European Journal of Operational Research 294 (2021) 651–672

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Comparing group performance over time through the Luenberger productivity indicator: An application to school ownership in European countries

Juan Aparicio[a], Lidia Ortiz[a,*], Daniel Santín[b]

[a] Center of Operations Research (CIO), Miguel Hernandez University of Elche, Elche, Alicante 03202, Spain
[b] Complutense Institute of Economic Analysis, Complutense University of Madrid, Campus de Somosaguas, Madrid 28223, Spain

A R T I C L E I N F O

Article history:
Received 11 October 2019
Accepted 2 February 2021
Available online 8 February 2021

Keywords:
DEA
Group performance
Directional Distance Function
Luenberger Productivity Indicator
School ownership

A B S T R A C T

This paper extends the Camanho and Dyson (2006) one-period Malmquist-type index (CDMI) and the recent pseudo-panel Malmquist index (PPMI) by Aparicio et al. (2017) and Aparicio and Santín (2018) to a context where additive efficiency measures are used. In particular, we apply the Luenberger productivity indicator. Unlike the CDMI, the new approach is based upon the directional distance function, allowing non-equiproportional changes in the input and output mix and variable returns to scale for comparing the efficiency and technology gaps of two or more groups of production units over time. To illustrate this methodology, we estimate how the productivity gaps between publicly funded private schools (PFPS) and public schools (PS) in eight European Union countries changed over the 2009–15 period using PISA data. Our results suggest that the performance of PFPS is better in Belgium, Ireland, the Netherlands and Spain in both waves, while PS productivity outperforms PFPS in the Czech Republic, Hungary and Slovakia. Both school types operate with a productivity gap close to zero in Denmark. In addition, we observe that despite being less efficient, PS are more productive than PFPS, thanks to their better production technology. Finally, we find that school autonomy is positively related to school productivity explaining why PFPS present higher productivity than PS in some countries.

1. Introduction

Public education can be produced by either public schools (PS) or by publicly funded private schools (PFPS). In PS, school ownership is public, and schools are monitored and managed by a public education principal, selected in most cases by the public sector. PFPS are owned by a non-public organization,[1] and the governing board is not elected by a government agency. Furthermore, PFPS have wider decision-making powers than PS concerning management and more flexibility for hiring and firing teachers or for deciding budget priorities. In some European countries, PFPS represent a non-negligible percentage of educational production, meaning that many families choose this option for educating their children. For this reason, an immediate question for policymakers is whether PFPS are more efficient at producing cognitive skills than their PS counterparts having accounted for student and school resources.

On one hand, some of the previous educational literature argues that, according to economic theory, PFPS are likely to perform better than public schools because market competitive pressure, combined with school choice freedom and a more flexible management, should lead to a more efficient use of resources (Epple, Romano & Zimmer, 2016; Hoxby, 2003; Rouse & Barrow, 2009). On the other hand, as PFPS are private entities free to choose their location, some authors claim that PFPS do not cover all populations because they tend to be disproportionally placed in middle- and high-income neighborhoods, where expected profitability is higher. This raises concerns about the coverage of minorities and the generation of inequalities (Frankenberg, Siegel-Hawley & Wang, 2010).

One way to tackle the analysis of differences between both models of educational production and how they evolve over time within a country and across different countries is through the measurement of efficiency and productivity (Farrell, 1957; Levin, 1974). Benchmarking schools is a good strategy to detect best practices and reference units in order to measure the degree of efficiency

* Corresponding author.
E-mail address: lidia.ortiz@umh.es (L. Ortiz).

[1] The organizations are mostly Catholic schools or other religious schools, teachers' cooperatives, non-for-profit organizations, trade union or simply private enterprises. In this paper, we do not include fully private schools because they are funded mainly by student families. Webbink (2005), Urquiola (2016) and Green (2020) provide an excellent overview on the findings of educational research with regard to a public versus private school performance comparison from a theoretical and empirical viewpoint.

https://doi.org/10.1016/j.ejor.2021.02.005
0377-2217/

of inefficient units and analyze how total factor productivity varies among groups of decision-making units (DMUs) over time. Previous research has analyzed and compared the efficiency of public and private schools from an international perspective using schools as production units. For example, Agasisti and Zoido (2018) derive efficiency measures for about 8500 schools in 30 countries using PISA 2012 data concluding, after performing a second stage regression, that private schools were more efficient than public schools. Similar results were found in Aparicio et al., 2018 analyzing around 11,700 schools from 34 OECD countries using the same database.

In this paper, we are particularly interested in studies explicitly distinguishing between private independent and private government dependent schools (Dronkers & Robert, 2008). This distinction is relevant because PFPS receive most of their core funding from the public sector and the comparison with PS allows monitoring the efficiency of public spending in education. In this context, several papers have considered the PS versus PFPS efficiency comparison using cross-sectional data through either parametric stochastic frontier analysis (Cordero, Crespo, Pedraja & Santín, 2011; Crespo-Cebada, Pedraja-Chaparro & Santín, 2014; Perelman & Santín, 2011a, 2011b) or non-parametric production frontiers, especially data envelopment analysis (DEA) (Cordero, Crespo-Cebada & Santín, 2010; Kirjavainen & Loikkanen, 1998; Mancebón & Muñiz, 2008; Mancebón, Calero, Choi & Ximenez, 2012; Segovia-Gonzalez, Dominguez & Contreras, 2020). The results of this literature are inconclusive and provide mixed evidence about the superiority of either school type (Cherchye, De Witte, Ooghe & Nicaise, 2010; De Witte & López-Torres, 2017; Sutherland, Price & Gonand, 2009).

Despite the importance of this issue for monitoring the efficiency of schools for evidence-based educational policy decision making, there are only a small number of papers examining total factor productivity (TFP) changes of schools over time (Aparicio, Lopez-Torres & Santín, 2018; Bradley, Johnes & Little, 2010; Brennan, Haelermans & Ruggiero, 2014; Essid, Oullette & Vigeant, 2014; Johnson & Ruggiero, 2014; Maragos & Despotis, 2004; Portela, Camanho & Keshvari, 2013). The main explanation for this dearth of research is perhaps the lack of national accountability systems for gathering school data to perform longitudinal studies on a panel of schools. While measuring TFP analysis within a country or region is challenging, it is even more complicated to benchmark education systems internationally. Recent international databases like TIMSS (Trends in International Mathematics and Science Study), PISA (Programme for International Student Assessment) or PIRLS (Progress in International Reading Literacy Study) might be a way to overcome the shortage of data (Cordero, Cristobal & Santín, 2018a). Nevertheless, their use has the added problem of dealing with pseudo-panel data, i.e. the analysis of random representative samples of cross-sectional international data, where participant schools and students differ from one wave to another.

Recently, in an effort to overcome this problem, Aparicio, Crespo-Cebada, Pedraja-Chaparro and Santín (2017) extended the well-known Camanho and Dyson (2006) group performance index (CDMI) for analyzing productivity gaps among two or more groups of production units over time using a pseudo-panel Malmquist index (PPMI). The PPMI can deal with both panel data and pseudo-panel data. Additionally, Aparicio and Santín (2018) enhanced both of the above proposals by introducing a new index that assumes a baseline group as reference technology. In this manner, the new base-group CDMI and the base-group base-period PPMI indexes can satisfy the circular relation property directly.[2] All these previous efforts are based on a Malmquist index that can be decomposed into technology and efficiency gaps. However, the underlying technology of both indexes assumes constant returns to scale (CRS), which rules out a more flexible variable returns to scale (VRS) assumption. Additionally, it is based on the very rigid Shephard distance function (Shephard, 1953), which imposes equiproportional input or output changes for reaching the production frontier.

In this paper, we propose a new approach, based on the Luenberger productivity indicator, for comparing the performance of two or more groups of production units over time using panel data or pseudo-panel data. This new index inherits some interesting features from previous indexes, especially the property of circularity, the use of a baseline group as the reference technology, and the decomposition of the productivity gap into technology and efficiency gaps. Additionally, it provides two new advantages. First, in contrast to the Shephard distance functions used by the Malmquist index, the Luenberger indicator is based upon the directional distance function (DDF), a graph measure that permits non-equiproportional changes in the input and output mix. Second, unlike the Malmquist index, the assumption of CRS can be relaxed when the Luenberger indicator is computed. In our context, this implies that it is possible to compare the best-practice technologies of two or more assessed groups of units under VRS, the most usual assumption when DEA is applied in practice. Additionally, the new methodology allows undesirable outputs to be incorporated directly, a problem whose solution is not so obvious in the case of using Malmquist indexes (see, for example, Färe, Grosskopf & Margaritis, 2008).

To illustrate the potential of our index, this paper includes an empirical application to address the following questions. First, we measure productivity gaps between PS and PFPS across eight European countries between two time periods (from 2009 to 2015) using PISA data. Second, we decompose these productivity gaps in order to analyze whether differences within and between school ownerships and countries are explained by efficiency gaps or by technical gaps. Third, as this period coincides with the beginning and the end of the financial crisis, we analyze how the initial gaps evolved over time. And finally, we provide a robustness check for exploring how different would the results be if another reference would have been chosen. In short, our study offers a basis for developing a new tool for monitoring and evaluating the efficiency and the equality of educational opportunities in education systems funded by the public sector.

To avoid any confusion, we only measure the productivity gaps by school type and how they evolve over time, but we do not try to prove a straight causal relationship for pointing out that the school ownership is the final cause of the productivity differences found. School ownership is not exogenously distributed among the population and parents choose their favorite school based on factors such as their location, religion, size, ideology or expectations and these factors might also vary across countries. Public administrations can also modulate the percentage of public schools with respect to the total, through the legal framework. However, an additional appealing factor of this methodology is that initial productivity differences could be used in future applications as a baseline to build up a difference-in-differences analysis, using efficiency of productivity scores as dependent variables (Agrell, Mattsson & Månsson, 2020; Bravo-Ureta, González-Flores, Greene & Solís, 2020). Therefore, in the case of natural experiments or changes in the legal framework affecting some groups of schools but not others, this monitoring system might be used to measure the impact of such changes on total factor productivity. For illustrating the potential use of this methodology, we estimate the effect of school autonomy on differences in productivity and its

[2] Circularity (Frisch, 1936) implies that, in frameworks where it is relevant to compare the average performance of more than two groups, the direct comparison between two groups is equivalent to their indirect comparison through a third group, whatever the third group selected for the assessment is.

components, within the context of different countries and types of schools over the two PISA waves.

The remainder of this paper is organized as follows. Section 2 reviews the previous background introducing the CDMI, the base-group CDMI, the base-group base-period PPMI and the traditional Luenberger productivity indicator. Section 3 introduces the ideas of the new indexes and their properties, based on the previous research. Section 4 briefly describes the database, the variables included in the analysis, and discusses the issue of the reference choice. Section 5 provides the empirical results and the main findings together with a robustness check of the results. Finally, Section 6 discusses the main conclusions and implications of this research for educational policymakers and future research lines.

2. Background

In this section, we give a brief description of the Malmquist index that measures the change in productivity, as well as the adaptation of this index proposed by Camanho and Dyson (2006) to compare the relative performance of two groups in a single period of time. We also provide the adaptation by Aparicio and Santín (2018) to measure this same relative performance between two or more groups over more than one period of time. All these definitions will be useful in order to introduce the new approach in Section 3.

We consider n DMUs that use m inputs to produce q outputs at period t. Let us define an input vector as $X^t \in R^m_+$ and an output vector as $Y^t \in R^q_+$, which come from a reference technology $T^t = \{(X^t, Y^t) \in R^m_+ \times R^q_+ : X^t \text{ produces} Y^t\}$. Particularly, in this paper, T^t is estimated using DEA assuming VRS (Banker, Charnes & Cooper, 1984) as $T^t_V = \{(X^t, Y^t) \in R^m_+ \times R^q_+ : \sum_{j=1}^n \lambda_j X^t_j \le X^t,\ \sum_{j=1}^n \lambda_j Y^t_j \ge Y^t, \sum_{j=1}^n \lambda_j = 1, \lambda_j \ge 0\}$ and under CRS as $T^t_C = \{(X^t, Y^t) \in R^m_+ \times R^q_+ : \sum_{j=1}^n \lambda_j X^t_j \le X^t,\ \sum_{j=1}^n \lambda_j Y^t_j \ge Y^t,\ \lambda_j \ge 0\}$ following Charnes, Cooper and Rhodes (1978).

When market prices are not available, the most popular approach for evaluating productivity change in a set of DMUs over time is the Malmquist productivity index (Caves, Christensen & Diewert, 1982; Färe, Grosskopf, Lindgren & Roos, 1994). The Malmquist index is a ratio-based index that uses Shephard (1953) distance functions to represent the technology.

The output-oriented Malmquist productivity index for two periods of time t and $t+1$ under CRS technology is defined as:

$$M^0_{t,t+1} = \left[\frac{D^t_C\left(X^{t+1}_0, Y^{t+1}_0\right)}{D^t_C\left(X^t_0, Y^t_0\right)} \cdot \frac{D^{t+1}_C\left(X^{t+1}_0, Y^{t+1}_0\right)}{D^{t+1}_C\left(X^t_0, Y^t_0\right)} \right]^{1/2}, \tag{1}$$

where $D^s_C(X^h_0, Y^h_0) = \inf\{\theta : (X^h_0, \frac{Y^h_0}{\theta}) \in T^s_C\}$ is the Shephard output distance function calculated from the period h observation (X^h_0, Y^h_0), $h = t, t+1$, to the frontier of the technology at time s, $s = t, t+1$, and under CRS. In order to calculate Eq. (1) it is necessary to estimate its four constituent Shephard output distance functions. This can be achieved using DEA where, for example, the value of $D^t_C(X^t_j, Y^t_j) \le 1$ obtained is the radial technical efficiency score of DMU j with data and technology referred to period t. From a value of $D^t_C(X^t_j, Y^t_j) = 0.85$ it turns out that in period t the DMU j is able to expand the production of all its outputs by 15% without altering its inputs. This interpretation only holds when data and technology belong to the same period, otherwise the Shephard output distance functions could have efficiency values greater than one. The Malmquist index values in Eq. (1) are interpreted as total factor productivity changes. Values of the Malmquist index greater than one signal productivity improvements whilst values less than one point out productivity losses. For example, a Malmquist index value of 1.10 (0.95) means that productivity improved (declined) 10% (5%) between t and $t+1$.

Except in trivial cases, the assumption of CRS or VRS when calculating distance functions leads to different results. However, several voices have been raised in recent years calling attention to the fact that the Malmquist index expression does not adapt to the type of returns to scale that best fits the technology estimated by the data. Grifell-Tatjé and Lovell (1995) give a two-dimensional example to demonstrate that, in the presence of VRS, the Malmquist productivity index does not adequately measure productive change. In this same vein, Ray and Desli (1997) state that "the Malmquist productivity index is correctly measured by the ratio of CRS distance functions even when the technology exhibits variable returns to scale". In Balk (2001), on the other hand, the final result of constructing a productivity index based on the definition of its components separately is the Malmquist productivity index based on CRS distance functions, regardless of the type of returns to scale that the true technology exhibits in periods t and $t+1$. Lovell (2003) also emphasizes and supports this same idea. Therefore, we will assume CRS in the definition of (1).

2.1. The Camanho-Dyson Malmquist index

In 2006, Camanho and Dyson proposed an adaptation of the Malmquist index in order to compare the relative performance of two groups of DMUs, group A and group B, in a given period of time. The index described by Camanho and Dyson (2006) is an extension of the Malmquist-type indices used by Berg, Førsund and Jansen (1992, 1993) and Pastor, Perez and Quesada (1997) for comparing the performance of financial institutions in different countries. Ana Camanho and Robert Dyson developed a measure that can be used to compare the relative performance of two groups of DMUs operating under different technologies. Camanho and Dyson (2006, p.36) claim that the main advantage of their approach is to avoid mixing technologies, whereas other alternatives are based on pooling all units together to form a common merged frontier or meta-frontier.

The index introduced by Camanho and Dyson (2006) is an adaptation of the Malmquist index in such a way that the index does not evaluates the productivity change between two periods of time, but the transversal comparison of the relative average performance of groups of DMUs operating under different conditions or circumstances in a given period of time. Specifically, they assumed the observation of two groups of DMUs, group A and group B. In group A, $n^{t,A}$ DMUs produce q outputs, $Y^{t,A} \in R^q_+$, in period t through m inputs, $X^{t,A} \in R^m_+$, and, in B, $n^{t,B}$ DMUs produce q outputs, $Y^{t,B} \in R^q_+$, in period t through m inputs, $X^{t,B} \in R^m_+$. The DMUs operating in group A in period t are represented by their input-output vector as $(X^{t,A}_j, Y^{t,A}_j)$, $j = 1, \ldots, n^{t,A}$. In the same way, $(X^{t,B}_k, Y^{t,B}_k)$ denotes the input-output vector of DMU_k $k = 1, \ldots, n^{t,B}$, belonging to group B in period t. $D^{t,A}_C(X^{t,B}_k, Y^{t,B}_k) = \inf\{\theta : (X^{t,B}_k, \frac{Y^{t,B}_k}{\theta}) \in T^{t,A}_C\}$ is the Shephard output distance function calculated from observation $(X^{t,B}_k, Y^{t,B}_k)$ in group B in period t to the frontier of the technology of group A in period t,$T^{t,A}_C$. A similar notation is used for the distance for a unit in group A with respect to the technology of group B, and for the distance from a unit that belongs to the same group as the reference technology. Resorting to the Shephard distance function, Camanho and Dyson (2006) defined the CDMI as follows:

$$CDMI_t^{AB} = \left[\frac{\left(\prod_{j=1}^{n^{t,A}} D_C^{t,A}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}}{\left(\prod_{k=1}^{n^{t,B}} D_C^{t,A}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}} \cdot \frac{\left(\prod_{j=1}^{n^{t,A}} D_C^{t,B}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}}{\left(\prod_{k=1}^{n^{t,B}} D_C^{t,B}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}} \right]^{1/2}. \quad (2)$$

In terms of the interpretation of the $CDMI_t^{AB}$, a value greater (less) than unity indicates better productivity in group A (B) than in group B (A) while a value equal to one indicates that there is no productivity gap between both groups. For example, a $CDMI_t^{AB}$value of 1.05 (0.92) means that, on average, DMUs belonging to group A are 5% (8%) more (less) productive than DMUs in group B.

The CDMI can be decomposed into the following two terms:

$$CDMI_t^{AB} = \underbrace{\frac{\left(\prod_{j=1}^{n^{t,A}} D_C^{t,A}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}}{\left(\prod_{k=1}^{n^{t,B}} D_C^{t,B}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}}}_{EG_t^{AB}}$$

$$\cdot \underbrace{\left[\frac{\left(\prod_{j=1}^{n^{t,A}} D_C^{t,B}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}}{\left(\prod_{j=1}^{n^{t,A}} D_C^{t,A}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}} \cdot \frac{\left(\prod_{k=1}^{n^{t,B}} D_C^{t,B}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}}{\left(\prod_{k=1}^{n^{t,B}} D_C^{t,A}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}} \right]^{1/2}}_{TG_t^{AB}} \quad (3)$$

The term EG_t^{AB}measures the efficiency gap between both groups, whereas the term TG_t^{AB}evaluates the technology gap between the frontiers of the two analyzed groups, A and B, assuming that both frontiers have been estimated under constant returns to scale. The interpretation of the index values for EG_t^{AB} and TG_t^{AB} follows the same pattern as for the $CDMI_t^{AB}$; values greater (less) than one mean that group A is more (less) productive than group B in terms of efficiency or technology, respectively.

The CDMI has two features that require improvement. The first one is that the CDMI does not satisfy the property of circularity for more than two groups. This means that, for a fixed period of time s and three groups of DMUs, -A, B and C-, $CDMI_s^{AC} \neq CDMI_s^{AB} \cdot CDMI_s^{BC}$. The second one is that the index is only applicable to a single period of time.

Regarding these two issues, Aparicio et al. (2017) and Aparicio and Santín (2018) recently extended the CDMI. In particular, Aparicio and Santín (2018) introduced a base reference group technology (R) in a base period of time h to compare the relative performance between two or more groups of DMUs over time. This reference group solves the two disadvantages of the CDMI. We show this extension below.

Let us assume that we have observed $n^{s,A}$ DMUs in group A in period s, $s = t, t+1$, which produce output $Y^{s,A} \in R_+^q$ from input $X^{s,A} \in R_+^m$ and that we have also observed $n^{s,B}$ DMUs in group B in period s, $s = t, t+1$, which produce output $Y^{s,B} \in R_+^q$ from input $X^{s,B} \in R_+^m$. The DMUs operating in group A in period s are represented by their input-output vector as $(X_j^{s,A}, Y_j^{s,A}), j = 1, \ldots, n^{s,A}$. In the same way, $(X_k^{s,B}, Y_k^{s,B})$denotes the input-output vector of DMU_k, $k = 1, \ldots, n^{s,B}$, belonging to group B in period s.

With the aim of evaluating the relative performance between A and B over a period of time s, the base-group CDMI is defined with a base reference group (R) in the base time period h, as follows:

$$CDMI_s^{AB}(R^h) = \frac{\left(\prod_{j=1}^{n^{s,A}} D_C^{h,R}\left(X_j^{s,A}, Y_j^{s,A}\right) \right)^{1/n^{s,A}}}{\left(\prod_{k=1}^{n^{s,B}} D_C^{h,R}\left(X_k^{s,B}, Y_k^{s,B}\right) \right)^{1/n^{s,B}}}. \quad (4)$$

The base-group $CDMI_s^{AB}(R^h)$ measures the productivity gap between A and B where both groups are measured with respect to an arbitrary fixed base-group reference technology. Therefore, this base-group index in Eq. (4) is a ratio of the average technical efficiency measures of DMUs belonging to groups A (numerator) and B (denominator) with respect to the base-group technology. The values of this index are interpreted as in Eq. (2), however, the results do not depend on the production frontiers of groups A and B but rather on the base reference group technology R.

Likewise, $CDMI_s^{AB}(R^h)$ can be decomposed into the following two terms:

$$CDMI_s^{AB}(R^h) = \underbrace{\frac{\left(\prod_{j=1}^{n^{s,A}} D_C^{s,A}\left(X_j^{s,A}, Y_j^{s,A}\right) \right)^{1/n^{s,A}}}{\left(\prod_{k=1}^{n^{s,B}} D_C^{s,B}\left(X_k^{s,B}, Y_k^{s,B}\right) \right)^{1/n^{s,B}}}}_{EG_s^{AB}}$$

$$\cdot \underbrace{\left[\frac{\left(\prod_{j=1}^{n^{s,A}} D_C^{h,R}\left(X_j^{s,A}, Y_j^{s,A}\right) \right)^{1/n^{s,A}}}{\left(\prod_{j=1}^{n^{s,A}} D_C^{s,A}\left(X_j^{s,A}, Y_j^{s,A}\right) \right)^{1/n^{s,B}}} \cdot \frac{\left(\prod_{k=1}^{n^{s,B}} D_C^{s,B}\left(X_k^{s,B}, Y_k^{s,B}\right) \right)^{1/n^{s,B}}}{\left(\prod_{k=1}^{n^{s,B}} D_C^{h,R}\left(X_k^{s,B}, Y_k^{s,B}\right) \right)^{1/n^{s,B}}} \right]}_{TG_s^{AB}(R^h)}. \quad (5)$$

The ratio EG_s^{AB} compares within-group efficiency spreads, measuring the technical efficiency gap between both groups, while the ratio $TG_s^{AB}(R^h)$ evaluates the technical gap between the frontiers of the two analyzed groups, A and B, measured on the base reference technology (R) in the base time period h. We note that the EG_s^{AB}in Eq. (5) coincides with the same term in Eq. (3). The difference between both indexes evolves from the technical gap. While in Eq. (3) the technical gap is built directly by comparing technologies of groups A and B, this comparison is performed through the reference group in Eq. (5).

2.2. The base-group base-period pseudo-panel Malmquist index

Aparicio et al. (2017) extended the CDMI through the pseudo-panel Malmquist index (PPMI) for determining group performance evolution in panel or pseudo-panel databases. The PPMI had two important drawbacks. First, in general, the PPMI could not be interpreted as the ratio of aggregated productivity changes of two groups of DMUs over time. Second, the PPMI is the ratio of two Malmquist-type indices inheriting the non-circularity of the original CDMI.

Regarding the analysis of the evolution of the gap of relative performance between two groups over several periods of time and in order to overcome both problems, Aparicio and Santín (2018) proposed the base-group base-period pseudo-panel Malmquist index PPMI(R):

$$PPMI_{t,t+1}^{AB}(R^h) = \frac{CDMI_{t+1}^{AB}(R^h)}{CDMI_t^{AB}(R^h)}. \quad (6)$$

They showed that the $PPMI_{t,t+1}^{AB}(R^h)$ matches the ratio between the change in aggregate productivity growth of the units of group A from t to $t+1$, and the change in aggregate productivity growth of the units of group B from t to $t+1$, measured on the technology of the reference group R observed in the period of time h.

$$PPMI_{t,t+1}^{AB}(R^h) = \left[\frac{\left(\prod_{j=1}^{n^{t+1,A}} D_C^{h,R}\left(X_j^{t+1,A}, Y_j^{t+1,A}\right) \right)^{1/n^{t+1,A}}}{\left(\prod_{j=1}^{n^{t,A}} D_C^{h,R}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}} \right] \Bigg/ \left[\frac{\left(\prod_{k=1}^{n^{t+1,B}} D_C^{h,R}\left(X_k^{t+1,B}, Y_k^{t+1,B}\right) \right)^{1/n^{t+1,B}}}{\left(\prod_{k=1}^{n^{t,B}} D_C^{h,R}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}} \right]. \tag{7}$$

Following Aparicio et al. (2017), in order to interpret $PPMI_{t,t+1}^{AB}(R^h)$ in a suitable way, it is also necessary to analyze the $CDMI_t^{AB}(R^h)$and the$CDMI_{t+1}^{AB}(R^h)$values in four possible settings:

Setting 1: $CDMI_t^{AB}(R^h) < 1$,$CDMI_{t+1}^{AB}(R^h) < 1$. This means that, on average, group B had a better relative performance than group A both in t and $t+1$. As a result, if $PPMI_{t,t+1}^{AB}(R^h) < 1$ the relative performance gap was opened up by B over A. In fact, $100 \times (1 - PPMI_{t,t+1}^{AB}(R^h))$ indicates the percentage in which the relative performance of A compared to B has worsened. If $PPMI_{t,t+1}^{AB}(R^h) > 1$ then group A is catching up on group B. In this case, $100 \times (PPMI_{t,t+1}^{AB}(R^h) - 1)$indicates the percentage in which the relative performance of A compared to B has improved.

Setting 2: $CDMI_t^{AB}(R^h) > 1$,$CDMI_{t+1}^{AB}(R^h) > 1$. This means that group A had a better relative performance than group B both in t and $t+1$. As a result, when $PPMI_{t,t+1}^{AB}(R^h) < 1$ then group A is catching up on group B. In fact, $100 \times (1 - PPMI_{t,t+1}^{AB}(R^h))$ indicates the percentage in which the relative performance of A compared to B has decreased. If $PPMI_{t,t+1}^{AB}(R^h) > 1$ the group A has opened up its performance gap with respect to the group B. In this case, $100 \times (PPMI_{t,t+1}^{AB}(R^h) - 1)$indicates the percentage in which the relative performance of A compared to B has increased.

Setting 3: $CDMI_t^{AB}(R^h) > 1$ and $CDMI_{t+1}^{AB}(R^h) < 1$. Under this scenario, group A had a better relative performance than group B in period t but the reverse happened in the second period, $t+1$. In this case, $PPMI_{t,t+1}^{AB}(R^h) < 1$ indicates that the performance of group A worsened drastically from period t to period $t+1$. The value of $100 \times (1 - PPMI_{t,t+1}^{AB}(R^h))$% measures how much the relative performance gap between the two analyzed groups of units reduced over time.

Setting 4: $CDMI_t^{AB}(R^h) < 1$ and $CDMI_{t+1}^{AB}(R^h) > 1$. In this case, B had a better relative performance than A in period t but the relation between both groups changed drastically in period $t+1$, where A had a better relative performance than B. In this case $PPMI_{t,t+1}^{AB}(R^h) > 1$ always, meaning that the status of group A improved from period t to period $t+1$ and the value of $100 \times (PPMI_{t,t+1}^{AB}(R^h) - 1)$% signals how much the relative performance gap between the two analyzed groups of units increased over time.

For the sake of simplicity, we did not interpret settings where the indices are equal to one. However, its interpretation is straightforward. When $CDMI_t^{AB}(R^h) = 1$ no group stands out over the other in the considered period. A $PPMI_{t,t+1}^{AB}(R^h) = 1$ would imply that the initial productivity gap in t holds in $t+1$.

The $PPMI_{t,t+1}^{AB}(R^h)$ can also be decomposed into efficiency gap change (EGC) and technology gap change (TGC) as follows:

$$PPMI_{t,t+1}^{AB}(R^h) = EGC_{t,t+1}^{AB} \cdot TGC_{t,t+1}^{AB}(R^h), \tag{8}$$

where

$$EGC_{t,t+1}^{AB} = \frac{EG_{t+1}^{AB}}{EG_t^{AB}} = \left[\frac{\left(\prod_{j=1}^{n^{t+1,A}} D_C^{t+1,A}\left(X_j^{t+1,A}, Y_j^{t+1,A}\right) \right)^{1/n^{t+1,A}}}{\left(\prod_{j=1}^{n^{t,A}} D_C^{t,A}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}} \right] \Bigg/ \left[\frac{\left(\prod_{k=1}^{n^{t+1,B}} D_C^{t+1,B}\left(X_k^{t+1,B}, Y_k^{t+1,B}\right) \right)^{1/n^{t+1,B}}}{\left(\prod_{k=1}^{n^{t,B}} D_C^{t,B}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}} \right] \tag{9}$$

$$\begin{aligned} TGC_{t,t+1}^{AB}(R^h) &= \frac{TG_{t+1}^{AB}(R^h)}{TG_t^{AB}(R^h)} \\ &= \frac{\left[\frac{\left(\prod_{j=1}^{n^{t+1,A}} D_C^{h,R}\left(X_j^{t+1,A}, Y_j^{t+1,A}\right) \right)^{1/n^{t+1,A}}}{\left(\prod_{j=1}^{n^{t+1,A}} D_C^{t+1,A}\left(X_j^{t+1,A}, Y_j^{t+1,A}\right) \right)^{1/n^{t+1,A}}} \cdot \frac{\left(\prod_{k=1}^{n^{t+1,B}} D_C^{t+1,B}\left(X_k^{t+1,B}, Y_k^{t+1,B}\right) \right)^{1/n^{t+1,B}}}{\left(\prod_{k=1}^{n^{t+1,B}} D_C^{h,R}\left(X_k^{t+1,B}, Y_k^{t+1,B}\right) \right)^{1/n^{t+1,B}}} \right]}{\left[\frac{\left(\prod_{j=1}^{n^{t,A}} D_C^{h,R}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}}{\left(\prod_{j=1}^{n^{t,A}} D_C^{t,A}\left(X_j^{t,A}, Y_j^{t,A}\right) \right)^{1/n^{t,A}}} \cdot \frac{\left(\prod_{k=1}^{n^{t,B}} D_C^{t,B}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}}{\left(\prod_{k=1}^{n^{t,B}} D_C^{h,R}\left(X_k^{t,B}, Y_k^{t,B}\right) \right)^{1/n^{t,B}}} \right]} = \\ &= \frac{\left[\left(\prod_{j=1}^{n^{t+1,A}} \frac{D_C^{h,R}\left(X_j^{t+1,A}, Y_j^{t+1,A}\right)}{D_C^{t+1,A}\left(X_j^{t+1,A}, Y_j^{t+1,A}\right)} \right)^{1/n^{t+1,A}} \cdot \left(\prod_{j=1}^{n^{t,A}} \frac{D_C^{t,A}\left(X_j^{t,A}, Y_j^{t,A}\right)}{D_C^{h,R}\left(X_j^{t,A}, Y_j^{t,A}\right)} \right)^{1/n^{t,A}} \right]}{\left[\left(\prod_{k=1}^{n^{t+1,B}} \frac{D_C^{h,R}\left(X_k^{t+1,B}, Y_k^{t+1,B}\right)}{D_C^{t+1,B}\left(X_k^{t+1,B}, Y_k^{t+1,B}\right)} \right)^{1/n^{t+1,B}} \cdot \left(\prod_{k=1}^{n^{t,B}} \frac{D_C^{t,B}\left(X_k^{t,B}, Y_k^{t,B}\right)}{D_C^{h,R}\left(X_k^{t,B}, Y_k^{t,B}\right)} \right)^{1/n^{t,B}} \right]} \end{aligned} \tag{10}$$

For interpreting $EGC_{t,t+1}^{AB}$ and $TGC_{t,t+1}^{AB}(R^h)$ we proceed by following the same logic as above for the $PPMI_{t,t+1}^{AB}(R^h)$ measure. Likewise, Aparicio and Santín (2018) demonstrated that the $PPMI_{t,t+1}^{AB}(R^h)$ and their components meet the property of circularity when more than two groups are evaluated, that is,

$$PPMI_{t,t+1}^{AB}(R^h) \cdot PPMI_{t,t+1}^{BC}(R^h) = PPMI_{t,t+1}^{AC}(R^h). \tag{11}$$

2.3. The Luenberger productivity indicator

One of the objectives in this paper is to define a new type of CD index based on the Luenberger indicator. Therefore, we will briefly introduce the concepts of the directional distance function (DDF) and the Luenberger productivity indicator below.

Chambers, Färe and Grosskopf (1996b) and Chambers and Pope (1996) defined the Luenberger productivity indicator as an alternative to the standard Malmquist productivity index. The Luenberger productivity indicator is a difference-based measure of directional distance functions (DDFs) (see Chambers, Chung & Färe, 1996a, 1998) and, therefore, is additive. The DDF is in fact, closely related to the shortage function introduced by Luenberger (1992a) using a netput formulation.[3]

[3] Luenberger (1992a) introduced the concept of benefit function as a representation of the amount that an individual is willing to trade, in terms of a specific reference commodity bundle g, for the opportunity to move from a consumption bundle to a utility threshold. Luenberger also defined a so-called shortage function (1992a, p. 242, Definition 4.1), which basically measures the amount by which a specific plan is short of reaching the frontier of the technology. Later, Chambers et al., 1996a, 1998) redefined the benefit function and the shortage function as the directional distance function.

The DDF for time t can be implemented in DEA under VRS as[4]:

$$\overrightarrow{D}_V^t(X_0^t, Y_0^t; g) = \max \quad \beta^t$$
$$\begin{aligned} s.t. \quad & \textstyle\sum_{j=1}^n \lambda_j X_{ij}^t \le X_{i0}^t - \beta^t g_i^I, & i = 1, \dots, m \\ & -\textstyle\sum_{j=1}^n \lambda_j Y_{rj}^t \le -Y_{r0}^t - \beta^t g_r^O, & r = 1, \dots, q \\ & \textstyle\sum_{j=1}^n \lambda_j = 1, & \\ & \lambda_j \ge 0, & j = 1, \dots, n \end{aligned} \tag{12}$$

where $g = (g^I, g^O) \in R_+^m \times R_+^q$ is a directional vector for inputs and outputs, respectively, and β^t measures the degree of technical inefficiency.[5] Directional distance function projects any input and output vector onto the technology frontier in a pre-assigned direction given by the directional vector.

One of the advantages of using DDF over the Shephard distance function is that DDF is an unoriented (graph) measure. It is, therefore, a more flexible function since inputs can be reduced and outputs increased, simultaneously (Chambers, Chung & Färe, 1998). Therefore, contrary to Shephard distance functions, an unnecessary equiproportional output or input change does not have to be imposed.

Another advantage of using DDF and its Luenberger indicator application over the Malmquist index is that it can assume VRS in the calculation of distances (Barros, Peypoch & Solonandrasana, 2009; Briec & Kerstens, 2009a; Juo, Fu, Yu & Lin, 2015; Nakano & Managi, 2008). This is discouraged on theoretical and practical grounds when the Malmquist index is used (Grifell-Tatjé & Lovell, 1995; Ray & Desli;, 1997). This will be a very important point in our approach, since we will be able to compare the technologies and performances of two groups, A and B, assuming VRS instead of CRS on a mandatory basis, as is the case if the Malmquist index and the definition of the original CDMI are used. This means that we can compare the best-practice frontier of two technologies linked to VRS instead of contrasting the conical frontiers associated with CRS.[6]

Following Luenberger (1992a), the DDF subtracts $\beta^t g_i^I$units from the observed inputs and adds $\beta^t g_r^O$units to the observed outputs of the evaluated unit in order to reach the frontier technology. If the DMU is fully efficient, then the value of the DDF will be zero, otherwise it would be positive.

Based on the DDF, it is possible to define a measure of productivity change over time through the Luenberger productivity indicator as follows (Chambers et al., 1996b, and Chambers & Pope, 1996):

$$\begin{aligned} L_{t,t+1}^0 = \frac{1}{2}\Big[& \left(\vec{D}_V^{t+1}\left(X_0^t, Y_0^t; g\right) - \vec{D}_V^{t+1}\left(X_0^{t+1}, Y_0^{t+1}; g\right)\right) \\ & + \left(\vec{D}_V^{t}\left(X_0^t, Y_0^t; g\right) - \vec{D}_V^{t}\left(X_0^{t+1}, Y_0^{t+1}; g\right)\right)\Big] \end{aligned} \tag{13}$$

Additionally, the traditional Luenberger indicator is decomposed into efficiency change (ECH) and frontier shift (TCH) as follows:

$$\begin{aligned} ECH &= \vec{D}_V^{t}\left(X_0^t, Y_0^t; g\right) - \vec{D}_V^{t+1}\left(X_0^{t+1}, Y_0^{t+1}; g\right),\ TCH \\ &= \frac{1}{2}\Big[\left(\vec{D}_V^{t+1}\left(X_0^{t+1}, Y_0^{t+1}; g\right) - \vec{D}_V^{t}\left(X_0^{t+1}, Y_0^{t+1}; g\right)\right) \\ &\quad + \left(\vec{D}_V^{t+1}\left(X_0^t, Y_0^t; g\right) - \vec{D}_V^{t}\left(X_0^t, Y_0^t; g\right)\right)\Big] \end{aligned} \tag{14}$$

3. New Camanho-Dyson and pseudo-panel Luenberger indicators

In this section, we present a new CD-type index based on the Luenberger indicator (CDLI) and the DDF to compare the relative performance between two or more groups of DMUs.

Let $\vec{D}_V^{s,A}(X_k^{h,B}, Y_k^{h,B}; g)$ represent the DDF calculated for the point $(X_k^{h,B}, Y_k^{h,B})$, $h = t, t+1$, to the frontier of the technology of group A in period s, $s = t, t+1$, under VRS, where $g = (g^I, g^O) \in R_+^m \times R_+^q$ is a directional vector (Chambers et al., 1996a).

Following Aparicio et al. (2017) and Aparicio and Santín (2018); Camanho and Dyson (2006) and based on the Luenberger indicator shown in Section 2, we define the $CDLI_s^{AB}(R^h)$ as a joint measure for the relative comparison of the performance between two groups of production units, A and B, with a base reference group R fixed in a base period of time h as follows:

$$CDLI_s^{AB}\left(R^h\right) = \frac{1}{n^{s,B}} \sum_{k=1}^{n^{s,B}} \vec{D}_V^{h,R}\left(X_k^{s,B}, Y_k^{s,B}; g\right) - \frac{1}{n^{s,A}} \sum_{j=1}^{n^{s,A}} \vec{D}_V^{h,R}\left(X_j^{s,A}, Y_j^{s,A}; g\right). \tag{15}$$

A positive indicator value points out that group A performed relatively better on average than group B, and a negative value indicates a better relative performance on average for group B over group A. According to Färe et al. (2008), the choice of the direction vector is an empirical issue. These authors argue that some potential specifications for the directional vector $g = (g^I, g^O)$ include a self-direction based on the observed data $g = (g^I, g^O) = (X, Y)$; or alternatively either an output direction $g = (g^I, g^O) = (0, Y)$ or an input direction $g = (g^I, g^O) = (X, 0)$ based on the observed data; mean values of all inputs and outputs data $g = (g^I, g^O) = (\bar{X}, \bar{Y})$ or an optimal direction to minimize the distance to the production frontier.

For benchmarking groups of DMUs through the $CDLI_s^{AB}(R^h)$, we suggest choosing a constant common direction vector g for evaluating all observations over time. This choice has three distinct advantages. First, assuming the same direction for all DMUs is akin to an "egalitarian" evaluation (Färe et al., 2008). Second, the constant direction vector facilitates aggregation for obtaining a meaningful average measure for comparing groups of DMUs (Luenberger, 1992b). Third, in this case the $CDLI_s^{AB}(R^h)$ can be interpreted in terms of the physical reduction (increase) of units of inputs (outputs) of one of the groups, say A, to reach the average performance of the other group, say B, and both in comparison with the reference technology R^h. We will go on to discuss this interpretation in the empirical application.

[4] A similar optimization model is used for determining $\overrightarrow{D}_V^s(X_0^t, Y_0^t; g)$ when $s \neq t$. However, in that case, the input-output vectors that appear on the left-hand side of the constraints are observed in period s, while the assessed input-output vector is observed in a different period, t. It is recognized in the literature that 'mixed period' directional distance functions can yield infeasible and unbounded results (see Briec and Kerstens 2009a, 2009b). Indeed, Briec and Kerstens, 2009a, 2009b) showed that this weakness can also occur even in single period (contemporaneous) calculations when the output directional vector is nonzero and the number of inputs is larger than or equal to two, or the directional input vector is non-full dimensional whenever the output direction is null. In addition, Briec and Kerstens (2009a) noticed that the computation of mixed-period DDF can lead to projections with a negative output, which in general have little meaning in standard economic production applications. In order to avoid these problems, one needs to add an additional constraint into program (12): $Y_{r0}^t + \beta^t g_r^O \ge 0$, $r = 1, \dots, q$. In the empirical study carried out in this paper, we did not find any of the aforementioned problems, but researchers must always pay attention to these drawbacks of the DDF in practice.

[5] $D(X_0, Y_0)$ denotes the Shephard distance function calculated for the point (X_0, Y_0), while $\overrightarrow{D}(X_0, Y_0; g)$ denotes the directional distance function calculated for the point (X_0, Y_0) when the directional vector g is used. Note that we use the symbol D for denoting both distances, although, as usual, the directional distance function also includes an arrow symbol. We will use this notation throughout the paper.

[6] As a reviewer pointed out, when the direction is preassigned, the DDF is homogenous of degree one in the case of assuming CRS. It could be problematic under the CRS assumption. Fortunately, in this paper, VRS models are always considered. However, even in such a case, the efficiency and productivity scores might be affected. The in-depth study of these effects is beyond the scope of this paper.

It is easily derived that $CDLI_s^{AB}(R^h)$ can, like the CDMI in (3), be decomposed into two subcomponents $EG_s^{AB}$and $TG_s^{AB}(R^h)$ as follows:

$$CDLI_s^{AB}(R^h) = \underbrace{\frac{1}{n^{s,B}}\sum_{k=1}^{n^{s,B}}\vec{D}_V^{s,B}\left(X_k^{s,B},Y_k^{s,B};g\right) - \frac{1}{n^{s,A}}\sum_{j=1}^{n^{s,A}}\vec{D}_V^{s,A}\left(X_j^{s,A},Y_j^{s,A};g\right)}_{EG_s^{AB}}$$
$$+\underbrace{\left[\frac{1}{n^{s,B}}\sum_{k=1}^{n^{s,B}}\vec{D}_V^{h,R}\left(X_k^{s,B},Y_k^{s,B};g\right) - \frac{1}{n^{s,B}}\sum_{k=1}^{n^{s,B}}\vec{D}_V^{s,B}\left(X_k^{s,B},Y_k^{s,B};g\right) + \frac{1}{n^{s,A}}\sum_{j=1}^{n^{s,A}}\vec{D}_V^{s,A}\left(X_j^{s,A},Y_j^{s,A};g\right) - \frac{1}{n^{s,A}}\sum_{j=1}^{n^{s,A}}\vec{D}_V^{h,R}\left(X_j^{s,A},Y_j^{s,A};g\right)\right]}_{TG_s^{AB}(R^h)} \tag{16}$$

The term EG_s^{AB} measures the efficiency gap between both groups in the time period s, whereas the term $TG_s^{AB}(R^h)$ evaluates the productivity gap between the frontiers of the two analyzed groups, A and B, in the time period s, measured on the base reference technology (R) in the base time period h and assuming that both frontiers have been estimated under variable returns to scale.

To determine the proposed index, $CDLI_s^{AB}(R^h)$, we must calculate a series of DDFs, such as $\vec{D}_V^{s,A}(X_j^{s,A},Y_j^{s,A};g)$, $\vec{D}_V^{s,B}(X_k^{s,B},Y_k^{s,B};g)$, $\vec{D}_V^{h,R}(X_j^{s,A},Y_j^{s,A};g)$ and $\vec{D}_V^{h,R}(X_k^{s,B},Y_k^{s,B};g)$, for all $j=1,\dots,n^A$, $k=1,\dots,n^B$. Note that, unlike the CDMI index, they are all calculated assuming VRS.

By analogy with the PPMI shown in Section 2, we suggest that, to evaluate the performance of two or more groups over time using pseudo-panels, this new base-group base-period PPLI should be used to measure the relative performance gap change between groups A and B from t to $t+1$ if a non-equiproportional graph efficiency measure has to be used:

$$PPLI_{t,t+1}^{AB}(R^h) = CDLI_t^{AB}(R^h) - CDLI_{t+1}^{AB}(R^h). \tag{17}$$

The straightforward interpretation is that a negative value of $PPLI_{t,t+1}^{AB}(R^h)$ means that the performance of group A has improved with respect to group B from t to $t+1$, while a positive value means the opposite. To further interpret the value of $PPLI_{t,t+1}^{AB}(R^h)$ in order to find out if the productivity gap between the two groups is opening or closing, we have to adapt the case-based reasoning defined by Aparicio et al. (2017) to the case of our index. Therefore, we have to analyze the value of $CDLI_t^{AB}(R^h)$ and $CDLI_{t+1}^{AB}(R^h)$ in the following cases:

Setting 1: $CDLI_t^{AB}(R^h), CDLI_{t+1}^{AB}(R^h) > 0$. On average group A had a better relative performance than group B in both periods from t to $t+1$. Regarding the value of PPLI, there are two possibilities:

1a) $PPLI_{t,t+1}^{AB}(R^h) > 0$, which means that the relative performance gap has narrowed, and group B has improved with respect to group A from t to $t+1$.

1b) $PPLI_{t,t+1}^{AB}(R^h) < 0$, which means that the relative performance gap has narrowed, and group B has worsened with respect to group A from t to $t+1$.

Setting 2: $CDLI_t^{AB}(R^h), CDLI_{t+1}^{AB}(R^h) < 0$. On average, relative performance was better for group B than for group A in both periods from t to $t+1$. Regarding the value of PPLI, there are two possibilities:

2a) $PPLI_{t,t+1}^{AB}(R^h) > 0$, which means that the relative performance gap has widened, and group A has worsened with respect to group B from t to $t+1$.

2b) $PPLI_{t,t+1}^{AB}(R^h) < 0$, which means that the relative performance gap has narrowed, and, therefore, group A has improved with respect to group B from t to $t+1$.

Setting 3: $CDLI_t^{AB}(R^h) > 0$ and $CDLI_{t+1}^{AB}(R^h) < 0$. In this scenario, relative performance was better for group A than for group B in period t but better for group B than for group A in period $t+1$. For the value of PPLI, we have only one possibility: $PPLI_{t,t+1}^{AB}(R^h) > 0$. In this case, the performance of group A dropped with respect to group B from t to $t+1$.

Setting 4: $CDLI_t^{AB}(R^h) < 0$ and $CDLI_{t+1}^{AB}(R^h) > 0$. In this scenario, relative performance was better for group B than group A in period t, but better for group A than group B in period $t+1$. For the value of PPLI, we have only one possibility: $PPLI_{t,t+1}^{AB}(R^h) < 0$. In this case, group A improved with respect to group B from t to $t+1$.

As for the components of $PPLI_{t,t+1}^{AB}(R^h)$, this index can be decomposed into EGC and TGC, as shown in (18):

$$\begin{aligned} PPLI_{t,t+1}^{AB}(R^h) &= CDLI_t^{AB}(R^h) - CDLI_{t+1}^{AB}(R^h) \\ &= \left(EG_t^{AB} + TG_t^{AB}(R^h)\right) - \left(EG_{t+1}^{AB} + TG_{t+1}^{AB}(R^h)\right) \\ &= EGC_{t,t+1}^{AB} + TGC_{t,t+1}^{AB}(R^h) \end{aligned} \tag{18}$$

$$\begin{aligned} &EGC_{t,t+1}^{AB} \\ &= EG_t^{AB} - EG_{t+1}^{AB} \\ &= \left(\frac{1}{n^{t,B}}\sum_{k=1}^{n^{t,B}}\vec{D}_V^{t,B}\left(X_k^{t,B},Y_k^{t,B};g\right) - \frac{1}{n^{t+1,B}}\sum_{k=1}^{n^{t+1,B}}\vec{D}_V^{t+1,B}\left(X_k^{t+1,B},Y_k^{t+1,B};g\right)\right) \\ &-\left(\frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\vec{D}_V^{t,A}\left(X_j^{t,A},Y_j^{t,A};g\right) - \frac{1}{n^{t+1,A}}\sum_{j=1}^{n^{t+1,A}}\vec{D}_V^{t+1,A}\left(X_j^{t+1,A},Y_j^{t+1,A};g\right)\right) \end{aligned} \tag{19}$$

As for $TGC_{t,t+1}^{AB}(R^h)$, we have that:

$$\begin{aligned} &TGC_{t,t+1}^{AB}(R^h) = TG_t^{AB}(R^h) - TG_{t+1}^{AB}(R^h) \\ &= \left[\frac{1}{n^{t,B}}\sum_{k=1}^{n^{t,B}}\vec{D}_V^{h,R}\left(X_k^{t,B},Y_k^{t,B};g\right) - \frac{1}{n^{t,B}}\sum_{k=1}^{n^{t,B}}\vec{D}_V^{t,B}\left(X_k^{t,B},Y_k^{t,B};g\right)\right. \\ &\left.+\frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\vec{D}_V^{t,A}\left(X_j^{t,A},Y_j^{t,A};g\right) - \frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\vec{D}_V^{h,R}\left(X_j^{t,A},Y_j^{t,A};g\right)\right] \\ &-\left[\frac{1}{n^{t+1,B}}\sum_{k=1}^{n^{t+1,B}}\vec{D}_V^{h,R}\left(X_k^{t+1,B},Y_k^{t+1,B};g\right) - \frac{1}{n^{t+1,B}}\sum_{k=1}^{n^{t+1,B}}\vec{D}_V^{t+1,B}\left(X_k^{t+1,B},Y_k^{t+1,B};g\right)\right. \\ &\left.+\frac{1}{n^{t+1,A}}\sum_{j=1}^{n^{t+1,A}}\vec{D}_V^{t+1,A}\left(X_j^{t+1,A},Y_j^{t+1,A};g\right) - \frac{1}{n^{t+1,A}}\sum_{j=1}^{n^{t+1,A}}\vec{D}_V^{h,R}\left(X_j^{t+1,A},Y_j^{t+1,A};g\right)\right] \\ &= \left[\frac{1}{n^{t,B}}\sum_{k=1}^{n^{t,B}}\left(\vec{D}_V^{h,R}\left(X_k^{t,B},Y_k^{t,B};g\right) - \vec{D}_V^{t,B}\left(X_k^{t,B},Y_k^{t,B};g\right)\right)\right. \\ &\left.+\frac{1}{n^{t+1,B}}\sum_{k=1}^{n^{t+1,B}}\left(\vec{D}_V^{t+1,B}\left(X_k^{t+1,B},Y_k^{t+1,B};g\right) - \vec{D}_V^{h,R}\left(X_k^{t+1,B},Y_k^{t+1,B};g\right)\right)\right] \\ &-\left[\frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\left(\vec{D}_V^{h,R}\left(X_j^{t,A},Y_j^{t,A};g\right) - \vec{D}_V^{t,A}\left(X_j^{t,A},Y_j^{t,A};g\right)\right)\right. \\ &\left.+\frac{1}{n^{t+1,A}}\sum_{j=1}^{n^{t+1,A}}\left(\vec{D}_V^{t+1,A}\left(X_j^{t+1,A},Y_j^{t+1,A};g\right) - \vec{D}_V^{h,R}\left(X_j^{t+1,A},Y_j^{t+1,A};g\right)\right)\right] \end{aligned} \tag{20}$$

Next, we show that CDLI with a base reference group R in a base period of time h fulfills the property of additive circularity when comparing more than two groups, namely, group A, group B and group C:

$$\begin{aligned} &CDLI_s^{AB}(R^s) + CDLI_s^{BC}(R^s) \\ &= \frac{1}{n^{s,B}}\sum_{k=1}^{n^{s,B}}\vec{D}_V^{s,R}\left(X_k^{s,B},Y_k^{s,B};g\right) - \frac{1}{n^{s,A}}\sum_{j=1}^{n^{s,A}}\vec{D}_V^{s,R}\left(X_j^{s,A},Y_j^{s,A};g\right) \end{aligned}$$

Table 1
School sample and percentage of fully private schools, PFPS and PS.

Country	PISA wave	N	N valid	Fully Private*	PFPS	PS
Belgium	2009	278	257	0.78%	64.59%	34.63%
	2015	288	87	0.00%	54.02%	45.98%
Czech	2009	261	247	0.00%	5.26%	94.74%
Republic	2015	344	334	0.60%	8.08%	91.32%
Denmark	2009	285	274	1.82%	13.87%	84.31%
	2015	333	255	1.96%	15.29%	82.75%
Hungary	2009	187	184	0.54%	10.87%	88.59%
	2015	245	230	1.74%	15.65%	82.61%
Ireland	2009	144	128	7.03%	48.44%	44.53%
	2015	167	157	2.55%	50.32%	47.13%
Netherlands	2009	186	173	0.00%	60.12%	39.88%
	2015	187	109	0.92%	55.96%	43.12%
Slovakia	2009	189	189	0.00%	8.99%	91.01%
	2015	290	289	0.00%	11.76%	88.24%
Spain	2009	889	836	3.59%	35.17%	61.24%
	2015	201	194	5.67%	27.84%	66.49%

N: Number of schools in PISA; N valid: Number of schools with information about their ownership. PFPS: Publicly Funded Private Schools. PS: Public Schools.
* In some countries there are also fully private schools, which receive the total or almost the total of their core funding from student fees. In this paper, we focus our attention on the PFPS.

$$+\frac{1}{n^{s,C}}\sum_{l=1}^{n^{s,C}}\vec{D}_V^{s,R}\left(X_l^{s,C},Y_l^{s,C};g\right)-\frac{1}{n^{s,B}}\sum_{k=1}^{n^{s,B}}\vec{D}_V^{s,R}\left(X_k^{s,B},Y_k^{s,B};g\right)$$
$$=\frac{1}{n^{s,C}}\sum_{l=1}^{n^{s,C}}\vec{D}_V^{s,R}\left(X_l^{s,C},Y_l^{s,C};g\right)-\frac{1}{n^{s,A}}\sum_{j=1}^{n^{s,A}}\vec{D}_V^{s,R}\left(X_j^{s,A},Y_j^{s,A};g\right)$$
$$=CDLI_s^{AC}(R^s) \tag{21}$$

Consequently, base-group base-period PPLI also fulfills the property of additive circularity:

$$\begin{aligned}
&PPLI_{t,t+1}^{AB}(R^h)+PPLI_{t,t+1}^{BC}(R^h)=\\
&=\left(CDLI_t^{AB}(R^h)-CDLI_{t+1}^{AB}(R^h)\right)+\left(CDLI_t^{BC}(R^h)-CDLI_{t+1}^{BC}(R^h)\right)\\
&=\left(CDLI_t^{AB}(R^h)+CDLI_t^{BC}(R^h)\right)-\left(CDLI_{t+1}^{AB}(R^h)+CDLI_{t+1}^{BC}(R^h)\right)\\
&=CDLI_t^{AC}(R^h)-CDLI_{t+1}^{AC}(R^h)=PPLI_{t,t+1}^{AC}(R^h)
\end{aligned} \tag{22}$$

The same procedure can be used to prove that the efficiency gap term and the technology gap change component also satisfy circularity.

On the other hand, the change in productivity of the same group can be assessed for two periods of time, t and $t+1$, through:

$$\begin{aligned}
CDLI_{t,t+1}^{A}(R^h)=&\frac{1}{n^{t+1,A}}\sum_{k=1}^{n^{t+1,A}}\vec{D}_V^{h,R}\left(X_k^{t+1,A},Y_k^{t+1,A};g\right)\\
&-\frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\vec{D}_V^{h,R}\left(X_j^{t,A},Y_j^{t,A};g\right).
\end{aligned} \tag{23}$$

A positive indicator value points out that the relative performance for group A was better on average in period t than in period $t+1$, and a negative value indicates that the relative performance was better on average in period $t+1$ than in period t.

In the same vein, the above expression can be decomposed into two subcomponents:

$$\begin{aligned}
CDLI_{t,t+1}^{A}(R^h)=&\underbrace{\frac{1}{n^{t+1,A}}\sum_{k=1}^{n^{t+1,A}}\vec{D}_V^{t+1,A}\left(X_k^{t+1,A},Y_k^{t+1,A};g\right)-\frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\vec{D}_V^{t,A}\left(X_j^{t,A},Y_j^{t,A};g\right)}_{EG_{t,t+1}^{A}}\\
&\underbrace{\begin{aligned}+\Big[&\tfrac{1}{n^{t+1,A}}\textstyle\sum_{k=1}^{n^{t+1,A}}\vec{D}_V^{h,R}\left(X_k^{t+1,A},Y_k^{t+1,A};g\right)-\tfrac{1}{n^{t+1,A}}\textstyle\sum_{k=1}^{n^{t+1,A}}\vec{D}_V^{t+1,A}\left(X_k^{t+1,A},Y_k^{t+1,A};g\right)\\
&+\frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\vec{D}_V^{t,A}\left(X_j^{t,A},Y_j^{t,A};g\right)-\frac{1}{n^{t,A}}\sum_{j=1}^{n^{t,A}}\vec{D}_V^{h,R}\left(X_j^{t,A},Y_j^{t,A};g\right)\Big]\end{aligned}}_{TG_{t,t+1}^{A}(R^h)}
\end{aligned} \tag{24}$$

Note, that the $CDLI_{t,t+1}^{A}(R^h)$ (one group and two periods) is halfway between the CDLI (two groups and one period) and the PPLI that needs to deal with four sets of units (two groups and two periods). However, regarding the self-changes in productivities of two groups over two periods, say $CDLI_{t,t+1}^{A}(R^h)$ and $CDLI_{t,t+1}^{B}(R^h)$, there is an interesting equivalence between these changes and the $PPLI_{t,t+1}^{AB}(R^h)$ defined in Eqs. (17) and (18), thanks to the circularity property that can be expressed as follows:

$$PPLI_{t,t+1}^{AB}(R^h)=CDLI_{t,t+1}^{A}(R^h)-CDLI_{t,t+1}^{B}(R^h)=CDLI_t^{AB}(R^h)-CDLI_{t+1}^{AB}(R^h). \tag{25}$$

This means that another way of looking at $PPLI_{t,t+1}^{AB}(R^h)$is to interpret the index as the difference of the productivity changes of two groups, A and B, over two periods, t and $t+1$, with respect to a reference technology.

4. Empirical application. sample, data description and the ‘reference’ choice

This section includes an empirical illustration of the use of the new methodology proposed in this paper. The CDLI and PPLI are applied to a set of European Union countries in the education sector in order to compare the performance of PS and PFPS over time.

4.1. Data and variables

To compare the productivity gap between PS and PFPS, we selected information from eight European Union countries participating in the PISA 2009 and 2015 waves: Belgium, Czech Republic, Denmark, Hungary, Ireland, Netherlands, Slovakia and Spain. On one hand, we selected PISA because it contains information about school ownership, student background, school resources and academic achievement. We use student data from the 2009 and 2015 waves aggregated at school level. On the other hand, we only include countries in which there is at least a representative five percent of PFPS in both waves.[7] Table 1 shows the proportion of high schools by declared ownership in PISA 2009 and 2015 in the eight countries.

Table 1 shows how PFPS represent more than fifty percent of publicly funded schools in Belgium, the Netherlands and, more recently, Ireland, followed by Spain where this percentage is around thirty percent. The sample of schools in every PISA wave is a representative snapshot of the educational situation in each country, although the schools that compose the sample are not the same in each wave. Therefore, a pseudo-panel has to be used to compare and benchmark the countries.

Regarding the output side, PISA focuses on measuring the extent to which students are able to apply their knowledge and skills to fulfill future real-life challenges rather than evaluating how well they have mastered a specific school curriculum. The evaluation addresses three knowledge areas: reading, mathematics and science. Fair comparisons of the groups of schools over time are possible because PISA uses a common scale for the purpose of trends (OECD 2014, p. 159).

To select the input and output variables to be considered in the production of education, we follow the standard selection made in the literature (for a review, see De Witte & López-Torres, 2017). On the input side, we include the classical economics of education inputs required to carry out the learning process, including family background and school resources (Levin, 1974). Although PISA provides some indexes for measuring these dimensions, for example, economic, social and cultural status (ESCS) or quality of the school educational resources (SCMATEDU), they are composite indexes built by categorical principal component analysis and centered at zero in each wave. This means that these numbers cannot be directly introduced in the analysis and compared to measure how productivity gaps evolve over time from one wave to another. Therefore, the selected inputs must be comparable over time.

For student resources, we use two variables, which, averaged over students, represent student family background at school level, that is, the raw material for producing education (Bradley, Johnes & Millington, 2001, p.554). PARED is the index of the highest level of parental education, measured by the number of years of schooling according to the International Standard Classification of Education (ISCED; OECD, 1999). HISEI is the index of the highest parental occupational status according to International Socio-Economic Index of Occupational Status (ISEI; Ganzeboom, De Graaf, Treiman & De Leeuw, 1992).

For school resources, we use three inputs. First, we approximate the labor factor using the teacher/student ratio (TEACHSTUD) defined as the number of teachers per hundred students. The ratio is built as the total number of teachers weighted by their working hours (part-time teachers contribute 0.5 and full-time teachers 1) to the total number of pupils and multiplied by a hundred. Furthermore, like Aparicio et al. (2017) and Aparicio and Santín (2018), Santín and Sicilia (2015), we use school principal responses in the two PISA weights to build two indexes related with the quality of school resources for their use as inputs. Therefore, the second input 'MATERIAL' is derived from principal perceptions of potential factors, like shortages of educational materials, infrastructure or teaching staff, hindering the provision of instruction at school.[8] The third school input 'CLIMATE' captures school principal perceptions of the school climate related to teacher and student behavior that might influence the provision of instruction at school. This includes factors like students skipping classes, students bullying other students, teacher absenteeism or teachers being too strict with students.[9] Table 2 summarizes all the variables used in the analysis.

Continuing with the research, we selected the performance of Finnish public schools in PISA 2009 as the base-group base-period reference technology (R) defined in Eqs. (15)–(17) for the analysis. It is widely recognized in the economics of education literature that the public Finnish education system is one of the best performers in cross-country secondary education evaluations inside the European Union and around the world (see Afonso & St. Aubyn, 2005, 2006; Bogetoft, Heinesen & Tranæs, 2015; Cordero, Santín & Simancas, 2017; De Jorge & Santín, 2010 for a discussion of this result). The final number of schools included in the analysis together with descriptive statistics for each country by school types for outputs and inputs are provided in Tables 3 and 4, respectively.

Table 3 shows that in all countries and in both periods, outputs are higher in PFPS than in PS except in the Netherlands. Table 4 points out the same pattern with respect to parents' background (PARED and HISEI) and the school climate (CLIMATE) where PFPS dominate PS, the Netherlands being the exception. Regarding the number of teachers per hundred students (TEACHSTUD), we observe that PS use more human resources than PFPS in Belgium, Denmark, Spain and Belgium while the reverse happens in the Czech Republic, Hungary and Slovakia. The Netherlands is the only country in which this figure was slightly higher in PS than in PFPS in 2009, although the contrary was observed in 2015. Finally, the variable related with school resources available for education (MATERIAL) is always better in PFPS than in PS except in Ireland.

4.2. The 'reference' choice

As discussed in Aparicio and Santín (2018), there are two main strategies available in the literature that can be followed to ensure that circularity holds when comparing the performance of more than two groups of DMUs over time using productivity indexes. The first one is to build a common reference technology. This approach has been at the forefront of empirical applications in education since the beginning of DEA analysis and was originally proposed by Charnes, Cooper and Rhodes (1981). The idea was to distinguish between management efficiency and program efficiency, which were two groups of public schools applying and not applying 'Program Follow Through' in various parts of the United States. The idea behind this approach is to combine the best performers of each group included in the analysis to somehow draw the well-known meta-frontier[10] (see, for example, Battese & Rao, 2002, Battese, Rao & O'Donnell, 2004 and Johnes & Virmani,

[7] While the proportion of PFPS in France (Sweden) was greater than five percent in 2015 (2009), no information for identifying school ownership was provided in 2009 (2015). Other countries like Austria, Germany, Italy or Portugal do not exceed the minimum five percent threshold in either wave.

[8] The four response categories are weighted differently in order to build the index: 'not at all' – 1, 'very little' – 0.75, 'to some extent' – 0.5, and 'a lot' – 0.25. Similar questions were asked in both waves, but PISA 2009 asks about thirteen items (Question 11 in the school questionnaire) and PISA 2015 about eight (Question 17 in the school questionnaire). For the purposes of comparison, we have normalized the index to PISA 2009. Therefore, the maximum value is eight as in PISA 2015.

[9] The principal responses for 'CLIMATE' are weighted as described for 'MATERIAL' in such a way that more input reflects better school climate. While PISA 2009 asks about thirteen items (Question 17), PISA 2015 asks about ten items (Question 61). Therefore, we normalize the index in PISA 2009 to ten, the maximum value as in PISA 2015.

[10] Originally, Charnes et al. (1981) dubbed for the combination of production frontiers as 'inter-envelope', although meta-frontier is the term normally used in the literature.

Table 2
Inputs and outputs included in the analysis. Student average at school level.

Variable	Description
Outputs	
MATHS	Test score in mathematics
READING	Test score in reading
SCIENCE	Test score in science
Inputs	
PARED	Students' highest parental education level expressed as years of schooling
HISEI	Index of the highest parental occupational status
TEACHSTUD	Number of teachers per hundred students
MATERIAL	Index of principal perception of having good school resources for education. It includes staff, infrastructure and educational material
CLIMATE	Index of principal perception of having a good school climate related to student and teacher behavior.
Ownership	
PS	Public Schools. Owned and governed by the public sector.
PFPS	Publicly funded private schools. Privately owned schools owned that receive more than 50% of funds from the public sector.

Table 3
School sample and average PISA student scores in mathematics, reading and science.

	Sample size		Mathematics		Reading		Science	
Country & Ownership	PISA 2009	PISA 2015	PISA 2009	PISA 2015	PISA 2009	PISA 2015	PISA 2009	PISA 2015
Belgium PS	80	36	482	468	473	459	474	462
Belgium PFPS	141	41	538	517	527	512	529	514
Czech. Rep PS	182	217	512	500	495	495	517	501
Czech. Rep PFPS	8	19	550	510	548	505	571	507
Denmark PS	193	183	483	493	477	483	476	482
Denmark PFPS	28	26	518	520	517	511	521	511
Hungary PS	119	138	494	481	498	473	507	482
Hungary PFPS	17	28	506	493	508	485	516	490
Netherlands PS	65	40	539	522	523	513	538	517
Netherlands PFPS	93	55	533	512	515	503	531	509
Slovakia PS	127	164	497	482	478	459	492	466
Slovakia PFPS	11	23	511	489	504	470	507	479
Spain PS	449	121	483	480	476	488	485	487
Spain PFPS	262	52	510	502	506	517	509	511
Ireland PS	55	70	472	492	477	508	489	490
Ireland PFPS	61	76	492	509	503	528	513	509
Reference country								
Finland PS	148	120	539	512	534	527	551	530

Table 4
Average family background and school inputs in PISA by country and year.

	PARED		HISEI		TEACHSTUD		MATERIAL		CLIMATE	
Country & Ownership	PISA 2009	PISA 2015	PISA 2009	PISA 2015	PISA 2009	PISA 2015	PISA 2009	PISA 2015	PISA 2009	PISA 2015
Belgium PS	13.82	14.53	46.13	47.01	13.42	12.40	5.84	5.28	7.44	6.40
Belgium PFPS	14.38	15.04	51.56	55.56	11.68	10.90	6.26	5.65	8.01	7.46
Czech. Rep PS	13.70	13.40	49.84	48.12	8.02	7.94	6.47	6.35	7.55	7.73
Czech. Rep PFPS	14.69	13.90	58.21	54.16	9.66	9.09	6.71	6.70	7.62	8.41
Denmark PS	13.74	15.64	48.71	52.53	8.99	8.17	6.62	6.45	7.74	7.59
Denmark PFPS	14.51	16.68	54.58	60.35	8.12	7.91	6.97	7.03	8.72	8.22
Hungary PS	12.82	13.60	47.21	46.75	9.94	12.47	6.87	5.14	7.87	7.96
Hungary PFPS	13.62	14.22	51.38	52.66	13.16	14.54	7.06	6.14	8.31	8.58
Netherlands PS	14.20	14.12	52.93	55.56	6.61	5.83	6.32	6.06	6.98	6.57
Netherlands PFPS	13.96	13.90	51.52	53.20	6.36	6.03	6.44	6.11	6.92	6.58
Slovakia PS	13.26	14.69	45.95	48.08	7.11	7.86	6.10	6.34	7.27	7.93
Slovakia PFPS	13.52	15.00	48.71	52.37	9.19	9.34	6.25	6.40	7.83	8.18
Spain PS	11.86	11.70	43.28	42.39	12.18	9.64	6.32	5.05	7.35	7.38
Spain PFPS	13.31	13.84	50.16	56.61	6.73	6.45	6.41	6.36	8.30	8.55
Ireland PS	12.93	13.75	45.72	49.21	7.93	8.01	5.68	5.67	7.24	7.25
Ireland PFPS	13.27	14.04	50.08	53.84	7.28	7.13	5.54	5.62	7.53	7.54
Reference country										
Finland PS	14.92	15.07	54.18	53.65	9.15	9.58	6.12	5.86	7.35	7.36

2020 for a recent overview of meta-frontiers applications in education). Many strategies can be used to define the meta-frontier. For example, Pastor and Lovell (2005) proposed, as one possible solution, a Malmquist productivity index based on a reference technology constructed by using all observations for all the periods, known as global Malmquist. Other approaches, like the global Malmquist-Luenberger productivity index (Giménez, Thieme, Prior & Tortosa-Ausina, 2019), have been used to draw the meta-frontier in education.

All these approaches assume that the union of DMUs from different groups satisfies convexity. To relax the convexity assumption, Afsharian, Ahn and Harms (2019) proposed the use of a non-convex meta-frontier, named meta-technology, in order to evaluate the performance of different groups of DMUs in one period. This approach combines the production frontiers of the different groups without incorporating additional convex combinations of observations belonging to different group technologies. The same idea is behind the non-convex meta-frontiers drawn by Cordero et al. (2017, 2018b) in the context of education. They estimate the non-convex frontiers of every group, in their case schools within different education systems, and the non-convex meta-frontier using the robust order-m partial frontier analysis (Cazals, Florens & Simar, 2002) based on the nonparametric free disposal hull (FDH) methodology (Deprins, Simar & Tulkens, 1984), which does not assume convexity either.

The second idea for making group comparisons relies on resorting to a fixed-group fixed-period production technology. This strategy was previously developed by Berg et al. (1992, 1993) and Balk and Althin (1996). These papers propose new insights for calculating the Malmquist index, where a base technology, a particular production unit, is taken as the comparison reference, or a fixed period is taken as the baseline in order to calculate multi-period Malmquist indices.

From our point of view, the original idea of Camanho and Dyson (2006) for evaluating and comparing the performance of different groups of production units is to avoid the first strategy, i.e. the different versions of meta-frontiers, focusing the attention on the second strategy. The spirit of this approach is clearly stated in Camanho and Dyson (2006, p. 36) as follows: "In contrast with earlier methods, our method makes comparisons relative to groups-specific frontiers only, without pooling the DMUs together to form a common frontier". As they claim subsequently, "An advantage of this approach is that it does not assume convex combinations of group-specific frontiers to be feasible. Specifically, even if group-specific production sets satisfy convexity, there is no reason why the union of these sets should be convex." Of course, one way to define a more reliable meta-frontier is to relax the convexity problem, as in Cordero et al. (2017, 2018b) and Afsharian et al. (2019), because it does not assume the feasibility of convex combinations of group frontiers. However, it does not get around the fact that the meta-frontier is still drawn after pooling all DMUs from different groups together.

We think that both strategies, meta-frontiers and the use of a fixed-group fixed-period production technology, are valid, and both can be applied in real-world problems after evaluating and assuming the pros and cons of both approaches. The idea of our paper directly stems from Camanho and Dyson (2006) approach of avoiding pooling DMUs from different groups to evaluate the groups discarding to draw a meta-frontier. The reason is that we believe that, in some empirical problems, for example in education, it makes more sense from an economic point of view to compare different groups facing different circumstances using a real reference that also has its own legal framework, budget constraints and other clear rules and environment to be copied. To some extent the comparison of real groups provides a view about what is feasible. We also think that the "All-Star technology"[11] approach is engaging from a mathematical point of view because it defines the best feasible synthetic technology even if DMUs included in such technology do not operate together.

A further question is whether the reference choice changes the results. It is well known in the productivity literature that the price for gaining circularity is paid with reference dependency (for a review of this issue see Althin, 2001). This means that choosing another reference will bring about another set of productivity differences among the evaluated groups. At this point, we can always drop methodologies based on a fixed technology, such as meta-frontiers or base-group base-period references and pool all DMUs from all groups together to draw a common technology for each period in order to average out results by groups. However, recall that resorting to the adjacent Malmquist index (Färe et al., 1994) guarantees that results are independent of a base technology but fails the circularity test.

For our empirical application in education, we know that the cost of our approach is that results might vary depending on the chosen reference. Nevertheless, for monitoring and evaluating education systems over time, the relevance of holding the circular test for policymakers is high. For example, within a Federal State responsible for monitoring productivity differences among regions or states with fully decentralized educational competencies, it makes sense that the productivity gap of public schools between region A and region B and region B and C should give the same result as going directly from region A to C.

Knowing that results will depend on the reference, it is important to define at least two rough guidelines for choosing a reference technology that can be used for policy purposes. On one hand, the reference should be an education system viewed by the other education systems, stakeholders, or an expert system, by consensus or majority vote, as a key benchmark for all other education systems. On the other hand, as educational technology and school management might change over time, a good practice will be to predefine a number of years for holding the same reference. Once this period has elapsed, the stakeholders should decide whether to keep the same reference or adopt a new one for the next period. In a research paper, this choice is made by the paper authors based on previous literature. To conclude, we think that, at the end of the day, the final choice is an empirical issue and will depend on the final targets of each real application.

5. Results

To conduct the analysis, we apply the new CDLI and PPLI. To do this, we first need to establish the directional vector in the different periods of time and in the different groups. In particular, we fix a common vector equal to the average value of the set of all the data in all the periods for those variables which have economic sense to increase or decrease their values from an economic point of view. The vector g in Eq. (12) is defined as $g =$ *(MATERIAL, TEACHSTUD, CLIMATE, PARED, HISEI, MATH, READ, SCIE)*, taking the following values $g = (0, 9.29, 0, 0, 0, 500.94, 493.73, 500.58)$. Therefore, the DDF allows schools to reach the production frontier by increasing outputs but also by reducing 'teachers per hundred students' for those groups of schools that have more teachers per hundred students than other groups. Other inputs cannot be reduced, as it makes less economic sense in this sector.

5.1. The CDLI. productivity gaps between PS and PFPS across countries

The first research question of this analysis is to measure the productivity gaps between PS and PFPS across countries through the CDLI. From the taxpayer and the policy maker viewpoints, it is important to explore which are the productivity differences across Europe to identify the characteristics of the best performers in terms of the educational management of public resources. For calculating the CDLI we use Eq. (15). Furthermore, using Eq. (16) we decompose the CDLI in an efficiency gap (EG) and a technological gap. Table 5 provides CDLI results and its components for the two PISA waves.

In Table 5, we compare the performance gap of PS in each country with respect to their PFPS counterparts. A positive (negative) sign in CDLI is interpreted as a better average performance of PS (PFPS) with respect to PFPS (PS) inside the analyzed country. The higher the absolute value, the greater the gap will be.

[11] We regard meta-frontiers as the NBA (National Basketball Association) 'All-Star Game'. It is a pleasure to watch the game between the best NBA players from the different teams able to draw the best basketball technology every year. These dream teams are a reference, and, of course, the reference is somehow feasible because the game takes place every year. However, in practice, real NBA teams face salary caps, making the All-Star Game just an unfeasible synthetic reference.

Table 5
Base-group Luenberger index in 2009 and 2015 for ownership comparisons by country and its decomposition in efficiency gap (EG) and technology gap (TG).

	2009			2015		
PS/PFPS	CDLI	EG	TG	CDLI	EG	TG
Belgium	−0.0507	0.0080	−0.0587	−0.0829	−0.0206	−0.0622
Czech Republic	0.1482	−0.0506	0.1988	0.1152	−0.0417	0.1569
Denmark	0.0023	−0.0196	0.0219	−0.0095	−0.0291	0.0197
Hungary	0.0720	−0.0392	0.1112	0.0309	−0.0466	0.0775
Ireland	−0.0941	−0.0074	−0.0868	−0.0244	−0.0186	−0.0058
Netherlands	−0.0238	0.0150	−0.0387	−0.0640	0.0067	−0.0707
Slovakia	0.1092	−0.0352	0.1445	0.0887	−0.0337	0.1224
Spain	−0.1465	−0.0319	−0.1146	−0.0534	−0.0248	−0.0287

As our directional vector g is common and fixed in both periods, its interpretation is straightforward. For example, the first value in Table 5 is the CDLI between PS and PFPS in Belgium, equal to −0.0507. As this value is negative, it indicates that PS perform worse than PFPS in Belgium. To be more specific multiplying the absolute value 0.0507 by $g=(0, 9.29, 0, 0, 0, 500.94, 493.73, 500.58)$ indicates that, on average, PS might reduce TEACHSTUD in $9.29 \times 0.0507 = 0.4710$ teachers by a hundred students and simultaneously increase their average results in mathematics, reading and science in $500.94 \times 0.0507 = 25.40$; $493.73 \times 0.0507 = 25.03$; $500.58 \times 0.0507 = 25.38$ PISA points, respectively, to reach the PFPS average performance.

As we can see in both CDLI columns of Table 5, PFPS performance is better in Belgium, Ireland, the Netherlands and Spain in both waves, whereas PS productivity outperforms PFPS in the Czech Republic, Hungary and Slovakia. In Denmark, the CDLI values are almost zero in both years, slightly positive in 2009 and negative in 2015, indicating very similar productivity levels for both school types.

Regarding the components that explain these gaps, first, we observe that efficiency gaps are practically negative in all countries in both years. This result suggests that PFPS are, on average, more efficient than PS. Therefore, their distance to their own frontier is generally lower than for PS. The exception is the Netherlands, where results are the opposite for both years, indicating a better PS efficiency. Similarly, the efficiency gap in Belgium benefitted PS in 2009, although this trend changed in 2015. Second, as Table 5 shows, the CDLI is driven, in most of cases, by the technology gap. Best performers account for a large variation in the average productivity difference between both school types. In the above countries, where PFPS (PS) outperform PS (PFPS) with respect to productivity, the technology gap is better too. Again, the exception is Denmark 2015, where the PS technology outperforms PFPS technology (positive sign). Nevertheless, the CDLI is negative as a consequence of the negative efficiency gap.

5.2. Productivity gaps between PS and PFPS over time

The second research question is testing whether the productivity gaps between PFPS and PS widen or narrow over the 2009 and 2015 periods in the eight European countries. Table 6 shows the PPLI estimated using Eq. (17).

To interpret Table 6 results, the first step is to remember that a positive (negative) PPLI value means that the performance of PFPS (PS) has improved with respect to PS (PFPS) from 2009 to 2015. In six out of the eight countries, the PPLI is positive, indicating a better performance of PFPS in relation to PS, although each case should be interpreted differently according to Table 5. First, the gap in favor of PFPS in Belgium and the Netherlands is wider in 2015 than in 2009. Second, as we mentioned above, the productivity gap between PFPS and PS for the Czech Republic, Hungary and Slovakia favors PS in 2015, but PFPS have closed the gap since 2009. Finally, while PS outperformed PFPS in Denmark in 2009, the situation has reversed in 2015, and PFPS are now slightly more productive than PS. However, the gap in both years is close to zero, and differences are the smallest compared with the other countries.

Table 6
Pseudo-panel Luenberger index over the period 2009–15 by ownership and country and its decomposition into efficiency gap change (EGC) and technology gap change (TGC).

	2009 – 2015		
PS/PFPS	PPLI	EGC	TGC
Belgium	0.0321	0.0286	0.0035
Czech Republic	0.0331	−0.0089	0.0419
Denmark	0.0118	0.0096	0.0022
Hungary	0.0411	0.0074	0.0337
Ireland	−0.0697	0.0112	−0.0810
Netherlands	0.0402	0.0082	0.0320
Slovakia	0.0205	−0.0015	0.0220
Spain	−0.0931	−0.0071	−0.0859

Analyzing the two components, we observe that in the above six countries with a positive PPLI, the TGC is positive too, indicating that the best PFPS have improved their technology with respect to PS. However, we observe that the sign for EGC in the Czech Republic and Slovakia is negative, leading to a better efficiency of PS with respect to PFPS in 2015 compared to the same difference in 2009. It is interesting to note that the TGC is largest in favor of PFPS in the Czech Republic. If only a small percentage of PFPS are improving the technology, this probably leads to more inefficiency inside the PFPS than in PS, although the final balance is positive for PFPS.

In two countries, Ireland and Spain, PS have managed to close the productivity gap, which still favors PFPS. In both countries, the best public schools have shifted up the production frontier, reducing the technology gap with respect to PFPS. While in Ireland this frontier shift has worsened the efficiency gap, PS in Spain have also reduced the efficiency gap with respect to PFPS. In the specific case of Spain, the severe economic crisis during the analyzed period led to large education budget cuts that might be behind this gain in productivity.[12] Fig. 1 summarizes the main results of Tables 5 and 6.

In Fig. 1, CDLI for 2009 with respect to the reference is measured on the x-axis, whereas the CDLI for every country in 2015 is captured on the y-axis. Within this framework, the origin of the coordinates (0, 0) in Fig. 1 represents the ideal situation for an education system, where there is no productivity gap between PFPS and PS. As we can see, Denmark is the only country close to this position.

The graph has four quadrants. For countries in the northeast quadrant, PS perform better than PFPS in both periods, whereas

[12] During the financial crisis, the Spanish Law for the Improvement of Educational Quality (*Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa (LOMCE)*) introduced measures like increasing the student-teacher ratio in high school classrooms or suspending teacher replacements after retirement, which helped to reduce the performance gap with respect to PFPS.

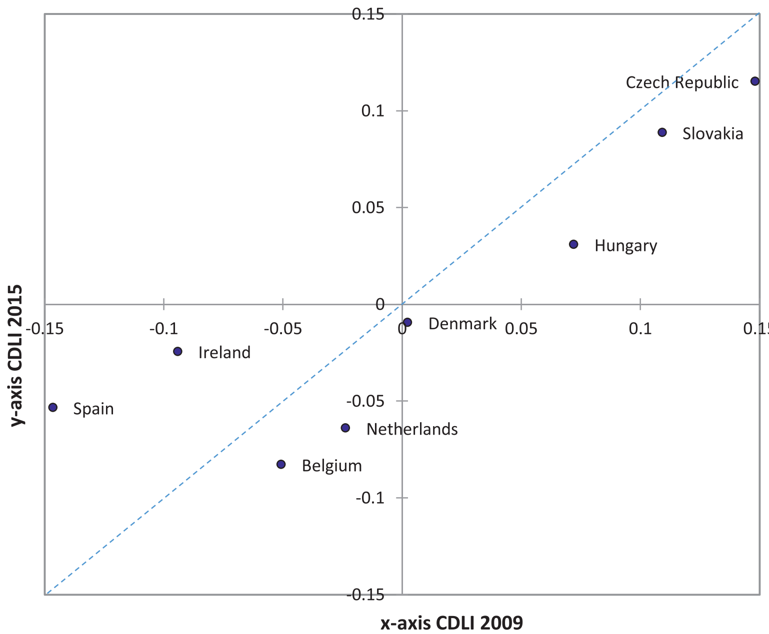

Fig. 1. CDLI by school ownership. Evolution across countries (PISA, 2009–2015).

the opposite applies to countries in the southwest quadrant: PS perform worse than PFPS in both years. Denmark is the only country in the southeast quadrant, indicating that, in this country, PS outperformed PFPS in terms of productivity in 2009, whereas the situation reversed in 2015, although the difference was close to zero in both years. In our empirical case, the northwest quadrant is empty, which means that we have not found any country where PFPS were more productive than PS in 2009, and the contrary was true in 2015. Additionally, a 45° degree line runs through the northeast and southwest quadrants, splitting them into two. The upper halves of these quadrants indicate a negative PPLI, as for Ireland and Spain, where PS are catching up with PFPS in terms of average productivity. The opposite applies to countries in the bottom halves, where the PPLI is positive, and PFPS did better than PS over the analyzed period. In the northeast quadrant, this means closing the gap with respect to the more productive PS, as applies to the Czech Republic, Slovakia and Hungary. In the southwest quadrant, PFPS in the Netherlands or Belgium are widening the original productive gap with respect to PS.

In general, plotting the CDLI for sequential periods is more informative than just providing the PPLI scores, as it allows observing which of the settings discussed in Section 3 occurred. The PPLI is just an overall score of the evolution of the productivity gap between groups that summarizes the detailed information and insights gained from the observation of Fig. 1.

5.3. A step backwards. productivity gaps evolution between PS and PFPS across countries

Apart from these results, the methodology can be used to monitor differences by school ownership not only within countries but also across countries. Eqs. (15) and (16) can be computed using the reference, PS in Finland 2009, as one group and the sixteen groups of schools, eight countries multiplied by two school types in both periods, as the second group. Table 7 shows the results.

We should underscore that the distances in Table 7 can be basically interpreted as shown in Table 5, i.e. a positive (negative) sign in CDLI is interpreted as the average performance of the analyzed group being better (worse) than the reference (Finnish PS in 2009). The higher the absolute value is, the wider the gap will be. Second, although the direct interpretation of these values is not clearly evident, the circularity of this index allows us to employ DDF values in Table 7 like 'bricks' for building all CDLI and PPLI measures presented in Tables 5 and 6. For example, the first value in Table 5 indicates the CDLI between PS and PFPS in Belgium 2009, and it was equal to -0.0507. From Eq. (15) and Table 7 we appreciate that the distance of PS in Belgium 2009 to the production frontier given by the reference technology is $\frac{1}{n^{s,B}}\sum_{k=1}^{n^{s,B}}\vec{D}_V^{h,R}(X_k^{s,B},Y_k^{s,B};g) = -0.0020$; the distance of PFPS in Belgium 2009 to the reference is $\frac{1}{n^{s,A}}\sum_{j=1}^{n^{s,A}}\vec{D}_V^{h,R}(X_j^{s,A},Y_j^{s,A};g) = 0.0487$. In this way, the $CDLI_s^{AB}(R^h) = -0.0020 - 0.0487 = -0.0507$.

Third, another way to looking at Table 7 is that we can straightforwardly find the performance gap between any pair of groups using again the circularity property discussed in Eqs. (21) and (22). Let us assume that we want to know, say, the productivity gap between PS in Hungary and PFPS in Slovakia during 2015. To do this, we have to subtract the distance between PS in Hungary with respect to PS in Finland from the distance between PFPS in Slovakia and Finland, i.e. 0.0245 – (– 0.0293) = 0.0538. As the sign is positive, the result means that PS in Hungary are more productive than PFPS in Slovakia.

The potential of these measures is illustrated in Fig. 2, where the distances represent the productivity gap among the groups in the two years.

In Fig. 2, the origin of the coordinates (0, 0) represents now the productivity of the reference, PS in Finland 2009. Taking an axis (2009 or 2015), we can sort groups by productivity and explore the productivity gaps between the school groups. This allows us to derive several conclusions. First, from Fig. 2 we conclude that PFPS and PS in the Netherlands clearly hold the top two positions in 2015 (y-axis), respectively. Interestingly, productivity differences between Dutch schools are moderate and quite constant constituting the educational production benchmark whose practices and law framework should be analyzed in depth. Second, in 2009 there was large productivity differences in favor of PFPS in Spain and Ireland, however, most of this gap was closed during the six-year period considered. As we have already mentioned the economic crisis brings about a shortcut of resources in public schools that possibly contributed to close the gap. Thirdly, we can see that in the Czech Republic and Slovakia PS outperform PFPS in both periods. Although some convergence has taken place, the gap remains sig-

Table 7
CDLI in 2009 and 2015 with respect to PS in Finland 2009 by ownership and country and their decomposition into efficiency gap (EG) and technology gap (TG).

		2009			2015		
		CDLI	EG	TG	CDLI	EG	TG
PS	Belgium	−0.0020	0.0036	−0.0056	−0.0734	−0.0118	−0.0616
	CzechRep	0.1598	−0.0150	0.1747	0.1384	−0.0175	0.1559
	Denmark	0.0000	0.0030	−0.0030	−0.0100	−0.0126	0.0027
	Hungary	0.0970	−0.0019	0.0989	0.0245	−0.0188	0.0432
	Ireland	0.0704	0.0175	0.0529	0.1271	0.0013	0.1257
	Netherlands	0.1741	−0.0097	0.1839	0.2629	0.0005	0.2623
	Slovakia	0.1668	−0.0062	0.1730	0.0594	−0.0143	0.0737
	Spain	−0.0305	−0.0357	0.0052	0.0573	−0.0075	0.0647
PFPS	Belgium	0.0487	−0.0044	0.0531	0.0095	0.0088	0.0007
	CzechRep	0.0115	0.0356	−0.0241	0.0232	0.0243	−0.0011
	Denmark	−0.0023	0.0225	−0.0249	−0.0005	0.0165	−0.0170
	Hungary	0.0250	0.0372	−0.0123	−0.0064	0.0279	−0.0343
	Ireland	0.1646	0.0249	0.1397	0.1515	0.0200	0.1315
	Netherlands	0.1979	−0.0247	0.2226	0.3269	−0.0062	0.3330
	Slovakia	0.0575	0.0290	0.0285	−0.0293	0.0194	−0.0487
	Spain	0.1160	−0.0038	0.1198	0.1107	0.0173	0.0934

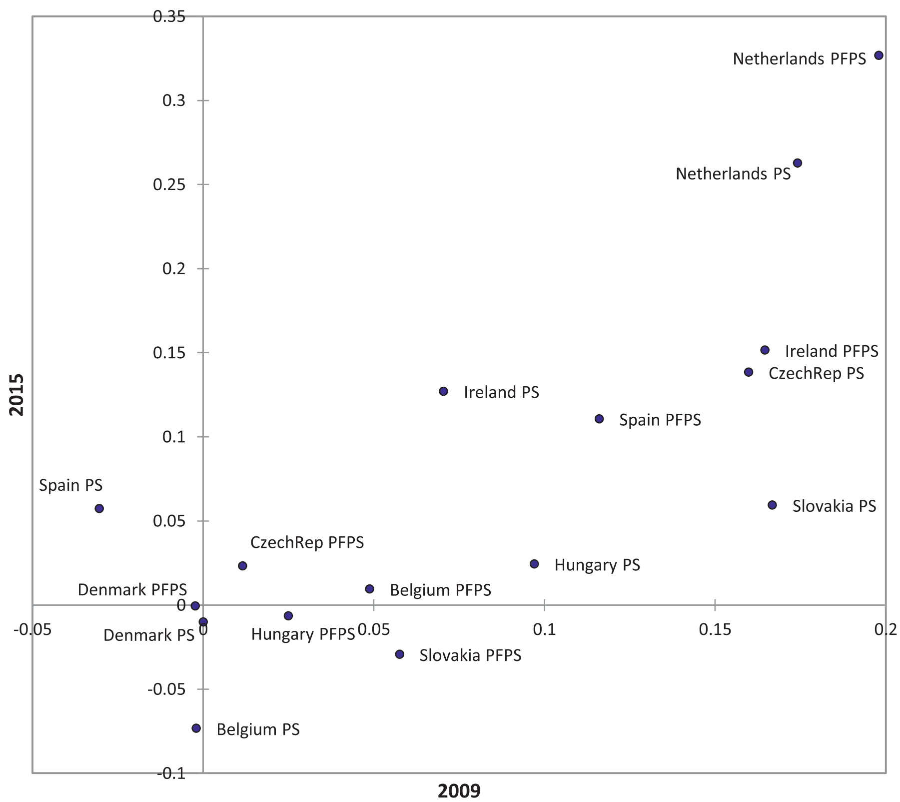

Fig. 2. Productivity gap evolution (CDLI) by school type across countries (PISA, 2009–2015).

nificant in 2015. Finally, it is interesting to see how the closest differences between both school types and years can be observed in Denmark. However in 2015, all schools in Denmark, together with PS in Belgium and PFPS in Hungary and Slovakia have a negative CDLI and are still below the reference.

Figs. 3 and 4 plot the efficiency gap and the technology gap evolution across countries in both periods to provide some insights into the main factors driving the productivity gaps.

Fig. 3 confirms how in terms of efficiency gaps PFPS are closer to their production frontier than PS. This finding is consistent with the economic theory, according to which, as private companies, PFPS should be more efficient at managing resources while, at the same time, delivering good outputs to assure that this option is appealing for the parents when they make their school choice. Curiously, the exception to this pattern is the group of PFPS in the Netherlands, having the best performers inside its group they still have room for further improvements. PS in Ireland and the Netherlands are the best PS systems and in 2015 they had a similar internal efficiency in comparison with the reference. Another interesting case is that the efficiency gap of PS in Spain 2009 was clearly the worst performer. However, in 2015 PS in Spain continues below the reference but with similar values to other PS systems. Looking at Fig. 3 axis values, we can also conclude that the efficiency gaps are low and differences between PS and PFPS remain quite stable in both periods.

We should highlight that the results in Fig. 4 are similar to results illustrated in Fig. 2, suggesting that the main driver of the productivity gap is the technology gap. Note also in this case that there is no clear pattern for directly identifying whether PS or PFPS are the best performers inside a country although we will come

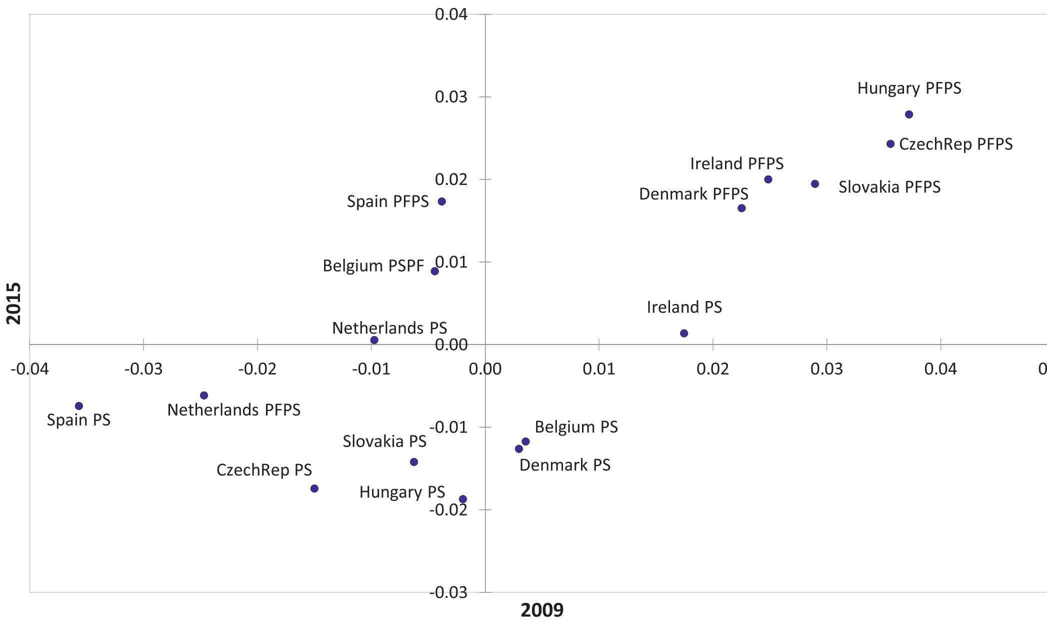

Fig. 3. Efficiency gap evolution by school type across countries (PISA, 2009–2015).

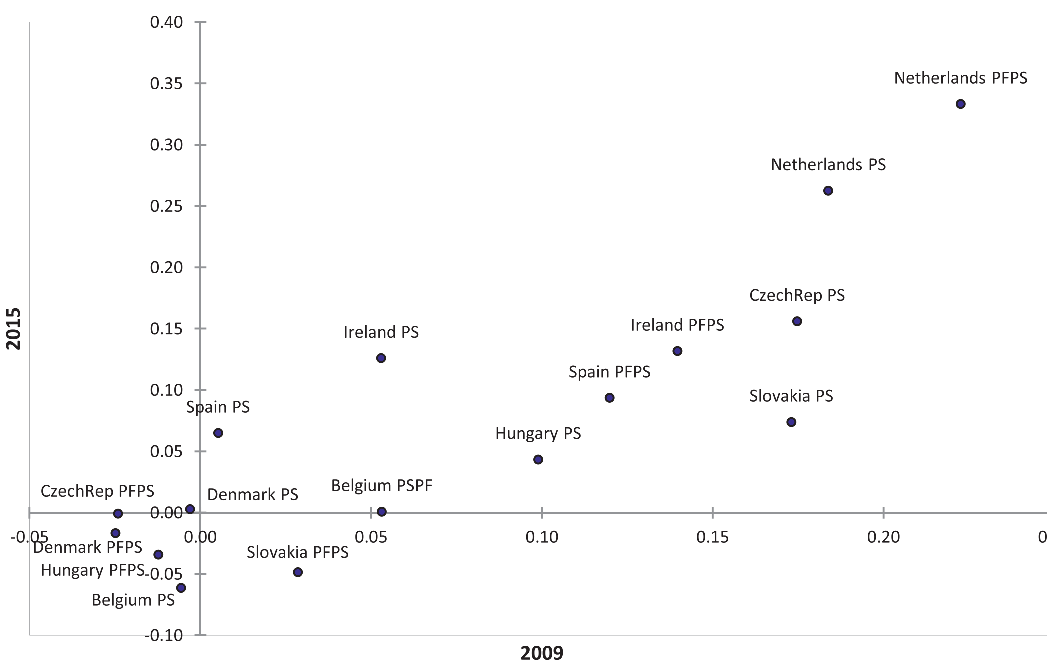

Fig. 4. Technology gap evolution by school type across countries (PISA, 2009–2015).

back on this issue on Section 5.5. However, Fig. 4 shows how the technical gap for Ireland and Spain was significantly reduced between both periods explaining the reduction in the productivity gap that we discussed in relation to Fig. 2. We also observe that four out of eight PFPS systems, Czech Republic, Denmark, Hungary and Slovakia (we could also include Belgium with a technical gap close to zero) had worse productivity than the reference, which clearly indicates that there is room in these countries for shifting up the production frontier in comparison with other countries.

Going back to Table 7, we can use Eqs. (23) and (24) to find out if at the end of the period the performance of a particular group of schools is better or worse than the reference. In other words, this measurement is the comparison of a country with itself in two different periods. For example, for PFPS in Belgium, we have $0.0095 - 0.0487 = -0.0392$, where the negative sign indicates that the performance of PFPS in Belgium was worse in 2015 than in 2009. The performance of PS in Belgium also deteriorated during the period and even more than PFPS: $-0.0734 - (-0.0020) = -0.0714$. As a result, the gap between PFPS and PS is $-0.0392 - (-0.0714) = 0.0322$, which matches the result in Table 4 for Belgium because of the circularity property.[13] Applying Eq. (17), Table 8 summarizes the productivity gap changes of all groups between both years.

For ten out of the sixteen groups, the positive sign in the PPLI indicates that the performance of these groups of schools was worse in 2015 than in 2009. The Netherlands is the country where school productivity grew most, regardless of school ownership.

13 There are slight differences in some cases due to rounding.

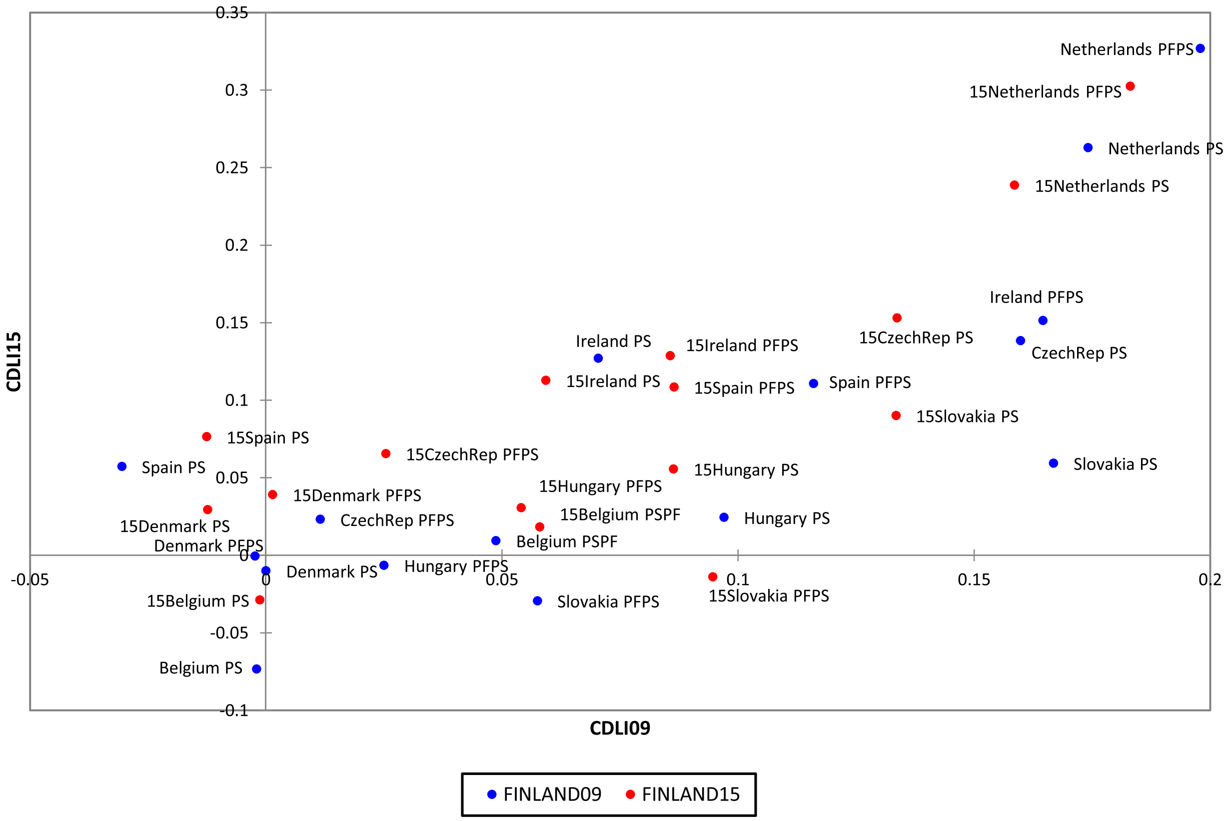

Fig. 5. CDLI by school type across countries (PISA, 2009–2015) using Finland 2009 and Finland 2015 as reference. The name of the countries begins with '15' when the estimations are done using Finland 2015 as reference.

Table 8
CDLI over the period 2009–2015 with respect to public schools in Finland (the reference) by ownership and country and decomposition into efficiency gap change (EGC) and technology gap change (TGC).

		2009–2015		
		$CDLI^{A}_{t,t+1}(R^h)$	$EG^{A}_{t,t+1}$	$TG^{A}_{t,t+1}(R^h)$
PS	Belgium	0.0713	0.0153	0.0560
	CzechRep	0.0214	0.0025	0.0189
	Denmark	0.0100	0.0156	−0.0056
	Hungary	0.0725	0.0168	0.0557
	Ireland	−0.0567	0.0162	−0.0728
	Netherlands	−0.0887	−0.0102	−0.0785
	Slovakia	0.1073	0.0080	0.0993
	Spain	−0.0877	−0.0282	−0.0595
PFPS	Belgium	0.0392	−0.0132	0.0524
	CzechRep	−0.0117	0.0113	−0.0230
	Denmark	−0.0018	0.0060	−0.0079
	Hungary	0.0314	0.0094	0.0220
	Ireland	0.0131	0.0049	0.0082
	Netherlands	−0.1289	−0.0185	−0.1104
	Slovakia	0.0868	0.0095	0.0773
	Spain	0.0053	−0.0211	0.0264

This growth is driven by technical change, suggesting that some schools and families successfully adapted the education process during the crisis to deliver similar results with fewer resources.

5.4. Robustness check

In Section 3.2 we highlighted that the price for having circularity in this monitoring system is paid with the technology dependence. Once we know that results will change with another reference, the question is raised as to what extent the new results would be different. Although an in-depth analysis devoted to this question is beyond the aim of this paper and deserves its own research, we have run all the analysis again using the set of Finnish public schools in PISA 2015 as the reference technology.[14] Table 9 reproduces Table 7 but now using the new reference.

The first obvious conclusion from Tables 7 and 9 is that the efficiency gaps coincide. This is not surprising because according to Eq. (16), this term is the same regardless of the reference. This result also holds for Eqs. (3) and (5). Table 10 presents the bivariate Pearson correlation coefficients for the CDLI and the TG component in both periods included in Tables 7 and 9. The main diagonal in Table 10 shows that correlations are high, positive and statistically significant, revealing that the initial underlying differences found among the groups of DMUs are still maintained after changing the reference.

We also explore the dissimilarities that we obtain when opting for another reference through two additional ways. First, using a Kolmogorov-Smirnov's test, we compare the cumulative distribution functions of the two empirical distributions of productivity gaps to assess how similar they are in shape and position. The *D* statistic ranges between 0 and 1, and a high value indicates a significant difference in the distributions of the two groups (see Banker, Zheng & Natarajan, 2010, p.234 for details). Second, we test whether there are any differences between the mean values of the productivity gaps found through a *t*-test for paired samples. Both results are shown in Table 11.

Regarding the Kolmogorov-Smirnov D values, we find that there are no statistically significant differences among the shape of the two paired empirical distributions of the four productivity gaps measured using the two references. However, the *t*-test indicates a different significant mean for the CDLI15 and TG15 depending on the reference used. Mean differences among these two productiv-

[14] For brevity we will use Finland 2009 and Finland 2015 for referring to PS from Finland in PISA 2009 and PS from Finland in PISA 2015 respectively,

Table 9
CDLI in 2009 and 2015 with respect to PS in Finland 2015 by ownership and country and their decomposition into efficiency gap (EG) and technology gap (TG).

		2009			2015		
		CDLI	EG	TG	CDLI	EG	TG
PS	Belgium	−0.0013	0.0036	−0.0049	−0.0289	−0.0118	−0.0171
	CzechRep	0.1337	−0.0150	0.1487	0.1530	−0.0175	0.1705
	Denmark	−0.0124	0.0030	−0.0154	0.0294	−0.0126	0.0421
	Hungary	0.0863	−0.0019	0.0883	0.0557	−0.0188	0.0745
	Ireland	0.0593	0.0175	0.0418	0.1128	0.0013	0.1115
	Netherlands	0.1586	−0.0097	0.1683	0.2387	0.0005	0.2382
	Slovakia	0.1334	−0.0062	0.1397	0.0902	−0.0143	0.1044
	Spain	−0.0126	−0.0357	0.0231	0.0765	−0.0075	0.0839
PFPS	Belgium	0.0580	−0.0044	0.0624	0.0182	0.0088	0.0094
	CzechRep	0.0254	0.0356	−0.0102	0.0656	0.0243	0.0413
	Denmark	0.0014	0.0225	−0.0211	0.0392	0.0165	0.0227
	Hungary	0.0540	0.0372	0.0168	0.0306	0.0279	0.0028
	Ireland	0.0856	0.0249	0.0608	0.1287	0.0200	0.1087
	Netherlands	0.1830	−0.0247	0.2077	0.3024	−0.0062	0.3086
	Slovakia	0.0947	0.0290	0.0657	−0.0139	0.0194	−0.0334
	Spain	0.0864	−0.0038	0.0902	0.1085	0.0173	0.0912

Table 10
Pearson correlation coefficients of CDLI and TG values using different reference technologies.

		(Reference) PS FINLAND 2015			
	Variables	CDLI09	TG09	CDLI15	TG15
(Reference) PS FINLAND 2009	CDLI09	**0.9402**	0.9079	0.7694	0.7682
	TG09	0.9134	**0.9527**	0.8081	0.8335
	CDLI15	0.7538	0.7924	**0.9883**	0.9649
	TG15	0.7532	0.8199	0.9857	**0.9871**

Table 11
Differences in empirical distributions and mean productivity gap values under two different references.

Variables	Kolmogorov-Smirnov D	Paired t-test
CDLI09	0.2500	1.0830
	(0.6325)	(0.2959)
TG09	0.1875	1.0807
	(0.9123)	(0.2969)
CDLI15	0.3125	2.3272*
	(0.3481)	(0.0344)
TG15	0.3125	2.3308*
	(0.3481)	(0.0341)

p-values are shown in parentheses.
* Means Difference is statistically significant at 5%.

ity gaps are around one point and a half higher when we use PS from Finland 2015 as reference with respect to using Finland 2009. To be more specific, the mean value for CDLI15 (TG15) when using Finland 2009 is 0.0732 (0.0703) while the average value is 0.0879 (0.0850) when the reference is Finland 2015.

Finally, to visualize the similarities and differences that arise from using different references in the CDLI and the TG, Figs. 2 and 4 are plotted again to include the results of Finland 2015 as reference.

In Figs. 5 and 6, the origin of the coordinates (0, 0) simultaneously represents the productivity of both references, Finland 2009 and Finland 2015, for evaluating the same sixteen groups when the reference is one or the other respectively. Figs. 5 and 6 confirm the previous findings as we knew that different references bring about different results. However, the scatter plots confirm that there is a high inertia among the CDLI, and the TG gaps measured under the two references. This means that although absolute productivity gaps with respect to the reference vary, the relative performance gaps of the groups hold quite stable. Of course, the variation in absolute numbers might derive in a slightly better off (or get worse) performance of some groups with respect to others but the results seem to be consistent for analysing how the groups of DMUs evolve over time.

Another type of robustness that should be checked is that associated with how sensitive the results are with respect to the selection of inputs and outputs. In this regard, some papers, like the recent contribution by Landete, Monge and Ruiz (2017), have suggested possible solutions. In particular, Landete et al. (2017) propose the calculation of a robust (radial) efficiency score that takes into account all the scenarios associated with all of the specifications of inputs and outputs that could be considered once a given set of input and output variables is defined. However, a new methodological extension of this technique to the case of considering directional distance functions instead of radial measures would be necessary. This extension is not trivial since the directional distance function determined for each scenario (combination of inputs and outputs considered) would depend on a different directional vector, which has consequences regarding how inappropriate it would be to directly aggregate all the inefficiency scores.

Another interesting robustness check would be adding weight restrictions to our models for incorporating experts' judgements on the importance of individual input and output variables (see, for example, Allen, Athanassopoulos, Dyson & Thanassoulis, 1997). Following this line, we could also check how results change using weight restrictions with respect to the free weights model calculated in this paper.[15]

5.5. Mind the gap. the Role of school autonomy on the CDLI and its components

Previous papers have analyzed the relationship between efficiency scores and some characteristics of educational systems

[15] In this empirical application, we solved thousands of optimization linear programs for each school in each country and with respect to the reference technology and the group-specific technology in 2009 and 2015, yielding also thousands of shadow weight vectors. The results are not reported but are available under request.

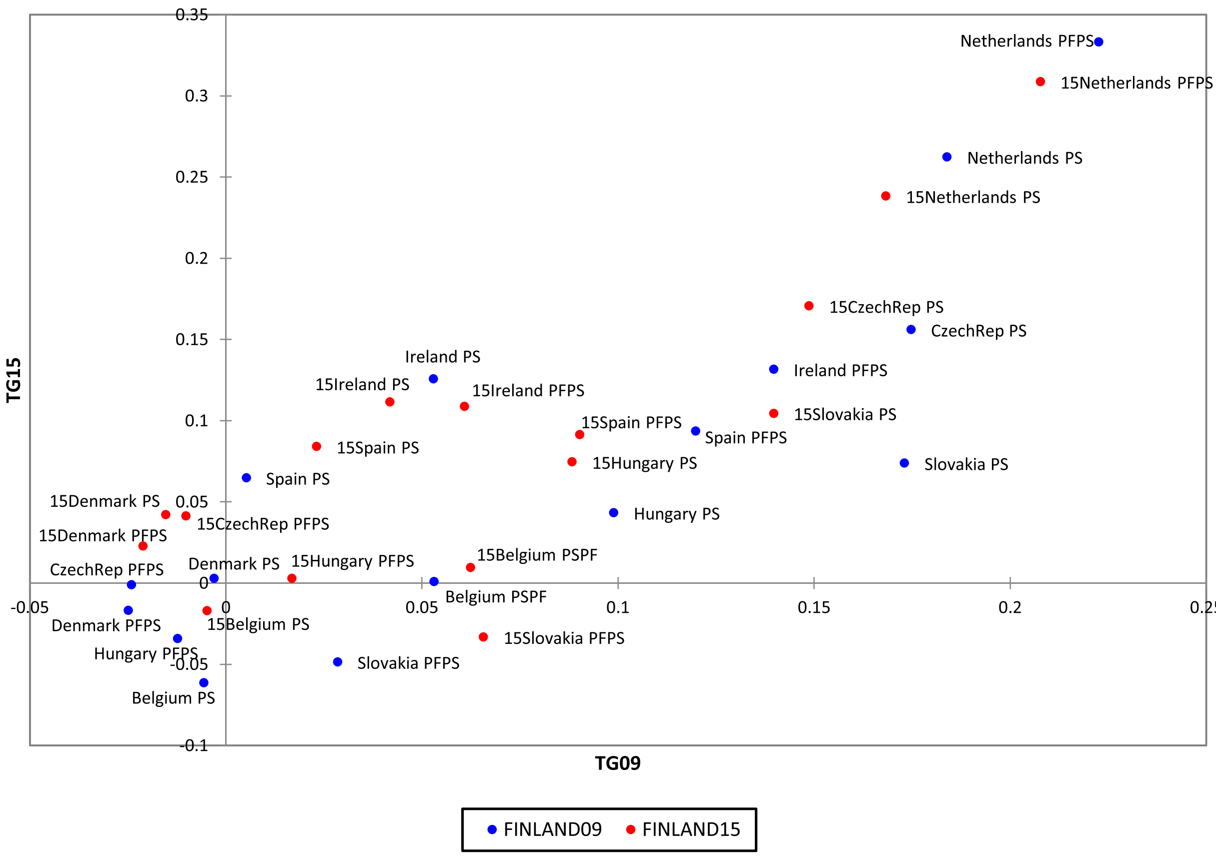

Fig. 6. Technology Gap by school type across countries (PISA, 2009–2015) using Finland 2009 and Finland 2015 as references. The name of the countries begins with '15′ when the estimations are done using Finland 2015.

across countries through correlations (Agasisti, Munda & Hippe, 2019), the conditional approach (Cordero et al., 2017, 2018b) or through a second stage analysis (Bogetoft et al., 2015). Hence, a question worth asking is why do productivity gaps widen or narrow between school types and across countries over time.

In addition to the underlying legal framework, the culture and social background of each country, differences in the school organization and management practices might be behind these differences. School autonomy has been pointed out in last years as one of the most powerful drivers for improving schools' productivity (Bloom, Lemos, Sadun & Van Reenen, 2015; Hanushek, Link & Woessmann, 2013; Woessmann, Luedemann, Schuetz & West, 2009). School ownership, whether public or private, implies in itself, differences in school autonomy levels, being general higher in PFPS in comparison with PS. However, we can exploit the fact that schools' autonomy varies across countries and over time to test whether a higher school autonomy affects productivity gaps while controlling country and year fixed effects.

In order to explore the variation of the CDLI and its components across the eight European countries, we estimate the following regression model.

$$P_{ist} = \alpha + \beta A_{ist} + \varphi S_{it} + \gamma_i + \gamma_t + \varepsilon_{ist}, \tag{26}$$

where P_{ist} stands for the productivity outcome of the group of schools s (PS and PFPS) in country i at period t being assessed. The productivity outcomes are the Luenberger distances against the reference reported in Table 7 (the CDLI, the efficiency gap and the technical gap). A_{ist} accounts for variables related with school autonomy and S_{it} is a dummy variable that takes value one if the school is public. Moreover, γ_t is a time dummy and γ_i includes a set of state dummies to take out countries fixed effects. Finally, ε_{ist} represents the error term.

We build two measures of school autonomy using two repeated set of questions from the school questionnaire in PISA 2009 and PISA 2015. In PISA 2009 (2015) the specific question was the question 24 in the school questionnaire (question 10 in the school questionnaire) formulated as follows[16]: "Regarding your school, who has a considerable responsibility for the following tasks?" The tasks include twelve items related to school management (salaries, promotion, hiring and firing), budget management and academic management (course contents, textbooks, disciplinary policies, assessment policies, admission policies and deciding which courses are offered).[17] Principals‘ answers to these questions include five options: principals; teachers; school governing board; regional or local education authority; National education authority. The principal could tick as many answers as appropriate for each of the twelve items.

From this information we construct two variables. First, for each school we count the number of times that the principal declares that 'principals', 'teachers' or 'the school governing board' take considerable responsibility in managing the different dimensions. We consider that ticking more than one option reinforces the necessary check and balances inside the school, therefore enhancing the school autonomy. Our hypothesis is that a higher value indicates a higher autonomy. This variable (AUTONOMY) ranges from a minimum of zero to a maximum of 36. The second variable is built

[16] For more details see PISA 2009 and PISA 2015 schools' questionnaires.
[17] The number of observations and the high correlation among the autonomy dimensions prevents us from breaking down the school autonomy into different categories.

Table 12
Average school autonomy by country and school ownership.

Country & Ownership	AUTONOMY		AUTHORITIES		(AUTONOMY - AUTHORITIES)	
	PISA 2009	PISA 2015	PISA 2009	PISA 2015	PISA 2009	PISA 2015
Belgium PS	12.40	9.05	6.99	10.13	5.42	−1.08
Belgium PFPS	15.47	12.11	4.83	6.40	10.64	5.70
Czech. Rep PS	16.79	15.54	1.97	2.07	14.81	13.47
Czech. Rep PFPS	17.38	16.37	0.69	0.33	16.69	16.04
Denmark PS	15.78	15.60	4.87	4.18	10.91	11.43
Denmark PFPS	18.53	18.49	2.05	1.67	16.47	16.82
Hungary PS	16.45	10.77	2.83	6.79	13.61	3.98
Hungary PFPS	16.75	16.19	1.60	2.94	15.15	13.25
Netherlands PS	15.78	15.51	0.80	0.85	14.99	14.66
Netherlands PFPS	15.12	16.36	0.85	0.84	14.27	15.52
Slovakia PS	14.96	13.68	4.09	3.38	10.87	10.30
Slovakia PFPS	15.71	15.00	4.29	2.29	11.41	12.71
Spain PS	8.01	8.59	7.85	8.17	0.16	0.42
Spain PFPS	14.10	12.22	4.92	5.39	9.18	6.83
Ireland PS	13.92	14.57	7.16	6.48	6.76	8.09
Ireland PFPS	16.05	16.72	3.29	3.42	12.76	13.30

Table 13
Effect of the autonomy variables on CDLI and its components.

	CDLI-R09	TG-R09	CDLI-R15	TG-R15	EG
Intercept	−0.5553**	−0.5195*	−0.4944**	−0.4590**	−0.0357
	(0.2532)	(0.2686)	(0.1868)	(0.2057)	(0.0615)
Public	0.0489*	0.0683**	0.0315	0.0509*	−0.0194**
	(0.0254)	(0.0311)	(0.0195)	(0.0255)	(0.0086)
Autonomy	0.0357**	0.0325**	0.0316***	0.0285**	0.0031
	(0.0130)	(0.0139)	(0.0091)	(0.0102)	(0.0034)
Authority	0.0149	0.0134	0.0186	0.0171	0.0016
	(0.0161)	(0.0175)	(0.0118)	(0.0129)	(0.0034)
Dummy year	Yes	Yes	Yes	Yes	Yes
Countries fixed effect	Yes	Yes	Yes	Yes	Yes
N	32	32	32	32	32
F	10.34	16.60	11.91	10.37	3.85
R^2	0.7993	0.7654	0.8225	0.7799	0.6265

Each column provides the results of a different regression. The dependent variables are the CDLI, TG and EG showed in Table 7 (R09). To check robustness, we also include distances of values included in Table 9 using PS in Finland 2015 as reference (R15). EG coincides with both references. All models include a set of dummy years and countries. Robust standard errors are shown in parentheses. ***, **, and * denote 1%, 5% and 10% statistical significance thresholds, respectively.

in the same vein but, in this case, counting the number of times in which the principal declares that a ‘regional or local education authority’ or a ‘National education authority’ have a considerable responsibility of managing the different dimensions. The assumption here is that more administrations influencing schools’ decisions introduce more constraints to develop the school autonomy. Accordingly, we expect that a higher value in this variable (AUTHORITIES) corresponds to less school autonomy since decisions are partly conditioned by the public administration. This variable ranges between zero and 24.

As in Hanushek et al. (2013), it is worth noting that these two individual school dimensions of autonomy are the self-perception of the school principal’s views about the reality of the school. Table 12 presents country-level means by school ownership of the two autonomy measures and their difference for PISA 2009 and 2015.

As expected, AUTONOMY (AUTHORITIES) is higher (lower) in PFPS than in PS. Two exceptions are the Netherlands where schools’ autonomy for both school types are very similar, and Slovakia where the perceived AUTHORITY is higher in PFPS than PS in PISA 2009. School autonomy is especially low in Belgian and Spanish PS compared with the autonomy of PS in other countries. Likewise, school autonomy has significantly decreased for PFPS in these two countries from PISA 2009 to PISA 2015. Interestingly, we find the lowest autonomy levels and the highest public administration intervention in these two countries.

In Belgium and Hungary, school autonomy has declined in both school types and, at the same time, the weight of authorities on school responsibilities has increased after the six years. The pattern for the Czech Republic and the Netherlands is a combination of few differences between PS and PFPS in the two variables with a relatively large difference between the variables AUTONOMY and AUTHORITY. Table 13 shows the results of estimating Eq. (26) using OLS accounting for country fixed effects, a time effect and robust standard errors.

The regression results reveal that school autonomy has a positive and significant impact on the CDLI, particularly through the Technical Gap (TG) while there is no significant effect on the efficiency gap (EG). This result confirms that more school autonomy pushes productivity upwards. The channel might be that more school autonomy enables schools to develop educational innovation and managerial practices without significant administrative barriers, allowing some successful schools to shift up the production frontier. The rest of schools will follow the successful practices in the following periods. The non-significance of the ‘Authority’ variable suggests that the degree of responsibility of public administrations in managing school decisions is less important by itself, just in relation with school autonomy, although these two variables are correlated.

We also find that PS schools are, on average, more productive than PFPS schools due to the technology gap. However, the efficiency spread among the PS in their groups is higher than for PFPS, limiting its global productivity. In other words, inside the group of PS, we find the best performers in terms of productivity. They reach the maximum outputs from their students, subject to their level of inputs. However, for some reason, their good practices are not followed by the remaining PS and this is detrimental to the efficiency of the whole group.

6. Conclusions

In this paper, we extend the Malmquist-type index proposed by Camanho and Dyson (2006) and the pseudo-panel Malmquist index put forward by Aparicio et al. (2017) and Aparicio and Santín (2018), which are used to measure the average relative performance divergences between different groups of units within the same year and over time, respectively. The extension is related to the use of the directional distance function and the Luenberger indicator instead of the Shephard distance function and the Malmquist index.

The directional distance function encompasses the Shephard approach where non-equiproportional improvements in both inputs and outputs can be used to determine technical efficiency, whereas the Luenberger indicator can assume VRS in contrast to the Malmquist index. This latter feature implies, in the context of the performance assessment of two or more groups, that it is possible to compare the best-practice technologies of these groups of units under VRS, the most usual assumption when DEA is applied in practice and the evaluated units differ with respect to size. Additionally, the new approach inherits some interesting features from previous indexes, especially the property of circularity, using a baseline group as the reference technology, and the decomposition of the productivity gap into technical and efficiency gaps.

To illustrate this new approach, we measure how the productivity gaps between publicly funded private schools (PFPS) and public schools (PS) across eight European countries changed over the 2009–15 period using PISA data. Indeed, we think that benchmarking schools and analyzing the productivity gaps inside European Union countries is one the most promising tools for monitoring education systems and learning from best managerial practices in order to improve education in the whole European Union.

Although our findings must be analyzed with appropriate caution, they point out that the performance of PFPS was better in Belgium, Ireland, the Netherlands and Spain in both waves, while PS productivity outperforms PFPS in the Czech Republic, Hungary and Slovakia, while productivity gaps between both school types in Denmark are close to zero. Moreover, PFPS improved their results with respect to PS in six out of the eight considered countries, where Ireland and Spain are the two exceptions, probably due to public expenditure cuts during the economic crisis ongoing during this period. Another general result is that PFPS are closer to their production frontier than PS, where variance is found to be greater. Moreover, the main driver of productivity gaps changes is the technology gap.

Additionally, in line with previous research, we conclude that school autonomy has a significant positive effect on school performance. It is worth stressing that when the principals, teachers and the school governing board all share considerable responsibility for managing important school decisions related with personnel, budget and academic contents, the productivity results of schools are higher. In terms of school type, we also confirm that PS are on average more productive than PFPS. Despite PS schools being, on average, more inefficient than PFPS schools, they possess a better technical gap that eventually benefits PS productivity. Consequently, we could argue that boosting school autonomy should be a priority for policy makers in countries such as Belgium or Spain, with relative low levels of school autonomy compared to other European countries.

Our results also suggest that, beyond school autonomy, more research is still needed in order to analyze which factors characterize best performers and which educational practices and environmental variables (also known as Zs variables) are able to shift up the production frontiers from one period to another, as in Bradley et al. (2010). This information may help policymakers to foster educational policies that work, by allocating more public expenditure to the most promising alternatives. In this sense, our framework could be used for running a difference-in-differences analysis over more than two periods in those situations where there are exogenous changes, or treatments, affecting some groups but not others, that would be the control group. This would allow to load the causal inference analysis into the benchmarking methods for understanding what cause productivity gap changes in some production units or groups with respect to others.

To conclude, there are other different avenues for further follow-up research on this issue. First, it is necessary to highlight the study of the directional distance function in the context of the evaluation of two or more groups of units when there are undesirable outputs. In this sense, a recent and interesting approach for dealing with bad outputs was introduced by Murty, Russell and Levkoff (2012) and Murty and Russell (2018), defining the by-production technology. As we already mentioned, a second interesting line of research would be to checking in depth the robustness of our findings. This includes the selection of the input and output variables (Landete et al., 2017) and the relative importance of individual input and output variables (see, for example, Allen et al., 1997) adding weight restrictions to our models for incorporating experts' judgements. Third, although in this paper we have explored how different the results would be if another reference had been chosen, it is necessary to devote specific research to this issue alone in order to properly define a suitable reference. Fourth, from a theoretical and empirical perspective, it would be worthwhile analyzing within this framework how to measure the relative performance of different groups of production units over time using the Hicks-Moorsteen index (Bjurek, 1996). This index, recently used in education (Aparicio et al. 2018, Becerra-Peña & Santín, 2020), has the potential of incorporating variable returns to scale as a new scale gap component in productivity gap decomposition in problems where size and returns to scale matter, following insights from O'Donnell (2012). Fifth, the main purpose of this paper was to introduce the Luenberger indicator for performing groups comparisons. Future empirical research should incorporate more PISA, TIMSS (Trends in International Mathematics and Science Study) or PIRLS (Progress in International Reading Literacy Study) waves and countries for analyzing other educational drivers behind the evolution of productivity gaps across countries over time. Finally, we also suggest developing a bootstrap procedure to build confidence intervals for the productivity gaps and their components. This analysis will be useful for managers introducing new measures in order to test whether the groups or schools under their management significantly improve their performance with respect to other groups.

Acknowledgments

J. Aparicio and L. Ortiz thank the financial support from the Spanish Ministry of Science and Innovation and the State Research Agency under grant PID2019–105952GB-I00/ AEI / 10.13039/501100011033. D. Santín is grateful for financial support from the Spanish *Ministerio de Ciencia, Innovación y Universidades* (ECO2017–83759-P).

References

Afonso, A., & St Aubyn, M. (2005). Non-parametric approaches to education and health efficiency in OECD countries. *Journal of Applied Economics, 8*(2), 227–246.
Afonso, A., & St Aubyn, M. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic Modelling, 23*(3), 476–491.

Afsharian, M., Ahn, H., & Harms, S. G. (2019). Performance comparison of management groups under centralised management. *European Journal of Operational Research, 278*(3), 845–854.
Agasisti, T., Munda, G., & Hippe, R. (2019). Measuring the efficiency of european education systems by combining data envelopment analysis and multiple-criteria evaluation. *Journal of Productivity Analysis, 51*(2–3), 105–124.
Agasisti, T., & Zoido, P. (2018). Comparing the efficiency of schools through international benchmarking: Results from an empirical analysis of OECD PISA 2012 data. *Educational Researcher, 47*(6), 352–362.
Agrell, P. J., Mattsson, P., & Månsson, J. (2020). Impacts on efficiency of merging the Swedish district courts. *Annals of Operations Research, 288*(2), 653–679.
Allen, R., Athanassopoulos, A., Dyson, R. G., & Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of operations research, 73*, 13–34.
Althin, R. (2001). Measurement of productivity changes: Two Malmquist index approaches. *Journal of productivity Analysis, 16*(2), 107–128.
Aparicio, J., Cordero, J. M., Gonzalez, M., & Lopez-Espin, J. J. (2018). Using non-radial DEA to assess school efficiency in a cross-country perspective: An empirical analysis of OECD countries. *Omega, 79*, 9–20.
Aparicio, J., Crespo-Cebada, E., Pedraja-Chaparro, F., & Santín, D. (2017). Comparing school ownership performance using a pseudo-panel database: a malmquist–type index approach. *European Journal of Operational Research, 256*(2), 533–542.
Aparicio, J., Lopez-Torres, L., & Santín, D. (2018). Economic crisis and public education. a productivity analysis using a Hicks-Moorsteen index. *Economic Modelling, 71*, 34–44.
Aparicio, J., & Santin, D. (2018). A note on measuring group performance over time with pseudo-panels. *European Journal of Operational Research, 267*(1), 227–235.
Balk, B. (2001). Scale efficiency and productivity change. *Journal of Productivity Analysis, 15*, 159–183.
Banker, R. D., Zheng, Z. E., & Natarajan, R. (2010). DEA-based hypothesis tests for comparing two groups of decision making units. *European Journal of Operational Research, 206*(1), 231–238.
Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*, 1078–1092.
Barros, C. P., Peypoch, N., & Solonandrasana, B. (2009). Efficiency and productivity growth in hotel industry. *International Journal of Tourism Research, 11*(4), 389–402.
Battese, G., & Rao, D. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business, 1*(2), 87–93.
Battese, G., Rao, D., & O'Donnell, C. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis, 21*(1), 91–103.
Becerra-Peña, D., & Santín, D. (2020). Measuring public primary education productivity across Mexican states using a Hicks-Moorsteen index. *Applied Economics*, 1–16. https://doi.org/10.1080/00036846.2020.1819951.
Berg, S. A., Førsund, F. R., Hjalmarsson, L., & Suominen, M. (1993). Banking efficiency in the Nordic countries. *Journal of Banking & Finance, 17*(2–3), 371–388.
Berg, S. A., Førsund, F. R., & Jansen, E. S. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. *The Scandinavian Journal of Economics*, S211–S228.
Bjurek, H. (1996). The Malmquist total factor productivity index. *The Scandinavian Journal of Economics*, 303–313.
Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? *The Economic Journal, 125*(584), 647–674.
Bogetoft, P., Heinesen, E., & Tranæs, T. (2015). The efficiency of educational production: a comparison of the Nordic countries with other OECD countries. *Economic Modelling, 50*, 310–321.
Bradley, S., Johnes, G., & Millington, J. (2001). The effect of competition on the efficiency of secondary schools in England. *European Journal of Operational Research, 135*(3), 545–568.
Bradley, S., Johnes, J., & Little, A. (2010). Measurement and determinants of efficiency and productivity in the further education sector in England. *Bulletin of Economic Research, 62*(1), 1–30.
Bravo-Ureta, B. E., González-Flores, M., Greene, W., & Solís, D. (2020). Technology and technical efficiency change: evidence from a difference in differences selectivity corrected stochastic production frontier model. *American Journal of Agricultural Economics*. https://doi.org/10.1111/ajae.12112.
Brennan, S., Haelermans, C., & Ruggiero, J. (2014). Nonparametric estimation of education productivity incorporating nondiscretionary inputs with an application to dutch schools. *European Journal of Operational Research, 234*(3), 809–818.
Briec, W., & Kerstens, K. (2009a). The Luenberger productivity indicator: an economic specification leading to infeasibilities. *Economic Modelling, 26*(3), 597–600.
Briec, W., & Kerstens, K. (2009b). Infeasibility and directional distance functions with application to the determinateness of the Luenberger productivity indicator. *Journal of Optimization Theory and Applications, 141*(1), 55.
Camanho, A. S., & Dyson, R. G. (2006). Data envelopment analysis and Malmquist indices for measuring group performance. *Journal of productivity Analysis, 26*(1), 35–49.
Caves, D., Christensen, L., & Diewert, W. (1982). The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica : journal of the Econometric Society, 50*(6), 1393–1414.
Cazals, C., Florens, J., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics, 106*, 1–25.
Chambers, R. G., Chung, Y., & Färe, R. (1996a). Benefit and distance functions. *Journal of economic theory, 70*(2), 407–419.
Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and nerlovian efficiency. *Journal of Optimization Theory and Applications, 98*, 351–364.
Chambers, R. G., Färe, R., & Grosskopf, S. (1996b). Productivity growth in APEC countries. *Pacific Economic Review, 1*, 181–190.
Chambers, R. G., & Pope, R. D. (1996). Aggregate productivity measures. *American Journal of Agricultural Economics, 78*(5), 1360–1365.
Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science, 27*(6), 668–697.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*, 429–444.
Cherchye, L., De Witte, K., Ooghe, E., & Nicaise, I. (2010). Efficiency and equity in private and public education: a nonparametric comparison. *European Journal of Operational Research, 202*(2), 563–573.
Cordero, J. M., Crespo-Cebada, E., & Santín, D. (2010). Factors affecting educational attainment: evidence from Spanish PISA 2006 results. *Regional and Sectoral Economic Studies, 10*(3), 55–76.
Cordero, J. M., Cristobal, V., & Santín, D. (2018a). Causal inference on education policies: A survey of empirical studies using PISA, TIMSS and PIRLS. *Journal of Economic Surveys, 32*(3), 878–915.
Cordero, J. M., Polo, C., Santín, D., & Simancas, R. (2018b). Efficiency measurement and cross-country differences among schools: A robust conditional nonparametric analysis. *Economic Modelling, 74*, 45–60.
Cordero, J. M., Santín, D., & Simancas, R. (2017). Assessing European primary school performance through a conditional nonparametric model. *Journal of the Operational Research Society, 68*(4), 364–376.
Cordero, J. M., Crespo, E., Pedraja, F., & Santín, D. (2011). Exploring educational efficiency divergences across Spanish regions in PISA 2006. *Revista de Economia Aplicada, 57*, 117–145.
Crespo-Cebada, E., Pedraja-Chaparro, F., & Santín, D. (2014). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. *Journal of Productivity Analysis, 41*(1), 153–172.
De Jorge, J., & Santín, D. (2010). Determinantes de la eficiencia educativa en la Unión Europea. *Hacienda Pública Española, 193*, 131–155.
De Witte, K., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society, 68*(4), 339–363.
Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor inefficiency in post offices. In P. Marchand, P. Pestieau, & H. Tulkens (Eds.), *Concepts and measurements* (pp. 243–267). Amsterdam: North Holland.
Dronkers, J., & Robert, P. (2008). Differences in scholastic achievement of public, private government-dependent, and private independent schools: A cross-national analysis. *Educational Policy, 22*(4), 541–577.
Epple, D., Romano, R., & Zimmer, R. (2016). Charter schools: A survey of research on their characteristics and effectiveness. In E. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 139–208). North-Holland: Elsevier Science & Technology 5.
Essid, H., Oullette, P., & Vigeant, S. (2014). Productivity, efficiency and technical change of Tunisian schools: A bootstrapped Malmquist approach with quasi–fixed inputs. *Omega, 42*(1), 88–97.
Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1994). Productivity developments in Swedish hospitals: a Malmquist output index approach. In A. Charnes, W. W. Cooper, A. Y. Lewin, & L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology and applications*. Boston: Kluwer Academic Publishers.
Färe, R., Grosskopf, S., & Margaritis, D. (2008). Efficiency and productivity: Malmquist and more. In H. O. Fried, C. K. Lovell, S. S. Schmidt, & S. S. Schmidt (Eds.), *The measurement of productive efficiency and productivity growth* (pp. 522–621). Oxford University Press Chapter 5.
Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253–290.
Frankenberg, E., Siegel-Hawley, G., & Wang, J. (2010). *Choice without equity: Charter school segregation and the need for civil rights standards*. Civil Rights Project, UCLA http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/choice-withoutequity-2009-report/frankenberg-choices-without-equity-2010.pdf.
Frisch, R. (1936). Annual survey of general economic theory: The problem of index numbers. *Econometrica Journal of the Econometric Society, 4*, 1–38.
Ganzeboom, H., De Graaf, P., Treiman, J., & De Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), 1–56.
Giménez, V., Thieme, C., Prior, D., & Tortosa-Ausina, E. (2019). Comparing the performance of national educational systems: inequality versus achievement? *Social Indicators Research, 141*(2), 581–609.
Green, F. (2020). Private schools: Choice and effects. *The economics of education* (pp. 519–530). Academic Press.
Grifell-Tatjé, E., & Lovell, C. A. K. (1995). A note on the malmquist productivity index. *Economics Letters, 47*, 169–175.
Hanushek, E. A., Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics, 104*, 212–232.
Hoxby, C. M. (Ed.). (2003). *The economics of school choice*. Chicago: University of Chicago Press.

Johnes, G., & Virmani, S. (2020). The efficiency of private and public schools in urban and rural areas: Moving beyond the development goals. *International Transactions in Operational Research, 27*(4), 1869–1885.
Johnson, A. L., & Ruggiero, J. (2014). Nonparametric measurement of productivity and efficiency in education. *Annals of Operations Research, 221*(1), 197–210.
Juo, J. C., Fu, T. T., Yu, M. M., & Lin, Y. H. (2015). Profit-oriented productivity change. *Omega, 57*, 176–187.
Kirjavainen, T., & Loikkanen, H. A. (1998). Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit analysis. *Economics of Education Review, 17*(4), 377–394.
Landete, M., Monge, J. F., & Ruiz, J. L. (2017). Robust DEA efficiency scores: A probabilistic/combinatorial approach. *Expert Systems with Applications, 86*, 145–154.
Levin, H. M. (1974). Measuring efficiency in educational production. *Public Finance Quarterly, 2*, 3–24.
Lovell, C. A. K. (2003). The decomposition of malmquist productivity indexes. *Journal of Productivity Analysis, 20*, 437–458.
Luenberger, D. G. (1992a). New optimality principles for economic efficiency and equilibrium. *Journal of Optimization Theory and Applications, 75*(2), 221–264.
Luenberger, D. G. (1992b). Benefit functions and duality. *Journal of Mathematical Economics, 21*(5), 461–481.
Mancebón, M. J., Calero, J., Choi, A., & Ximenez, D. (2012). Efficiency of public and publicly-subsidized high schools in Spain. Evidence from PISA 2006. *Journal of Operational Research Society, 63*, 1516–1533.
Mancebón, M. J., & Muñiz, M. A. (2008). Private versus public high schools in Spain: Disentangling managerial and programme efficiencies. *Journal of the Operational Research Society, 59*(7), 892–901.
Maragos, E. K., & Despotis, D. K. (2004). Evaluating school performance over time in the frame of regional socio-economic specificities. *WSEAS Transactions on Mathematics, 3*(3), 664–670.
Murty, S., & Russell, R. R. (2018). Modeling emission-generating technologies: Reconciliation of axiomatic and by-production approaches. *Empirical Economics, 54*(1), 7–30.
Murty, S., Russell, R. R., & Levkoff, S. B. (2012). On modeling pollution-generating technologies. *Journal of Environmental Economics and Management, 64*, 117–135.
Nakano, M., & Managi, S. (2008). Regulatory reforms and productivity: An empirical analysis of the Japanese electricity industry. *Energy Policy, 36*(1), 201–209.
O'Donnell, C. J. (2012). An aggregate quantity framework for measuring and decomposing productivity change. *Journal of Productivity Analysis, 38*(3), 255–272.
OECD. (1999). *Classifying educational programmes, manual for ISCED-97 implementation in OECD countries*. Paris: Organisation for Economic Co-operation and Development.
OECD. (2014). *PISA 2012*. Paris: OECD Publishing Technical Report.
Pastor, J. T., & Lovell, C. K. (2005). A global Malmquist productivity index. *Economics Letters, 88*(2), 266–271.
Pastor, J. M., Perez, F., & Quesada, J. (1997). Efficiency analysis in banking firms: An international comparison. *European Journal of Operational Research, 98*, 395–407.
Perelman, S., & Santín, D. (2011a). Measuring educational efficiency at student level with parametric stochastic distance functions: An application to Spanish PISA results. *Education Economics, 19*(1), 29–49.
Perelman, S., & Santín, D. (2011b). Imposing monotonicity on outputs in parametric distance function estimations. *Applied Economics, 43*(30), 4651–4661.
Portela, M. C., Camanho, A. S., & Keshvari, A. (2013). Assessing the evolution of school performance and value-added: Trends over four years. *Journal of Productivity Analysis, 39*(1), 1–14.
Ray, S. C., & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: comment. *American Economic Review, 87*, 1033–1039.
Rouse, C. E., & Barrow, L. (2009). School vouchers and student achievement: Recent evidence and remaining questions. *Annual Review of Economics, 1*(1), 17–42.
Santín, D., & Sicilia, G. (2015). Measuring the efficiency of public schools in Uruguay: Main drivers and policy implications. *Latin American Economic Review, 24*(1), 1–28 5.
Segovia-Gonzalez, M. M., Dominguez, C., & Contreras, I. (2020). An assessment of the efficiency of Spanish schools: Evaluating the influence of the geographical, managerial, and socioeconomic features. *International Transactions in Operational Research, 27*(4), 1845–1868.
Shephard, R. W. (1953). *Cost and production functions*. Princeton: Princeton University Press.
Sutherland, D., Price, R., & Gonand, F. (2009). Improving public spending efficiency in primary and secondary education. *OECD Journal: Economic Studies*, (1), 1–30.
Urquiola, M. (2016). Competition among schools: Traditional public and private schools, *Handbook of the economics of education* (5, pp. 209–237). Elsevier.
Webbink, D. (2005). Causal effects in education. *Journal of Economic Surveys, 19*(4), 535–560.
Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. (2009). *School accountability, autonomy, and choice around the world*. Cheltenham, UK: Edward Elgar.