 | 8.53 | 10.03 | 9.74 | 9.37 | 9.52 |
| Slovak Republic | 498 | 482 | 469 | 463 | 93 | 101 | 93 | 104 | 5.44 | 6.69 | 9.69 | 9.87 | 9.92 | 9.82 |
| South Korea | 542 | 554 | 534 | 536 | 92 | 99 | 83 | 87 | 3.32 | 6.21 | 10.57 | 9.82 | 9.90 | 10.01 |
| Spain | 485 | 484 | 481 | 488 | 88 | 88 | 95 | 92 | 7.21 | 7.94 | 10.13 | 10.01 | 9.70 | 9.81 |
| Sweden | 509 | 478 | 514 | 483 | 95 | 92 | 96 | 107 | 8.94 | 10.36 | 10.03 | 10.21 | 10.25 | 10.28 |
| United States | 483 | 481 | 495 | 498 | 95 | 90 | 101 | 92 | 6.75 | 6.93 | 10.29 | 10.46 | 10.30 | 10.17 |
| Brazil | 356 | 391 | 403 | 410 | 100 | 78 | 111 | 85 | 4.64 | 4.88 | 9.94 | 9.65 | 9.05 | 8.83 |
| Hong Kong-China | 550 | 561 | 510 | 545 | 100 | 96 | 85 | 85 | 5.06 | 6.93 | 9.99 | 9.98 | 9.24 | 9.21 |
| Indonesia | 360 | 375 | 382 | 396 | 81 | 71 | 76 | 75 | 4.93 | 5.38 | 9.47 | 9.48 | 8.74 | 8.20 |
| Latvia | 483 | 491 | 491 | 489 | 88 | 82 | 90 | 85 | 7.30 | 9.08 | 10.06 | 10.38 | 10.12 | 9.74 |
| Macao-China | 527 | 538 | 498 | 509 | 87 | 94 | 67 | 82 | 3.85 | 7.09 | 9.75 | 9.89 | 9.11 | 9.11 |
| Russian Federation | 468 | 482 | 442 | 475 | 92 | 86 | 93 | 91 | 6.01 | 5.10 | 9.90 | 10.17 | 9.91 | 9.89 |
| Tunisia | 359 | 388 | 375 | 404 | 82 | 78 | 96 | 88 | 4.64 | 5.83 | 9.66 | 8.75 | 8.66 | 8.81 |
| Uruguay | 422 | 409 | 434 | 411 | 100 | 89 | 121 | 96 | 4.80 | 8.02 | 9.35 | 9.59 | 9.65 | 9.12 |
| Mean | 481.75 | 484.32 | 479.09 | 486.16 | 92.69 | 90.01 | 95.59 | 92.22 | 6.48 | 7.61 | 9.97 | 9.91 | 9.78 | 9.73 |
| Std. Dev. | 56.15 | 47.97 | 43.71 | 40.99 | 6.72 | 7.75 | 11.31 | 7.85 | 1.86 | 2.02 | 0.29 | 0.36 | 0.50 | 0.54 |
| Mín | 356.02 | 375.11 | 374.62 | 396.12 | 80.51 | 71.36 | 66.93 | 75.38 | 3.32 | 3.57 | 9.35 | 8.75 | 8.66 | 8.20 |
| Max | 550.38 | 561.24 | 543.46 | 544.60 | 109.88 | 102.26 | 121.50 | 106.75 | 10.43 | 11.95 | 10.57 | 10.50 | 10.45 | 10.41 |

**Table 4** Temporal analysis results for the FDH type model

| Country | *FPI* | | $FPI^G$ | | *GNRMI* | *EC* | *BPC* |
|---|---|---|---|---|---|---|---|
| | 2003 | 2012 | 2003 | 2012 | | | |
| Austria | 0.7698 | 1.0000 | 0.7698 | 0.8106 | 1.0530 | 1.2991 | 0.8106 |
| Belgium | 0.8230 | 0.8530 | 0.7737 | 0.8530 | 1.1025 | 1.0364 | 1.0638 |
| Brazil | 1.0000 | 1.0000 | 0.6629 | 1.0000 | 1.5085 | 1.0000 | 1.5085 |
| Canada | 0.9145 | 1.0000 | 0.9145 | 0.8715 | 0.9529 | 1.0935 | 0.8715 |
| Czech Republic | 0.7879 | 1.0000 | 0.7879 | 0.8084 | 1.0259 | 1.2692 | 0.8084 |
| Finland | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Germany | 0.7087 | 0.9177 | 0.7087 | 0.9177 | 1.2948 | 1.2948 | 1.0000 |
| Greece | 1.0000 | 1.0000 | 1.0000 | 0.7558 | 0.7558 | 1.0000 | 0.7558 |
| Hong Kong-China | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Hungary | 0.7860 | 1.0000 | 0.7860 | 0.7803 | 0.9928 | 1.2723 | 0.7803 |
| Indonesia | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Ireland | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Italy | 0.7221 | 1.0000 | 0.7221 | 1.0000 | 1.3849 | 1.3849 | 1.0000 |
| Japan | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Latvia | 0.8216 | 1.0000 | 0.8216 | 1.0000 | 1.2171 | 1.2171 | 1.0000 |
| Luxembourg | 0.7680 | 1.0000 | 0.7680 | 1.0000 | 1.3021 | 1.3021 | 1.0000 |
| Macao-China | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mexico | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| New Zealand | 1.0000 | 0.8423 | 1.0000 | 0.8423 | 0.8423 | 0.8423 | 1.0000 |
| Poland | 0.7996 | 1.0000 | 0.7996 | 1.0000 | 1.2506 | 1.2506 | 1.0000 |
| Portugal | 0.7886 | 1.0000 | 0.7886 | 1.0000 | 1.2681 | 1.2681 | 1.0000 |
| Russian Fed. | 0.7375 | 1.0000 | 0.7375 | 1.0000 | 1.3559 | 1.3559 | 1.0000 |
| Slovak Republic | 1.0000 | 1.0000 | 1.0000 | 0.6931 | 0.6931 | 1.0000 | 0.6931 |
| South Korea | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Spain | 0.7934 | 1.0000 | 0.7934 | 0.8147 | 1.0269 | 1.2604 | 0.8147 |
| Sweden | 0.8142 | 0.8112 | 0.8142 | 0.7387 | 0.9073 | 0.9964 | 0.9105 |
| Tunisia | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| United States | 0.7549 | 0.8968 | 0.7549 | 0.8088 | 1.0714 | 1.1880 | 0.9019 |
| Uruguay | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mean | 0.8893 | 0.9766 | 0.8760 | 0.9205 | 1.0692 | 1.1149 | 0.9627 |
| Std. Dev. | 0.1143 | 0.0547 | 0.1208 | 0.1036 | 0.1855 | 0.1496 | 0.1421 |
| Max | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.5085 | 1.3849 | 1.5085 |
| Min | 0.7087 | 0.8112 | 0.6629 | 0.6931 | 0.6931 | 0.8423 | 0.6931 |

## 4 Results

### 4.1 Efficiency change, best practice gap change, and performance change

Table 4 shows the results of the evolution of performance for the educational systems analyzed during the 2003–2012 period by applying the FDH technology. On average, the evolution of productivity[7] in the period was positive, with *GNRMI* taking a value of 1.06. This improvement is mainly due to a positive efficiency change (*EC*), which offsets the negative technological change (*BPC*) observed during the period. Specifically, the *EC* was 1.11, representing an 11% improvement. However, compared to the global frontier, the frontier in 2012 fell over the period as shown by the average $BPC = 0.96$. Therefore, part of the efficiency improvements is due to worsening educational reference systems. The explanation for these results could lie in the severe economic crisis that affected many of the sample countries during part of the period, which probably forced many of them to control education spending by reducing inefficiencies. However, it seems that the situation could have adversely affected the frontier countries'

[7] We will refer to the concepts of productivity and performance interchangeably.

**Table 5** Temporal analysis results for the DEA type model

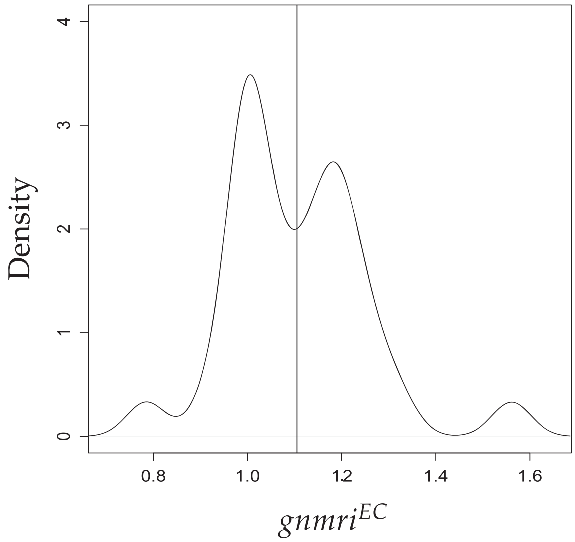
| Country | *FPI* | | *FPI*$^G$ | | *GNRMI* | *EC* | *BPC* |
|---|---|---|---|---|---|---|---|
| | 2003 | 2012 | 2003 | 2012 | | | |
| Austria | 0.7713 | 0.9061 | 0.7713 | 0.8165 | 1.0586 | 1.1747 | 0.9012 |
| Belgium | 0.7129 | 0.8393 | 0.7129 | 0.7576 | 1.0627 | 1.1772 | 0.9027 |
| Brazil | 0.5693 | 0.8887 | 0.5495 | 0.8396 | 1.5279 | 1.5610 | 0.9788 |
| Canada | 0.9123 | 0.9535 | 0.9123 | 0.8715 | 0.9553 | 1.0452 | 0.9140 |
| Czech Republic | 0.7938 | 0.9407 | 0.7938 | 0.8148 | 1.0264 | 1.1850 | 0.8661 |
| Finland | 1.0000 | 1.0000 | 1.0000 | 0.8993 | 0.8993 | 1.0000 | 0.8993 |
| Germany | 0.7119 | 0.8926 | 0.7119 | 0.8236 | 1.1570 | 1.2539 | 0.9227 |
| Greece | 0.8724 | 0.8464 | 0.7383 | 0.7573 | 1.0257 | 0.9702 | 1.0573 |
| Hong Kong-China | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Hungary | 0.7913 | 0.8569 | 0.7913 | 0.7848 | 0.9918 | 1.0828 | 0.9159 |
| Indonesia | 1.0000 | 1.0000 | 0.8122 | 1.0000 | 1.2313 | 1.0000 | 1.2313 |
| Ireland | 0.9357 | 1.0000 | 0.9031 | 0.9068 | 1.0040 | 1.0687 | 0.9395 |
| Italy | 0.7251 | 0.8877 | 0.7251 | 0.7966 | 1.0986 | 1.2243 | 0.8973 |
| Japan | 0.7569 | 1.0000 | 0.7569 | 0.8628 | 1.1399 | 1.3212 | 0.8628 |
| Latvia | 0.8269 | 1.0000 | 0.8269 | 0.9199 | 1.1125 | 1.2093 | 0.9199 |
| Luxembourg | 0.7705 | 0.8686 | 0.7705 | 0.7957 | 1.0326 | 1.1273 | 0.9160 |
| Macao-China | 1.0000 | 1.0000 | 1.0000 | 0.9486 | 0.9486 | 1.0000 | 0.9486 |
| Mexico | 1.0000 | 1.0000 | 0.7797 | 1.0000 | 1.2825 | 1.0000 | 1.2825 |
| New Zealand | 0.7803 | 0.8125 | 0.7803 | 0.7470 | 0.9574 | 1.0413 | 0.9194 |
| Poland | 0.8022 | 0.9638 | 0.8022 | 0.8778 | 1.0942 | 1.2014 | 0.9108 |
| Portugal | 0.7897 | 0.8890 | 0.7897 | 0.7874 | 0.9971 | 1.1258 | 0.8857 |
| Russian Fed. | 0.7413 | 0.9415 | 0.7413 | 0.8174 | 1.1026 | 1.2701 | 0.8681 |
| Slovak Republic | 0.8412 | 0.7723 | 0.8009 | 0.6958 | 0.8687 | 0.9182 | 0.9461 |
| South Korea | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Spain | 0.7952 | 0.8997 | 0.7952 | 0.8178 | 1.0285 | 1.1315 | 0.9090 |
| Sweden | 0.8142 | 0.8063 | 0.8142 | 0.7387 | 0.9073 | 0.9904 | 0.9161 |
| Tunisia | 1.0000 | 1.0000 | 0.6965 | 1.0000 | 1.4358 | 1.0000 | 1.4358 |
| United States | 0.7549 | 0.8907 | 0.7549 | 0.8147 | 1.0792 | 1.1799 | 0.9146 |
| Uruguay | 1.0000 | 0.7847 | 0.6691 | 0.7051 | 1.0539 | 0.7847 | 1.3431 |
| Mean | 0.8438 | 0.9187 | 0.8000 | 0.8482 | 1.0717 | 1.1050 | 0.9795 |
| Std. Dev. | 0.1176 | 0.0747 | 0.1052 | 0.0925 | 0.1465 | 0.1485 | 0.1488 |
| Max | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.5279 | 1.5610 | 1.4358 |
| Min | 0.5693 | 0.7723 | 0.5495 | 0.6958 | 0.8687 | 0.7847 | 0.8628 |

performance, which might explain the technological regress identified.

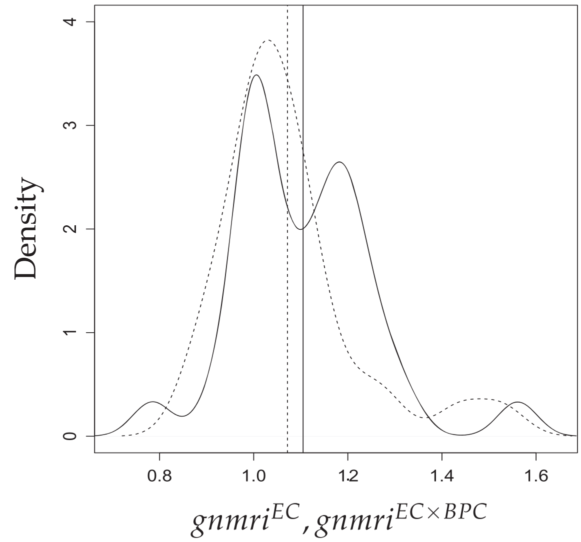
At the country-level analysis, one group of countries maintained productivity ($GNRMI = 1$). This group is made up of Finland, Hong Kong-China, Indonesia, Ireland, Japan, Macao-China, Mexico, South Korea, Tunisia and Uruguay. They lie on the contemporary frontier each year as indicated ($FPI = 1$). In another group of countries, Canada, Greece, Hungary, New Zealand, Slovak Republic and Sweden, performance worsened over the period, although the case of each country is different. Greece and Slovak Republic lie on the contemporary frontier in both years, although in 2012 their performance had worsened since 2003, as shown by $BPC < 1$. Hungary's efficiency improved, but the decline in its frontier area reference caused overall performance to fall. In fact, this country had almost no change in productivity because its efficiency improvement was accompanied by a similar fall of the frontier. Finally, the group of countries in which performance improved over the period includes Austria, Belgium, Brazil, Czech Republic, Germany, Italy, Latvia, Luxembourg, Poland, Portugal, Russian Federation, Spain and United States. Three countries, Brazil, Luxembourg and Russian Federation, has experienced the greatest improvement, with $GNRMI \geq 1.37$ in all the cases. The case of Brazil stands out as having the greatest improvement in performance in the sample ($GNRMI = 1.51$).

## Direction 1

(isolation of the efficiency change effect)



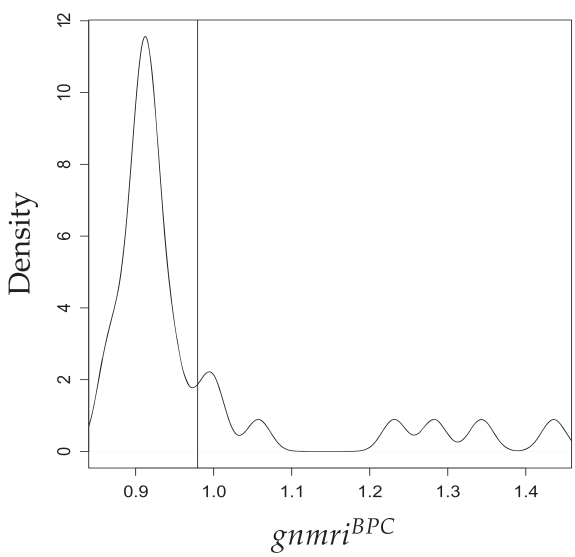

**a) Contribution of efficiency change**



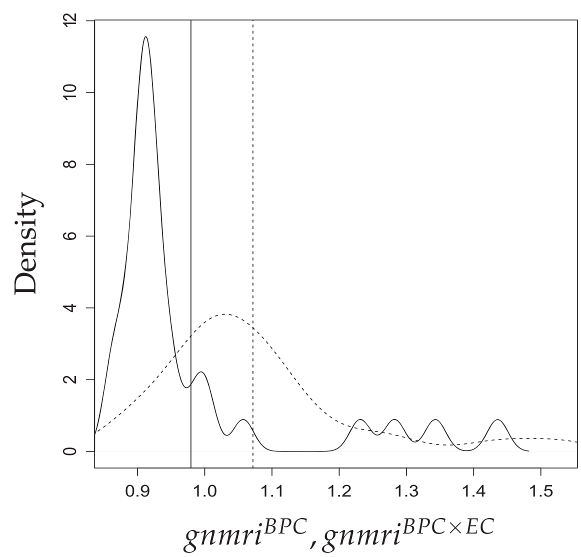

**b) Contribution of best practice gap change**

## Direction 2

(isolation of the best practice change effect)




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




**d) Contribution of efficiency change**

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

◀ **Fig. 1** Kernel density plots of the bipartite decomposition of educational improvement, DEA, global bandwidth. *Notes*: All figures contain densities estimated via kernel smoothing for the different components of the bipartite decomposition in expression (9), considering sequentially and in both directions how each component (*EC* or *BPC*) contribute to the change in educational performance (*gnmri*). The vertical lines in each plot represent the average for each component of the decomposition. Densities were estimated using a Gaussian kernel and the Sheather and Jones (1991) plug-in bandwidth (global bandwidth)

It lies on the contemporary frontier in both years, although with a substantial performance improvement in 2012. Austria, Belgium, Czech Republic, Spain and the United States saw increased productivity due to improvements in efficiency, and technological regress in their reference area on the frontier. Performance of the remaining countries in the group of best performers improved due exclusively to improvements in the efficiency.

As mentioned above, to check the robustness of the results we also report results for the DEA model assuming VRS (see Table 5). Overall, the results coincide with those reported in the preceding paragraphs. In this case, we also identify an improvement in productivity ($GNRMI = 1.07$), similar to that obtained for the FDH model. This change in productivity is the result of a positive change in efficiency ($EC = 1.10$) and a slight technological regress ($BPC = 0.98$). As expected, results at the country level show more marked differences, with the discriminatory power of the DEA models being the greatest. The group of countries where productivity fell is larger than the case of the FDH model, which contains Canada, Finland, Hungary, Macao-China, New Zealand, Portugal, Slovak Republic and Sweden. In most of these cases, the decline in productivity (performance) was mainly due to greater technological regress than deterioration in efficiency. Interesting cases are those of Finland or Macao-China, often regarded as educational benchmarks due to their students' performance. Both countries were part of the contemporary frontiers, although their performance deteriorated throughout the period. In fact, Finland dropped several positions in several indicators in the PISA report for 2012. Productivity in Hong Kong-China and South Korea stagnated during the period, with no remarkable changes in any of its components. The remaining countries experienced productivity improvements, caused mainly by increases in their efficiency levels.

The case of Brazil is worth mentioning: it has the best improvement in performance over the period with $GNRMI \leq 1.50$ in both models. In the Brazilian educational system students in both public and private schools take an external exam at the end of each academic year. If students fail to meet the minimum standards required, it is not unusual for them to repeat the grade, so it is relatively common to find students of different ages in the same class. This continuous external control could be one of the reasons to explain Brazil's improved productivity and technological change. Another feature of the Brazilian educational system is that students are not segregated according to their academic level. These two features might contribute to ensuring a good and relatively homogeneous academic performance.

For illustrative purposes, we also highlight the case of four countries that are contemporary efficient in both periods and models, FDH and DEA-VRS, namely, Finland, South Korea, Macao-China and Hong Kong-China. These are four culturally different countries, with differently focused and organized educational systems. For instance, the South Korean educational system introduces a high level of competition among students, while the Finnish model aims to achieve maximum levels of equality between students, facilitated by free (public) education, high teacher competence, students' personal development, and schools' autonomy. Meanwhile, the education systems of Hong Kong and Macao have been influenced in the final stage by Chinese educational culture, but are also significantly affected by British (in Hong Kong) and Portuguese (in Macao) colonial influences. In both countries' education systems, private, partially state-subsidized schools have a considerable share. Consequently, the results seem to suggest that no one educational model is clearly more efficient than another.

### 4.2 Bipartite decomposition of performance change

The results of the analysis proposed in Section 2.2 are reported in Fig. 1, whose sub-figures illustrate the sequential analysis of the contribution of each performance change component. Figure 1 is divided into two panels in order to show both directions of the sequential order of the analysis: the upper panel (Fig. 1a, b) corresponds to the decomposition in Eq. (9), whereas the lower panel (Fig. 1c, d) represents the decomposition in Eq. (11).

Given some of the particularities of the data used, and as indicated in Section 2.2, the densities were also estimated for different values of the smoothing parameter. Specifically, Fig. 2 reports an analogous analysis as that in Fig. 1 for a *global* bandwidth. In the case of Fig. 2, the amount of smoothing varies *locally*, depending on the structure of the data at a given point. In this regard, some problems related to the estimation of the smoothing parameter in the case of FDH, either in its local or global variants, impelled us to confine the analysis to DEA.

The analysis in the upper panel of Fig. 1 shows that the contribution of efficiency to the change of global non-radial

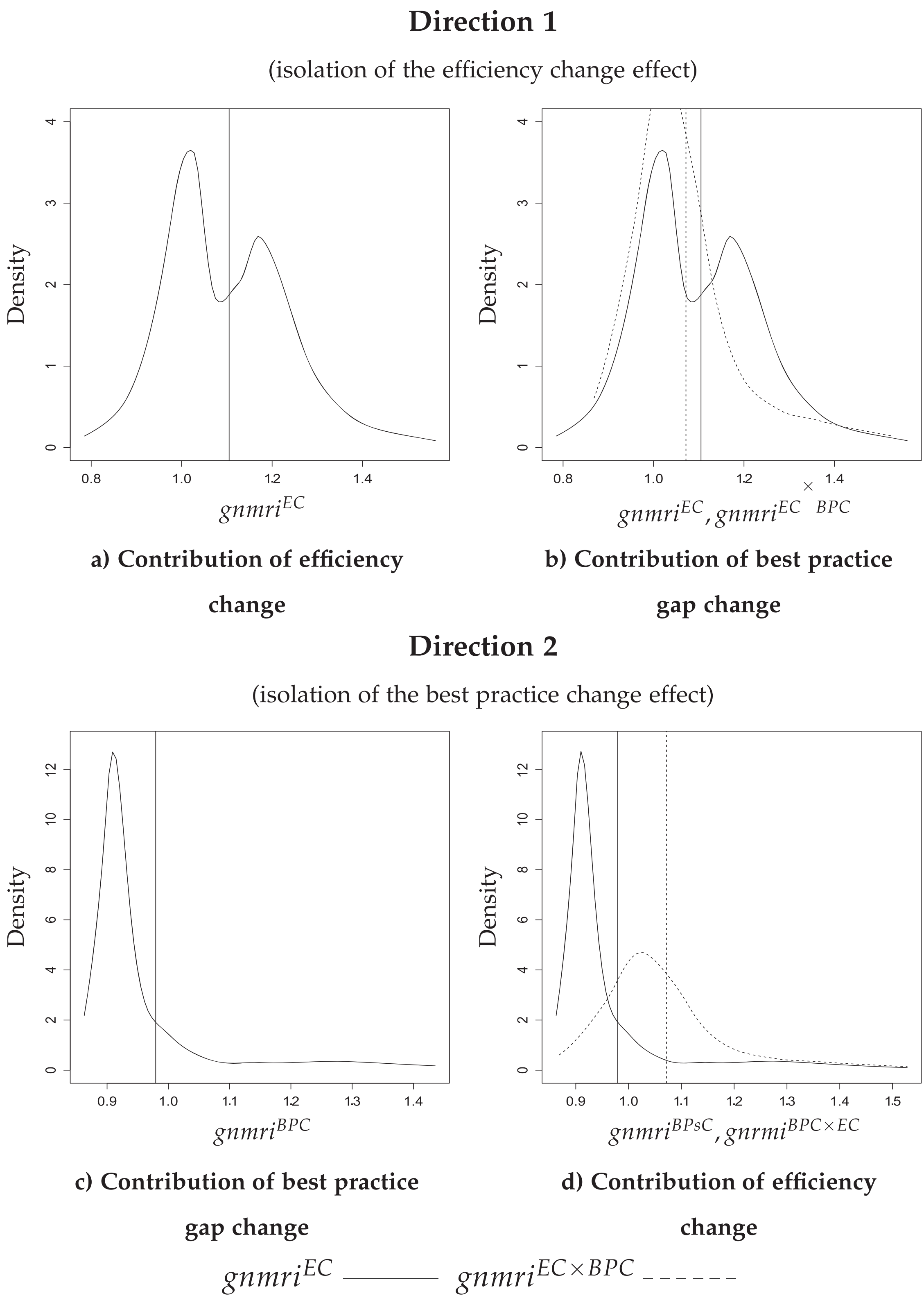



**Fig. 2** Kernel density plots of the bipartite decomposition of educational improvement, DEA, local bandwidth. *Notes*: All figures contain densities estimated using kernel smoothing for the different components of the bipartite decomposition in expression (9), considering sequentially and in both directions how each component (*EC* or *BPC*) contribute to the change in educational performance (*gnmri*). Densities were estimated using local likelihood methods (Loader 1996), and a Gaussian kernel was chosen

Malmquist index is very heterogeneous, as indicated by several bumps shown by the density corresponding to $gnmri^{EC}$, i.e., the counterfactual educational achievement change attributable to changes in efficiency (Fig. 1a). However, the contribution of the best practice change shown in Fig. 1b offsets the heterogeneity of $gnmri^{EC}$, leading to a much smoother density when the two effects are combined ($gnmri^{EC \times BPC}$). Actually, *on average*, as indicated by the vertical lines in Fig. 1a, b, although the effect of efficiency change ($gnmri^{EC}$) is positive (the solid vertical line is above 1), the contribution of the best practice gap leads to a more reduced combined effect (the dashed vertical line is closer to 1). The smoother lines depicted when choosing local bandwidths, as shown in Fig. 2a, b point in the same direction, excepting for the bumps corresponding to $gnmri^{EC}$, which are largely smoothed out in Fig. 2.

These discrepancies are also present when the sequential order is reversed, as shown in the lower panels of Fig. 1, for the global bandwidth, and Fig. 2, for the local bandwidth. The analysis performed in the reverse order indicates that the discrepancies for $gnmri^{BPC}$ are even higher than those for $gnmri^{EC}$; this is particularly apparent when a global bandwidth is chosen (Fig. 1). Therefore, countries follow very different paths to obtain their productivity change index.

## 5 Conclusions

In this paper we have implemented recent methodological proposals using data from the latest version of the Programme for International Student Assessment (PISA) to analyze the performance of educational systems consistent with the trend of public education policies calling for higher levels of academic achievement combined with lower levels of inequality. In this respect, PISA offers provides a framework for evaluating and describing students' learning processes in participating countries, for the two disciplines analyzed (mathematics and reading), providing relevant information on additional factors involved in these processes.

We considered the global non-radial Malmquist index (*GNRMI*), which is particularly interesting in the context of education. This index is not only appropriate for its highly desirable properties; it also suits our context because it incorporates bad outputs which, ideally, educational systems should minimize while simultaneously maximizing the outputs—or, more correctly, good/desirable outputs.

Since the variables used in the evaluation are fixed ratios, and following recent recommendations in the literature to deal with them, we used Free Disposal Hull (FDH) technology. In the same vein, and in order to check the robustness of the results, we also performed the calculations using the variable returns to scale version of Data Envelopment Analysis (DEA-VRS).

The results of the different evaluations of the global non-radial Malmquist index show, on average, a positive evolution in educational systems' performance during the 2003–2012 period, with the global non-radial Malmquist index taking a value of 1.06. This improvement is mainly due to a positive efficiency change ($EC = 1.11$), which offset the negative technological change observed ($BPC = 0.96$).

These results coincide regardless of the methodology considered (FDH or DEA-VRS). In this case, and as expected, the comparatively weaker ability of FDH to discriminate with respect to DEA is offset by its higher methodological rigor given the characteristics of the variables.

In light of the economic crisis faced by the vast majority of countries in our sample during this period, the average results obtained appear to be good news. Indeed, considering the arguments in the preceding paragraphs, the results from this analysis based on averages could be interpreted in the sense that tighter controls on educational spending might have implied a stronger effort to eliminate inefficiencies in the education sector. In addition, the gains in technical efficiency offset the technological regress. However, an average might provide a misleading view of what the data conceal. Deeper scrutiny of results at the country level shows great variations in performance from one country to another, and the conclusions drawn from the average analysis might not be generalized. Indeed, out of the 27 countries experiencing either stagnation or technological regress (i.e., best-practice gap change, $BPC \leq 1$), only ten saw improvements in overall performance—i.e., technological regress was offset by efficiency improvements.

This heterogeneity in performance can also be observed when classifying educational systems according to the components of the global non-radial Malmquist index (*GNMRI*)—efficiency change (*EC*) and best-practice gap change (*BPC*). The countries in the first group would be those whose overall performance (productivity) improves due to improvements in efficiency (Austria, Czech Republic, Germany, Italy, Latvia, Luxembourg, Poland, Portugal, Russian Federation, Spain and United States). Overall performance also increases in Brazil, although in this case due to technological progress. The third group contains 10 countries (Finland, Hong Kong-China, Indonesia, Ireland, Japan, Macao-China, Mexico, South Korea, Tunisia and Uruguay), all of which show stability for the three indicators. Overall performance deteriorates in the two remaining groups: countries in group 4 (Canada, Greece, Hungary, Slovak Republic ans Sweden) mainly due to unfavourable technological change, and group 5 (which is not technically a group since it has only one country, New Zealand) led by the negative efficiency change.

Although the study has achieved its objectives, there are three issues that should be considered in future investigations. The first one is related to the unit of analysis. Indeed, although, like ours, most other studies comparing the performance of educational systems consider information at the country level, there is also a recent trend to use data at either the school or the student level. The second issue is related to comparatively recent contributions such as the metafrontier and partial frontiers, which open up newer fields for methodological contributions. Finally, we also recognize as a limitation of the study the lack of any analysis of differences in performance, which would provide even greater insights in terms of public policy.

Taking into account these limitations, it is part of our immediate research agenda to continue with similar analyses, using order-*m* metafrontiers with school-level data as well as analyzing of the determinants. This would strengthen the analysis in terms of its explanatory power, as well as provide additional guidance for public policy-makers, in order to merge three increasingly critical issues in education: academic achievement, equality, and efficiency.

**Acknowledgements** Claudio Thieme and Emili Tortosa-Ausina thank FONDECYT (National Fund of Scientific and Technological Development, grant #1121164 and #1151313) for generous financial support. Víctor Giménez, Diego Prior and Emili Tortosa-Ausina acknowledge the financial support of the Ministerio de Economía y Competitividad (ECO2013-44115-P and ECO2014-55221-P). Emili Tortosa-Ausina also acknowledges the financial support of Generalitat Valenciana (PROMETEOII/2014/046) and Universitat Jaume I (P1.1B2014-17). All four authors are grateful to the Associate Editor and two anonymous referees, whose comments contributed to an overall improvement of the paper. The usual disclaimer applies.

**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

Afonso A, St. Aubyn M (2005) Non-parametric approaches to education and health effciency in OECD countries. J Appl Econ 8 (2):227–246

Afonso A, St. Aubyn M (2006) Cross-country effciency of secondary education provision: a semi-parametric analysis with non-discretionary inputs. Econ Model 23(3):476–491

Agasisti T (2014) The efficiency of public spending on education: an empirical comparison of EU countries. Eur J Educ 49(4):543–557

Agasisti T and Zoido P (2015). The efficiency of secondary schools in an international perspective: preliminary results from PISA 2012. OECD Education Working Papers 117, OECD, Paris.

Aparicio J, Crespo-Cebada E, Pedraja-Chaparro F, Santín D (2016a) Comparing school ownership performance using a pseudo-panel database: a Malmquist-type index approach. Eur J Oper Res 256 (2):533–542.

Aparicio J, Pastor JT, Vidal F (2016b) The directional distance function and the translation invariance property. Omega 58:1–3

Aristovnik A, Obadić A (2014) Measuring relative efficiency of secondary education in selected EU and OECD countries: the case of Slovenia and Croatia. TEDE 20(3):419–433

Badunenko O, Romero-Ávila D (2013) Financial development and the sources of growth and convergence. Int Econ Rev 54(2):629–663

Balk BM (2001) Scale efficiency and productivity change. J Prod Anal 15(3):159–183

Brown, G, Micklewright, J, Schnepf, SV, and Waldmann, R (2007). International surveys of educational achievement: how robust are the findings? J R Stat Soc A 170(3):623–646

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 DW, Christensen LR, Diewert WE (1982) The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50(6):1393–1414

Chung YH, Färe R, Grosskopf S (1997) Productivity and undesirable outputs: a directional distance function approach. J Environ Manage 51:229–240

Clements B (2002) How efficient is education spending in Europe? Eur Rev Econ Finance 1(1):3–26

Cooper, WW, Seiford, LM, and Tone, K (2007). Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. Springer Science & Business Media, New York

Cordero JM, Santín D, Simancas R (2017) Assessing European primary school performance through a conditional nonparametric model. J Oper Res Soc in press.

Denvir B, Brown M (1986) Understanding of number concepts in low attaining 7–9 year olds: Part I. development of descriptive framework and diagnostic instrument. Educ Stud Math 17(1):15–36

Deprins D, Simar L, Tulkens H (1984) Measuring labor-efficiency in post offices. In: Marchand M, Pestieau P, Tulkens H (eds) The performance of public enterprises: concepts and measurement. North-Holland, Amsterdam, p 243–267. Chapter 10

Deutsch J, Dumas A, Silber J (2013) Estimating an educational production function for five countries of Latin America on the basis of the PISA data. Econ Educ Rev 36:245–262

Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. Eur J Oper Res 132 (2):260–273

Emrouznejad A, Parker BR, Tavares G (2010) Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA. Socioecon Plann Sci 42 (3):151–157

Ercikan K (2006) Examining guidelines for developing accurate proficiency level scores. Can J Educ 29(3):823–838

Färe R, Grosskopf S (2004) Modeling undesirable factors in efficiency evaluation: comment. Eur J Oper Res 157(1):242–245

Färe R, Grosskopf S (2009) A comment on weak disposability in nonparametric production analysis. Am J Agric Econ 91 (2):535–538

Färe R, Grosskopf S, Lovell CAK (1994a) Production Frontiers. Cambridge University Press, Cambridge

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

Färe R, Grosskopf S, Norris M, Zhang Z (1994b) Productivity growth, technical progress, and efficiency change in industrialized countries. Am Econ Rev 84(1):66–83

Färe R, Grosskopf S, Pasurka C (1989) Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. Rev Econ Stat 71(1):90–98

Giambona F, Vassallo E, Vassiliadis E (2011) Educational systems efficiency in European Union countries. Stud educ eval 37 (2):108–122

Giménez V, Prior D, Thieme C (2007) Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. J Oper Res Soc 58(8):996–1007

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

Golany B, Thore S (1997) The economic and social performance of nations: efficiency and returns to scale. Socioecon Plann Sci 31 (3):191–204

Grifell-Tatjé E, Kerstens K (2008) Incentive regulation and the role of convexity in benchmarking electricity distribution: economists versus engineers. Ann Public Coop Econ 79(2):227–248

Grosskopf S, Hayes KJ, Taylor LL (2014) Efficiency in education: research and implications. Appl Econ Perspect Policy 36 (2):175–210

Gupta S, Verhoeven M (2001) The efficiency of government expenditure: experiences from Africa. J Policy Model 23(4):433–467

Hailu A, Veeman TS (2001) Non-parametric productivity analysis with undesirable outputs: an application to the Canadian pulp and paper industry. Am J Agric Econ 83(3):605–616

Henderson DJ, Parmeter CF (2015) Applied Nonparametric Econometrics. Cambridge University Press, Cambridge, MA

Henderson DJ, Russell RR (2005) Human capital and macroeconomic convergence: a production-frontier approach. Int Econ Rev 46 (4):1167–1205

Hollingsworth B, Smith P (2003) Use of ratios in data envelopment analysis. Appl Econ Lett 10(11):733–735

Jacob WJ, Holsinger DB (2008) Inequality in education: a critical analysis. In: Jacob WJ, Holsinger DB (eds) Inequality in education. Springer, Hong Kong, p 1–33

Johnes J (2015) Operational research in education. Eur J Oper Res 243 (3):683–696

Johnes J (2004) Efficiency measurement. In: Johnes G, Johnes J (eds) The international handbook on the economics of education. Edward Elgar, Cheltenham, UK

De Jorge J, Santín D (2010) Determinantes de la eficiencia educativa en la Unión Europea. Hacienda Publica Española/Revista de Economía Pública 193(2):131–155

Kumar S (2006) Environmentally sensitive productivity growth: a global analysis using malmquist–luenberger index. Ecol Econ 56 (2):280–293

Kumar S, Russell RR (2002) Technological change, technological catch-up, and capital deepening: relative contributions to growth and convergence. Am Econ Rev 92(3):527–548

Kuosmanen T (2005) Weak disposability in nonparametric production analysis with undesirable outputs. Am J Agric Econ 87:1077–1082

Kuosmanen T, Podinovski VV (2009) Weak disposability in nonparametric production analysis: reply to Färe and Grosskopf. Am J Agric Econ 91:539–545

Li Q, Racine JS (2007) Nonparametric econometrics: theory and practice. Princeton University Press, Princeton and Oxford

Loader CR (1996) Local likelihood density estimation. Ann Stat 24 (4):1602–1618

Luenberger D (1992) New optimality principles for economic efficiency and equilibrium. J Optim Theory Appl 75(2):221–264

Nakano M, Managi S (2008) Regulatory reforms and productivity: an empirical analysis of the Japanese electricity industry. Energy Policy 36(1):201–209

OECD (2014) PISA 2012 technical report. Technical report, OECD Publishing, Paris.

Oh D (2010) A global Malmquist-Luenberger productivity index. J Prod Anal 34(3):183–197

Olesen OB, Petersen NC, Podinovski VV (2015) Efficiency analysis with ratio measures. Eur J Oper Res 245(2):446–462

Pastor JT, Lovell CAK (2005) A global Malmquist productivity index. Econ Lett 88(2):266–271

Picazo-Tadeo AJ, Prior D (2009) Environmental externalities and efficiency measurement. J Environ Manage 90(11):3332–3339

Quah DT (1993a) Empirical cross-section dynamics in economic growth. Eur Econ Rev 37:426–434

Quah DT (1993b) Galton's fallacy and tests of the convergence hypothesis. Scand J Econ 95(4):427–443

Reinhard S, Lovell CAK, Thijssen GJ (2002) Analysis of environmental efficiency variation. Am J Agric Econ 84:1054–1065

Scarf HE (1994) The allocation of resources in the presence of indivisibilities. J Econ Perspect 8(4):111–128

Scott DW (1992) Multivariate density estimation: theory, practice, and visualization. Wiley, New York, NY

Seiford LM, Zhu J (2002) Modeling undesirable factors in efficiency evaluation. Eur J Oper Res 142(1):16–20

Sheather SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. J R Stat Soc B 53 (3):683–690

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London

Sueyoshi T, 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? Eur J Oper Res 211(1):76–89

Sutherland D, Price R, Gonand F (2009) Improving public spending efficiency in primary and secondary education. OECD Econ Stud 2009(4):1–30

Teddlie C, Reynolds D (2000) The international handbook of school effectiveness research. Routledge, London

Thieme C, Giménez V, Prior D (2012) A comparative analysis of the efficiency of national educational systems. APER 13:1–15

Tone K, Sahoo BK (2003) Scale, indivisibilities and production function in data envelopment analysis. Int J Prod Econ 84 (2):165–192

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

Weber WL, Domazlicky B (2001) Productivity growth and pollution in state manufacturing. Rev Econ Stat 83(1):195–199

Wenglinsky H (1998) Finance equalization and within-school equity: the relationship between education spending and the social distribution of achievement. Educ Eval Policy Anal 20(4):269–283

De Witte K, López-Torres L (2017) Efficiency in education: a review of literature and a way forward. J Oper Res Soc in press.

Worthington AC (2001) An empirical survey of frontier efficiency measurement techniques in education. Educ Econ 9:245–268

Xue M, Harker PT (2002) Note: ranking DMUs with infeasible super-efficiency DEA models. Manage Sci 48(5):705–710

Yörük BK, Zaim O (2005) Productivity growth in OECD countries: a comparison with Malmquist indices. J Comp Econ 33 (2):401–420

Zhang N, Choi Y (2013) Total-factor carbon emission performance of fossil fuel power plants in China: a metafrontier non-radial Malmquist index analysis. Energ Econ 40:549–559

Zhang N, Zhou P, Choi Y (2013) Energy efficiency, $CO_2$ emission performance and technology gaps in fossil fuel electricity generation in Korea : a meta-frontier non-radial directional distance function analysis. Energy Policy 56:653–662

Zhou P, Ang BW, Wang H (2012) Energy and $CO_2$ emission performance in electricity generation: a non-radial directional distance function approach. Eur J Oper Res 221(3):625–635