Socio-Economic Planning Sciences 81 (2022) 100927

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: http://www.elsevier.com/locate/seps

Assessing the efficiency of secondary schools: Evidence from OECD countries participating in PISA 2015

Jose Manuel Cordero *, Cristina Polo, Rosa Simancas

Department of Economics, University of Extremadura, Av. Elvas s/n, Badajoz, Spain

ARTICLE INFO

JEL Classification:
I21
H75
C14

Keywords:
Student attainment
Efficiency
Nonparametric methods
Cross-country analysis
Robust frontiers

ABSTRACT

One of the most important issues accompanying the publication of the main results of the well-known Programme for International Student Assessment (PISA) is the classification of countries according to student attainment. However, this ranking does not take into account some highly relevant factors, such as the different resource endowments of each education system or the heterogeneous context in which schools operate. This study aims to provide a fuller picture of education system operation worldwide by assessing the managerial efficiency of secondary schools in a cross-country framework. To do this, we use data from OECD countries participating in PISA 2015 and apply a robust nonparametric approach that accounts for the fact that schools were operating under heterogeneous conditions before the efficiency measures of performance were estimated. Our results suggest that the consideration of both school resources and environmental factors significantly modifies the country ranking based solely on student results.

1. Introduction

The increasing development of international large-scale assessments like PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study) or PIRLS (Progress in International Reading Literacy Study) has changed the landscape of educational research. These studies produce data used by researchers to analyse between- and within-country achievement differences, as well as to investigate the potential effects of multiple educational and societal factors on educational performance. As a result, comparative educational studies are increasingly popular in education sciences today, also attracting considerable media attention and having a profound influence on educational policy design [1].

Since Woessmann [2]published his seminal research, multiple studies have adopted a cross-country approach to explore the major factors influencing educational achievement from different perspectives [3,4]. Most of these studies use econometric techniques to identify links between student background, school-related variables and educational outcomes (typically represented by test scores).[1] However, exploring how resources are used is another key issue in science and technology management [5]. Indeed, the measurement of education system efficiency is now a burning issue among educational stakeholders because most countries now face constraints on public spending on education caused by shortages of resources raised by taxation. Therefore, the development of guidelines to help schools improve academic outcomes taking into account school factors is a key concern of both policy makers and researchers.

This study proposes an international comparison of education managerial efficiency using cross-country data on secondary schools from different countries participating in PISA 2015. Note, in this respect, that international comparisons are extremely challenging since schools operate in very different contexts. Some studies (e.g. [6,7] have addressed this problem by limiting the comparison group to similar countries. By contrast, the dataset used in this paper includes a large sample of schools belonging to 35 OECD countries with very heterogeneous conditions. Therefore, we need to take into account data about the diverse educational environments in which schools operate to reliably estimate the efficiency measures of school performance. In this way, each unit can be benchmarked against other units from different countries provided that they operate in a similar environment.

To do this, we could resort to the most recent developments of the nonparametric conditional frontier literature [8–10]. These authors

* Corresponding author.
E-mail address: jmcordero@unex.es (J.M. Cordero).

[1] More recently, some have started to apply more sophisticated methods in order to identify causal relationships in the international data on educational achievement (see Ref. [53]; for a review).

https://doi.org/10.1016/j.seps.2020.100927
Received 29 December 2019; Received in revised form 21 June 2020; Accepted 24 July 2020
Available online 11 August 2020
0038-0121/

extended the probabilistic formulation of the production process proposed by Cazals et al. [11] to account for the potential influence of heterogeneous contextual factors without assuming the restrictive separability condition, thus contextual variables might affect both the distribution of the efficiencies and the boundary of the attainable set [12].[2] To enact this procedure, smoothing techniques have to be applied to the environmental variables. This requires the selection of optimal smoothing parameters (bandwidths) like the ones proposed by Badin et al. [13] or, more recently, by Badin et [14]. Once conditional efficiency scores have been estimated, it is possible to investigate the direction of their effect (favourable or unfavourable) on the production process by examining the ratio between those conditional measures and the unconditional ones, i.e. without taking contextual variables into account [8,10]. Furthermore, Badin et al. [15] suggest using a two-stage flexible location-scale regression model to discount the effect of contextual conditions and produce rankings of units according to their managerial efficiency, i.e. their intrinsic ability to use their available resources after removing the effects associated with the presence of the environmental conditions in which they are operating. This results in cleaned or pure efficiency scores.

However, this approach may not be a good option if the number of contextual variables we intend to include in the model is very high, because the estimated efficiency scores might be struck by the so-called *"curse of dimensionality"*. In this paper, we circumvent this potential problem by applying the flexible location-scale model suggested by Florens et al. [16]. This model can be interpreted inversely to the above two-stage method. Thus, we first use a nonparametric regression model to eliminate the dependence of production inputs/outputs on external factors (Z) and we then estimate the frontier and the efficiencies of the units using pure inputs and outputs, i.e., whitened from the influence of Z. Those measures of "*pure or managerial efficiency*" are more reliable for producing international country or education system rankings or benchmarks, since they are assuming that all schools participating in PISA are operating in a similar environment.

Apart from estimating those cleaned efficiency measures and derive country ranking according to their values, we are also interested in investigating the influence of contextual variables on the performance of the schools under analysis so that we can test whether their impact on their efficiency is similar to the effect on students' attainment. Therefore, in a later analysis, we rely on the ratio analysis that we explained above. In order to implement it we need to estimate both unconditional and conditional measures of performance (different from the previous ones) using the original values of the inputs and the outputs.

The remainder of the paper is structured as follows. Section 2 reviews previous literature on cross-country studies using data from international large-scale assessments and applying frontier techniques to estimate efficiency measures of performance. Section 3 describes the methodology applied. Section 4 explains the key characteristics of the data and the variables selected to conduct our empirical analysis. Section 5 discusses the main results compared with previous literature. Finally, Section 6 outlines some concluding remarks.

2. Literature review

Frontier techniques have increasingly been used to estimate efficiency measures of performance in the education sector worldwide over the last decades.[3] Until quite recently, most empirical studies focused on analysing the performance of schools from one and the same country using national databases provided by each country's ministry of education or equivalent agencies [17]. Lately, however, researchers have access to comparable data from large-scale international assessments like PISA, TIMSS or PIRLS to conduct cross-country studies. This branch of research is a good monitor of efficiency differences across countries and the factors that influence education system performance. Table 1 summarizes a significant number of empirical studies that have been published in this rather young field of research focused on efficiency assessment in a cross-country framework classified according to the database used and the level of analysis (country, school or student).

As we can see, the OECD PISA database is the most popular source of information for most researchers conducting this type of studies. With regard to the analysis level, we find that very few researchers use micro data, i.e. consider the student as their unit of analysis. A possible explanation is that this approach fails to lead to results of use for educational policy decision making. In contrast, there are a large number of studies using country-level data, focusing primarily on identifying and understanding the differences in average efficiencies. Such papers can be rather uninformative, because they are unable to identify the source of potential inefficiency behaviours or the units within the educational system (regions, districts or schools) whose performance is good or bad. Thus, they are likely to include a blend of very wide-ranging behaviours that this approach fails to capture. Therefore, a more reasonable approach would be to use data about schools to assess the performance of education systems worldwide.

Most of the above studies use nonparametric techniques, like DEA or FDH, to estimate performance efficiency measures. These techniques are flexible enough to adapt to the characteristics of public services provision, and especially to the fact that such services have multiple inputs and outputs.[4] Moreover, a two-stage procedure is also applied in many cases to examine the potential influence of contextual variables on efficiency estimates (e.g., Refs. [7,18,23,24,26,29,32]. This procedure is troublesome primarily because it assumes the restrictive separability condition, i.e. the contextual variables might affect the shape of the distribution of inefficiencies (i.e., mean, variance, etc.) but not the attainable set or the estimated frontier (see Ref. [37,38]; for details). This assumption is often unrealistic, since contextual factors can be expected to influence both educational outcomes and resources and therefore also the shape of the frontier. Although the statistical tools

Table 1
Summary of previous studies using data from international large-scale in an efficiency assessment and adopting a cross-country approach.

	PISA	TIMSS/PIRLS
Student level	De jorge and Santin [18] Deutsch et al. [20]	Cordero et al. [19]
School level	Wilson [21] Sutherland et al. [22] Agasisti and Zoido [23] Aparicio et al. [24] Cordero et al. [25] Agasisti and Zoido [7]	Cordero et al. [19]
Country level	Afonso and St Aubyn [26] Verhoeven et al. [29] Giambona et al. [30] Thieme et al. [31] Agasisti [32] Aristovnik and Obadić [33] Coco and Lagravinese [34] Bogetoft rt al [35]. Giménez et al.[36]	Clements [27] Giménez et al. [28]

Source: Own elaboration.

[2] This approach has been extensively applied in multiple empirical studies conducted in different frameworks, in which assuming that environmental factors are not affecting the shape of the efficient frontier is quite unrealistic. Some examples can be found in Roudaut and Vanhems [57], Kourtzidis et al. [58], Mastromacro et al. [59] or Tzeremes [60].

[3] See Worthington [61] for an early review of this literature, and Johnes [62] for an updated revision.

[4] There are some exceptions using parametric methods (e.g., Ref. [20,22]).

developed by Daraio et al. [12] can be used to test the separability between the input–output space and the space of external variables in advance, none of the above empirical studies ran this test before applying this method.

If the two-stage procedure is found to be inappropriate, an alternative is to use the conditional nonparametric approach developed by Refs. [8–10]. Using this approach, the estimation of efficiency scores can account for the effect of contextual factors without assuming the aforementioned separability condition. To the best of our knowledge, only two empirical studies have applied this method. Cordero et al. [19] analysed the effect of several contextual factors at school and country level on the performance of primary schools from 16 European countries participating in PIRLS 2011. They decomposed the estimated inefficiency levels between two different sources (school and country) using the metafrontier framework [39]. Likewise, Cordero et al. [25] assessed the performance of a set of more than 12,000 secondary schools from 36 countries participating in PISA 2012. They explored the influence of a wide range of contextual factors, including variables at both school and country level. Moreover, they also obtain "whitened or cleaned" efficiency scores by applying the second-stage approach suggested by Badin et al. [15] to eliminate the effects of contextual conditions. Cleaning was the final step in the method applied by Cordero et al. [25], whereas in the present work we estimate the frontier and the managerial efficiencies of the units using pre-whitened inputs free from the influence of environmental variables. This has some advantages over the method proposed by Badin et al. [15], as explained in Section 3.

3. Methodology

Starting with the usual production process in which a set of inputs $X \in \mathbf{R}_+^p$ produces a set of outputs $Y \in \mathbf{R}_+^q$, the attainable set of feasible combinations of inputs and outputs can be defined by.

$$\Psi = \left\{(x, y) \in R_+^{p+q} | x \text{ can produce } y\right\} \tag{1}$$

Following Cazals et al. [11], this production function can be defined using a probabilistic formulation as

$$H_{X,Y}\left(X, y = prob\left(X \leq x, Y \geq y\right)\right. \tag{2}$$

Note that school performance might be affected by the presence of contextual or exogenous factors that are beyond their control. Thus, the efficiency analysis should account for the school environment in order to assure a fair comparison among units. These factors can be included as a set of contextual variables influencing both the input-output space (i.e., the frontier) and the distribution of the efficiencies. Following Daraio and Simar [8], [10], the above variables can be included in the joint distribution of (X, Y) conditional on $Z = z$ as

$$H_{X,Y|Z}(x, y|z) = Prob(X \leq x, Y \geq y|Z = z) \tag{3}$$

For an output conditional measure of efficiency, this function can be decomposed into two terms: the survival conditional function of outputs and the conditional distribution function of inputs, as follows:

$$H_{X,Y|Z}(x, y|z) = S_{Y|X,Z}\left(y|x, z\right)F_{X|Z}\left(x|z\right) \tag{4}$$

Thus, the conditional output-oriented efficiency measure can be defined as the proportionate increase in outputs required for the evaluated unit to have a zero probability of being dominated at the given input level and by other units facing the same environmental conditions z:

$$\lambda(x, y|z) = sup\left\{\lambda > 0 \middle| H_{X,Y|Z}(x, \lambda y|z) > 0\right\} = sup\left\{\lambda > 0 \middle| S_{Y|X,Z}(\lambda y|x, z) > 0\right\} \tag{5}$$

In terms of the Farrell-Debreu efficiency scores, a value equal to one indicates that the unit is fully efficient, whereas units with values greater than one are considered to be inefficient (higher values imply a longer distance to the frontier representing more inefficiency). We can use a plug-in rule to define the different conditional estimators of either the full frontier, such as the Free Disposal Hull (FDH) and Data Envelopment Analysis (DEA),[5] or partial frontiers, like order-m and order-α.

This is a well-established approach in the literature. However, it has some disadvantages when we are interested in building school or country rankings, which is a major objective of this study. Particularly, the estimators defined according to the above procedure have to be smoothed at the frontier, where there are fewer units than inside the data cloud. As a result, measures are more sensitive to outliers and extreme data. On the other hand, the resulting efficiency measures could be subject to endogeneity bias caused by reverse causality between external factors and input/output indicators. Besides, the dimension of Z could leave this conditional model open to the well-known curse of dimensionality problem.[6]

We address the above issues by adopting the methodology proposed by Florens et al. [16]. These authors assume flexible nonparametric location-scale models that link the input-output space to the contextual variables. They then remove the dependence on Z in the original inputs and outputs to generate two new sets of whitened variables. This approach has several points in common with the four-stage model proposed by Fried et al. [40], which has been applied extensively to evaluate managerial efficiency in different frameworks,[7] although it includes some improvements to deal with some its major weaknesses.[8] Specifically, we adopt the following nonparametric location-scale regression model:

$$\begin{cases} X_i = \mu_x(Z_i) + \sigma_x(Z_i)\varepsilon_{x,i} \\ Y_i = \mu_y(Z_i) + \sigma_y(Z_i)\varepsilon_{y,i} \end{cases} \tag{6}$$

The first equation in (6) includes p relations, one for each component of X, whereas the second one integrates q components as the dimension of Y. Likewise, μ_x, σ_x and $\varepsilon_{x,i}$ each have p relations, and μ_y, σ_y and $\varepsilon_{y,i}$ each have q relations. All vector products are understood component wise. The terms associated with the residuals in this model (i.e. ε_x, ε_y) are considered as the cleaned inputs and outputs, and they have mean zero and a standard deviation equal to 1. The vectors μ_x and μ_y stand for the locations, whereas σ_x and σ_y capture the scale effects, all of which are conditioned to Zs.

As pointed out in Florens et al. [16], the above mathematical specification of the methodology has some very important features. First, model construction respects its nonparametricity, since no hypothesis is assumed for the distributions of ε_x and ε_y. Furthermore, the procedure removes any dependence between the exogenous variables and vectors. Therefore, they can be accepted as the whitened version of X and $Y\varepsilon_x\varepsilon_y$. Additionally, we can derive from this well-defined model a pure efficiency measure that appears to be more robust than previous conditional nonparametric techniques that account for the impact of external factors. This is because smoothing takes place in the centre of the sample where there are more observations than on the boundary (see Ref. [15] for details).

From Model (6), we can estimate the location and scale vectors by running a double nonparametric regression. The first regression estimates the location functions $[\widehat{\mu}_x(Z_i),\ \widehat{\mu}_y(Z_i)]$ based on a local linear

[5] While the conditional efficiency estimator of the FDH frontier was developed in Daraio and Simar [8], the corresponding convex technology, i.e., the DEA estimator was established in Daraio and Simar [10].

[6] Florens et al. [16] indicate that smoothing Z to get the different nonparametric estimators, whereby n is to be replaced by $n\prod_{j=1}^{d_z} h_j$ (h_j being the corresponding bandwidth for each unit) when product kernels are used for smoothing the d_z components of Z (see Ref. [63] for details), degrades convergence rates.

[7] For example, see Wang [64] or Shyu and Chiang (2002) [65]

[8] See Cordero-Ferrera et al. [66] for details.

model. The square residuals resulting from this first regression are used to estimate the scale functions $[\widehat{\sigma}_x^2(Z_i), \widehat{\sigma}_x^2(Z_i)]$ from a subsequent local constant model to assure positive values of the variances. Both regressions use local polynomial estimators of degree p that are incorporated into a local minimization problem (see pp. 459–460 in Ref. [16] for a detailed description of the nonparametric estimation of the functions μ and σ). Thus, we can derive the residuals:

$$\widehat{\varepsilon}_{x,i} = \frac{X_i - \widehat{\mu}_x(Z_i)}{\widehat{\sigma}_x(Z_i)} \tag{7}$$

$$\widehat{\varepsilon}_{y,i} = \frac{Y_i - \widehat{\mu}_y(Z_i)}{\widehat{\sigma}_y(Z_i)} \tag{8}$$

These expressions represent the inputs and outputs for which the influence of the exogenous variables has been removed.[9] According to Mastromarco and Simar [41], this model asymptotically verifies the assumption of independence among $\widehat{\varepsilon}_x$, $\widehat{\varepsilon}_y$ and Zs since $Cov(X_i, \widehat{\varepsilon}_{x,i}) \rightarrow 0$ and $Cov(Y_i, \widehat{\varepsilon}_{y,i}) \rightarrow 0$ as $N \rightarrow \infty$. Input and output cleaning (i.e., removing the effect of the Zs) mitigates two of the abovementioned limitations of conditional estimators, since this approach is less sensitive to the curse of the dimensionality caused by the size of Z and eliminates the dependence of production inputs/outputs on common contextual factors, reducing the potential problem of endogeneity that might arise due to the existence of reverse causality in the production process[10]. The following step would be to estimate a pure measure of efficiency by using the whitened estimations ε_x and ε_y. To do this, we must transform Equation (1) to define the attainable set of ε_x and ε_y as follows:

$$\Psi_\varepsilon = \left\{ (e_x, e_y) \in \mathbb{R}^{p+q} \middle| H_{\varepsilon_x,\varepsilon_y}(e_x, e_y) = Prob(\varepsilon_x \le e_x, \varepsilon_y \ge e_y) > 0 \right\} \tag{9}$$

The different efficiency estimators, like DEA or FDH, can be derived by replacing the empirical counterparts of $\widehat{H}_{\varepsilon_x,\varepsilon_y}(e_x, e_y)$. Nonetheless, since pure inputs and outputs have mean zero, the measures based on radial distances from each observation to the efficient frontier are inappropriate. Directional distance functions would appear to be better suited in such a scenario.[11] Therefore, we apply:

$$\delta(e_x, e_y; d_x, d_y) = sup\left\{ \gamma \middle| H_{\varepsilon_x,\varepsilon_y}(e_x - \gamma d_x, e_y + \gamma d_y) > 0 \right\} \tag{10}$$

The terms d_x, d_y in (10) represent the desired direction of the projection over the efficient frontier. In this case, we select the output direction ($d_x = 0$ and $d_y = 1$) since schools are supposed to be interested in improving their performance by achieving better student results and not by reducing their use of resources. Then, the nonparametric pure efficiency estimator in the output direction can be obtained as:

$$\widehat{\delta}(\widehat{\varepsilon}_{x,i}, \widehat{\varepsilon}_{y,i};\ \ 0,\ \ 1) = \widehat{\phi}(\widehat{\varepsilon}_{x,i}, \widehat{\varepsilon}_{y,i}) - \widehat{\varepsilon}_{y,i} \tag{11}$$

where $\widehat{\phi}(e_x, e_y) = \max\limits_{\{i | \widehat{\varepsilon}_{x,i} \le e_x\}} \left\{ \min\limits_{j=1,\dots,q} \left(\frac{\widehat{\varepsilon}_{y,i}^{\,j}}{e_y^j} \right) \right\}$ is clearly identified as the output-oriented FDH estimator of the pure efficient frontier.

The efficiency derived from this estimator could be affected by outliers or extreme data within the sample. To avoid this problem, robust versions of the proposed estimator have been developed in the literature. One of the most used approaches is the order-m estimator, which uses just a set of m observations randomly drawn from the population of units using fewer inputs than x to define a partial frontier. Following Florens et al. [16], the pure version of this estimator is the expected value of the maximum output set with length m such that $\varepsilon_{x,i} \le e_x$:

$$\widehat{\phi}_m(e_x, e_y) = \widehat{E}\left[\max_{i=1,\dots,m} \left\{ \min_{j=1,\dots,q} \left(\frac{\widehat{\varepsilon}_{y,i}^{\,j}}{e_y^j} \right) \right\} \right] \tag{12}$$

We rely on these measures of pure efficiency for both the full and partial frontier estimators in order to mitigate potential problems related to both endogeneity bias and the dimensionality problem in traditional conditional models. This is possible because the influence of contextual variables has been removed from the original variables and hence from the estimations.

Although the main interest of our work focuses on obtaining representative measures of managerial school efficiency separated from the influence of external conditions, we are also interested in exploring whether school contextual variables affecting school performance are similar to those identified in previous literature as determinants of student attainment. In order to assess the impact of those exogenous variables on efficiencies, we need to resort to the conditional and unconditional efficiency measures (instead of managerial efficiencies) estimated using the original values of the inputs and the outputs, i.e., X and Y. In particular, we used the unconditional and conditional efficiency measures from robust output-oriented order-m estimators [10], which can be defined, respectively, as:

$$\widehat{\lambda}_m\left(x, y\right) = \int_0^\infty \left[1 - \left(1 - \widehat{S}_{Y|X}(uy|X \le x)\right)^m\right] du \tag{13}$$

$$\widehat{\lambda}_m(x, y|z) = \int_0^\infty \left[1 - \left(1 - \widehat{S}_{Y|X,Z}(uy|X \le x, Z = z)\right)^m\right] du \tag{14}$$

Note that the frontier built with these estimators will not envelope all observations, thus there might be efficiency scores with a value less than1.[12] In addition, the empirical survival functions, $\widehat{S}_{Y|X}$ and $\widehat{S}_{Y|X,Z}$, implicitly assume the estimation of bandwidths parameters that determine the relevance of each exogenous variable in the analysis.[13]

Once we have estimated those unconditional and conditional efficiency measures, we can explore the effect of Zs on efficiency following the procedure proposed in Badin et al., [15],[14] i.e. investigating the ratio between the conditional and the unconditional order-m efficiency measures:

$$\widehat{R}_m(x, y|z) = \frac{\widehat{\lambda}_m(x, y|z)}{\widehat{\lambda}_m(x, y)} \tag{15}$$

In particular, we focus on the impact of contextual factors on the attainable set using measures relative to the full frontier of the conditional and the unconditional attainable sets, that is, those represented by extreme order-m measures ($m \rightarrow \infty$).[15] Since the orientation selected is to maximize the outputs, most of the ratios are below 1, which is where the

[9] Florens et al. [16] propose a bootstrap-based procedure to test the independence between the whitened inputs and outputs and the Zs. In this paper, we report evidence of independence as in Mastromarco and Simar [51].

[10] We are aware that there might be still a certain problem of endogeneity due to the presence of unobserved heterogeneity. However, the possibility of correcting this using other methodological approaches (e.g. Ref. [67] or [68] is beyond the scope of this paper.

[11] For a more detailed explanation of directional distance functions (DDF), see Färe and Grosskopf [42], Simar and Vanhems [69], Daraio and Simar [70]: [71] or Daraio et al. [72].

[12] These units are classified as superefficient and are below the frontier built by the m units with which they were compared.

[13] To obtain the bandwidths, we have resorted to the data-driven selection method proposed by Badin et al. [13], based on the least squares cross validation (LSCV) procedure developed by Li and Racine [73]. This approach has the appealing feature of detecting the irrelevant factors and smoothing them out by providing them with large bandwidth parameters (see Refs. [74,75] or [76] for details).

[14] Mastromacro and Simar [51]: [41] also apply this procedure after using the pre-whitening approach suggested by Florens et al. [16]. They extend the method to a dynamic framework in which time (for a panel database) plays a role as an additional exogenous variable.

[15] In practice, this is equivalent to $m = n$ (total number of units).

marginal and the conditional frontiers meet. Nonetheless, as we build partial frontier estimators, the ratios are not bounded, and values above 1 are also possible.

The interpretation of the direction of the impact produced by *Zs* in an output-oriented model is as follows. An upward trend of the ratio when the conditioning variables increase would indicate a favourable effect (the conditional frontier approximates the marginal frontier, where the variables act as freely available inputs), whereas a downward trend when the variables increase denotes an unfavourable impact (the conditional frontier moves away from the unconditional frontier, where the *Zs* act as undesirable outputs).

4. Data and variables

The data used in this study come from the Programme for International Student Assessment (PISA) designed by the OECD in the late 1990s as a comparative, international and continuous study of specified characteristics and skills of students aged between 15 and 16 years. PISA assesses their performance in three core competencies: mathematics, reading and science.[16] PISA also gathers information about students' background, school environment or educational provision (factors potentially related to student performance) through separate questionnaires addressing students, parents, teachers and school principals. The first PISA survey took place in the year 2000. Since then, it has been repeated every three years. Although the three core competencies mentioned above are always assessed, each wave focuses on one domain. In this study we use data from PISA 2015, with a total of 72 participating countries (35 OECD members and 31 partners), where science was the main domain [43].

This survey implements a two-stage stratified sampling design [44]. In the first stage, schools are randomly selected from the population of all schools with 15-year-old students enrolled, where the probability of being selected is proportional to school size. A minimum of 150 schools are selected in each country. In the second stage, once the participating schools have been determined, a total of 42 students are randomly selected from each school [45]. This kind of sampling design can influence accuracy since similarities between students within schools are higher than between students across schools (intra-class correlation).[17]

PISA assesses student performance in mathematics, reading and science on a continuous scale with a mean of 500 and a standard deviation of 100. For each student and domain, ten plausible values are computed. Plausible values are randomly drawn from their distribution of results, which are estimated by means of Item Response Theory [46]. These ten plausible values are regarded as a representation of the range of student abilities [45].[18]

In this study, we concentrate on 35 OECD participating countries. Our dataset comprises a total number of 9,359 schools distributed across countries as shown in Table 2. As explained above, the minimum number of participating schools in each country is 150. However, our sample contains some exceptional cases with a lower number of observations because the number of schools in the country is limited (e.g., Luxembourg). Likewise, the sample is very large in several countries because it includes representative samples for different regions within the country (e.g. Australia, Canada, Italy or the United Kingdom).

As output indicators, we use the student results in the three competencies evaluated in PISA 2015 aggregated at school level. Specifically, the ten plausible values for each subject (PVMATH, PVREAD and PVSCIE) have been taken into account by applying replicate weights using the REPEST command provided by STATA® [47].

Table 2
Dataset composition: number of schools per country.

Country	Schools	Country	Schools
Australia	758	Korea	168
Austria	269	Latvia	250
Belgium	288	Luxembourg	44
Canada	759	Mexico	275
Chile	227	Netherlands	186
Czech Republic	344	New Zealand	183
Denmark	333	Norway	229
Estonia	206	Poland	169
Finland	168	Portugal	246
France	252	Slovak Republic	290
Germany	256	Slovenia	333
Greece	211	Spain	201
Hungary	245	Sweden	202
Iceland	124	Switzerland	227
Ireland	167	Turkey	187
Israel	173	United Kingdom	550
Italy	474	United States	177
Japan	198	**TOTAL**	**9,369**

One of the main challenges is the selection of the variables to be included in the empirical analysis due to the long list of potential indicators that international large-scale assessments usually provide. As inputs, we selected three variables that are in line with most previous empirical studies measuring school efficiency (see Refs. [17] for a recent literature review). Additionally, these inputs are generally accepted as objective measures of resources involved in the educational process and meet the requirement of monotonicity (inputs should be positively correlated with outputs). First, we select one variable representing human resources available such as the number of teachers per (hundred) students (TSRATIO), that is, the inverse of the student-teacher ratio provided by PISA. Second, we use an index representing the quality of school educational resources (SCHRES) computed as the inverse of EDUSHORT (the original PISA index that measures principals' perceptions of potential factors hindering the provision of instruction at school). Finally, the third input selected is the average economic, social and cultural status (ESCS) of students in the school,[19] since students are the "raw material" to be transformed through the learning process. Using this variable as an input is a common practice in several recent papers attempting to measure the efficiency of schools (e.g. Ref. [7,23,24,32,48–50]. Since the original values of SCHRES and ESCS presented positive and negative values, all of them were rescaled to show positive values[20].

Finally, we also selected several continuous and dummy variables representing the educational environment in which schools operate, as well as factors associated with the type of school management. Specifically, we included the proportion of fully certified teachers with respect to the total number of teachers (PROPCERT), the total number of students enrolled per school (SCHSIZE), the level of responsibility that school staff have in allocating resources (RESPRES), the percentage of students who have repeated at least one grade (%REPEAT) and the percentage of students who skipped a whole school day (%TRUANCY).

[16] The most recent waves of the PISA survey also evaluate other innovative skills, such as collaborative problem solving or financial literacy.

[17] If 42 students within a school are selected, they do not provide as much "information" as 42 students randomly selected from all schools [77].

[18] More detailed information about plausible values can be found in Mislevy et al. [78] or Wu [79].

[19] This variable is derived from an index created by PISA including the highest educational level of either of the student's parents, the highest labour occupation of either of the student's parents and an index of educational possessions related to household economy.

[20] The rescaling process was made by adding the minimum value to all the original values of the variables. This transformation does not alter the efficient frontier (or empirical production function) and hence the associated DEA model is translation invariant.

As dummy variables, we considered whether the school is located in a rural area, i.e., town with fewer than 15,000 inhabitants (RURAL), the type of school ownership (PRIVATE),[21] classroom size, i.e., if the school has fewer than 20 students per class (SMCLASS), school accountability policies, i.e., whether the school reports student average achievement publicly (ACCOUNT) and, finally, class distribution, i.e., if the school divides its students into classes based on their ability (ABGROUP). The main descriptive statistics for all these variables are summarized in Table 3.

5. Results

This section shows the results of applying the methodologies described in Section 3. First of all, we eliminate the dependence of production inputs/outputs on external factors (Z) by means of the nonparametric location-scale model. We use the model described in (6) and Equations (7) and (8) to derive the cleaned values for these new pure inputs and outputs. Table 4 shows the main descriptive statistics for these residuals. Note that, although there are some divergences in the dispersions of the new variables, they all contain both positive and negative values, confirming the predicted results for Model (6).

Before estimating the pure or managerial efficiencies from the new cleaned variables, we checked whether the whitening was performed correctly. Following Mastromarco and Simar [51], we calculated the Pearson and Spearman rank correlations coefficients between the vectors $\hat{\varepsilon}_x, \hat{\varepsilon}_y$ and the Zs, which are reported in Table 5. As expected, we found both positive and negative values, but the most important point is that they are all very low. Thus, we can conclude that the effect of the contextual variables has been totally removed from the original data. In this way, we mitigate the potential endogeneity bias and dimensionality problems occurring if contextual variables are included in a nonparametric conditional approach. Therefore, we can construct a ranking of countries according to their managerial efficiency, i.e. their intrinsic ability to use their available resources in a hypothetical scenario where all observed units are considered to be operating in the same context, regardless of the educational system to which they belong.

As detailed in Section 3, once we have whitened input and output values, we estimate both FDH and order-m pure efficiencies. Following Daraio and Simar [8], we determine the size of the partial frontier as the value of m at which the decrease in the number of superefficient observations stabilizes. In our application, we selected a value of $m = \sqrt[3]{N^2}$, following the rule of thumb suggested by Tauchmann [52]. This value corresponds to approximately 15% of the total number of schools. This implies that each school is compared to approximately 1,400 schools randomly drawn from observations in the whole dataset that consume at most the same amount of inputs. For statistical inference, we use 200 bootstrap replications. Fig. 1 shows the two histograms of the distributions of the pure inefficiencies corresponding to both frontiers.

For the full frontier, we observe a potential problem of dimensionality, since the proportion of efficient schools is very high (5,664 out of 9,369). In contrast, the distribution of the inefficiencies in the case of the order-m is much more dispersed. The discriminative power of this approach is greater, enabling us to better identify the best and worst performers among schools assuming that they are operating in the same environment. Specifically, we find 450 schools (around 4.8%) that are fully efficient and another 717 schools (around 7.65%) that are superefficient (below 1), i.e., they perform better than the 15% schools against which they are benchmarked.

Table 6 reports the ranking of countries according to the average estimated managerial efficiency order-m scores of their schools and the corresponding standard deviation.[22] Additionally, we also show the classification of countries according to the mean values of their science results in order to explore whether or not this classification changes taking into account inputs and contextual factors in our efficiency assessment.[23] The top performers in terms of efficiency include several countries that also have high average test scores for science, such as Japan, Korea, Canada, New Zealand or Poland, but also cover other nations that have poor results in that competence, like Slovenia, Greece or Israel. Our interpretation of this striking result is that schools in these countries are making the most in terms of academic achievement from students with lower socioeconomic status and using less educational resources. These results are in line with those obtained in other cross-country studies focused on measuring school efficiency with different approaches. For instance, in Cordero et al. [19,53] several countries with poor results in reading and mathematics (e.g. Romania, Mexico or Colombia) are ranked among the top performers according to efficiency measures estimated with a nonparametric conditional approach. Similarly, Aparicio et al. [24] and Agasisti and Zoido [7] also find that Turkish schools present high DEA efficiency scores despite having relatively low average test scores in mathematics.

The opposite applies to other countries with good results for science, which, however, are ranked in the middle of the classification based on managerial efficiency (e.g., Estonia or Finland) or even among the worst performers according to this indicator (Netherlands and Germany). This is because schools from these countries have not been able to achieve as good results (measured in terms of test scores) as could be expected from the higher level of quality and volume of their resources in relative terms. This is also a common finding in previous empirical studies with a similar approach to ours. Thus, for example, in) Gimenez et al. [28] and Agasisti [32] several developed countries with relatively high test scores in international large-scale assessments appear to be inefficient when resources are taken into account in an efficiency assessment.

To supplement the information regarding the average values recorded by the schools belonging to each education system, Table 7 shows the percentage of efficient (and superefficient) units found in the subsample of units belonging to each country. Here we observe that the top-performing countries are (more or less) the same countries as in Table 6, although there are some slight changes in their classification. For instance, Korea, Israel, Greece and Poland improve their position relatively, while Slovenia and Canada are ranked relatively lower. If we look at the countries with the lowest percentages, it is striking that Germany is placed last with less than 4% of schools being located on the frontier (or beyond). Likewise, we were also surprised to find Finland among the lowest ranked countries according to this criterion, bearing in mind that this education system has been considered as a benchmark in most comparative studies. Again, we believe that the explanation for these results lies in the fact that the schools belonging to these education systems have relatively better students and more resources. Therefore, managerial efficiencies estimated assuming equal conditions across schools in all OECD countries suggest that schools from these countries do not perform as well as might be expected taking into account the results achieved by their students.

The next step in our analysis was to analyse how contextual factors affect efficiency estimates. To do this, we estimated unconditional and conditional efficiency scores with the original levels of inputs and outputs. Table 8 summarizes the main descriptive statistics (mean, standard

[21] The term "private school" include both schools privately managed and funded and private-government dependent schools (privately managed but government funded).

[22] We report the classification using the order-m estimations only because they are more robust and have a greater discriminative power. The original values have been transformed into values between 0 and 1 in order to facilitate their interpretation (higher values indicate higher levels of efficiency).

[23] For comparative reasons, we use only average results for science, as this is the main competence assessed in PISA 2015 [43]. In any case, the three competencies assessed (science, mathematics, and reading) are highly correlated with each other.

Table 3
Descriptive statistics of variables included in the analysis.

Variables		Mean	SD	Min	1Q	Median	3Q	Max
Outputs	**PVMATH**	484.71	60.55	254.86	449.04	488.77	524.06	680.22
	PVREAD	486.26	64.14	226.81	447.81	491.92	530.56	678.87
	PVSCIE	488.63	63.60	265.69	447.80	493.29	532.83	697.87
Inputs	**ESCS**	4.82	0.66	0.80	4.47	4.88	5.27	6.44
	TSRATIO	9.87	8.70	1.00	6.69	8.28	10.24	100.00
	SCHRES	3.72	0.96	0.02	3.17	3.78	4.50	4.97
Continuous	**PROPCERT**	0.85	0.27	0.00	0.81	0.97	1.00	1.00
	SCHSIZE	695.56	472.33	2.00	329.00	636.00	925.00	2,490
	RESPRES	0.82	1.02	0.01	0.25	0.46	0.81	3.62
	%REPEAT	0.37	0.41	0.00	0.00	0.17	0.92	1.00
	%TRUANCY	0.27	0.23	0.00	0.10	0.22	0.38	1.00
		% Level 0	**% Level 1**					
Dummies	**RURAL**	70.06	29.94					
	PRIVATE	85.04	14.96					
	SMCLASS	76.84	23.16					
	ACCOUNT	60.76	39.24					
	ABGROUP	59.23	40.77					

Table 4
Descriptive statistics of whitened inputs and outputs.$\widehat{\varepsilon}_x, \widehat{\varepsilon}_y$

	Mean	SD	Min	1Q	Median	3Q	Max
Pure PVMATH	0.0186	0.4185	−1.2176	−0.2008	0.0071	0.2284	1.2915
Pure PVREAD	0.0167	0.4378	−1.4331	−0.1882	0.0178	0.2323	1.2753
Pure PVSCIE	0.0185	0.4081	−1.1704	−0.1950	0.0101	0.2234	1.2483
Pure ESCS	3.5889	34.3610	−106.5459	−12.6181	1.6437	18.3303	129.0121
Pure TSRATIO	3.6884	5.0647	0.0547	0.8870	1.9344	3.6842	21.2080
Pure SCHRES	3.5171	25.5664	−62.6006	−9.4867	1.6522	13.0147	108.7423

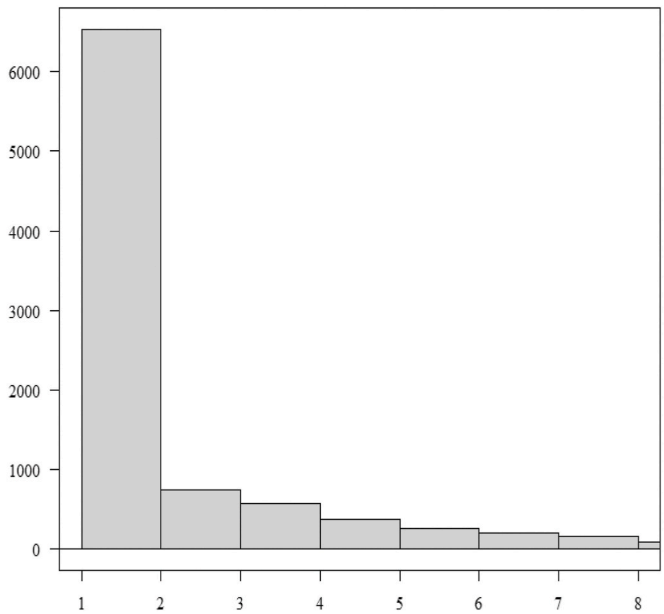
Table 5
Pearson and Spearman correlations between $\widehat{\varepsilon}_x, \widehat{\varepsilon}_y$ and *Zs*.

Pearson correlation	PROPCERT	SCHSIZE	RESPRES	RURAL	PRIVATE	SMCLASS	%REPEAT	%TRUANCY	ACCOUNT	ABGROUP
Pure PVMATH	0.0031	0.0055	−0.0127	0.0070	0.0576	−0.0089	0.0091	0.0058	0.0008	−0.0119
Pure PVREAD	0.0145	0.0130	−0.0079	0.0126	0.0385	−0.0129	0.0007	−0.0018	0.0091	−0.0014
Pure PVSCIE	0.0172	0.0233	−0.0140	0.0096	0.0213	−0.0172	0.0036	−0.0080	−0.0038	−0.0066
Pure ESCS	0.0073	−0.0006	−0.0413	0.0081	−0.0018	0.0197	−0.0313	−0.0323	−0.0094	−0.0314
Pure TSRATIO	−0.1785	0.2442	0.1112	0.1825	−0.1038	0.0268	0.0749	−0.1097	−0.0969	−0.0700
Pure SCHRES	0.0122	−0.0013	−0.0484	0.0016	−0.0571	0.0471	−0.0458	−0.0566	−0.0521	−0.0440
Spearman correlation	**PROPCERT**	**SCHSIZE**	**RESPRES**	**RURAL**	**PRIVATE**	**SMCLASS**	**%REPEAT**	**%TRUANCY**	**ACCOUNT**	**ABGROUP**
Pure PVMATH	−0.0271	0.0319	0.0725	0.0022	0.0792	−0.0207	0.0182	0.0112	0.0081	−0.0014
Pure PVREAD	0.0173	0.0331	0.0489	0.0227	0.0496	−0.0136	0.0010	−0.0010	0.0038	−0.0077
Pure PVSCIE	0.0075	0.0518	0.0655	0.0105	0.0472	−0.0159	0.0166	−0.0034	0.0045	−0.0011
Pure ESCS	−0.0250	0.0226	0.0727	0.0700	0.0306	0.0311	−0.0220	−0.0209	0.0046	−0.0117
Pure TSRATIO	−0.0440	0.4965	0.0694	0.1407	−0.1147	0.0193	−0.0003	−0.2657	−0.0357	0.0114
Pure SCHRES	0.0315	0.0056	0.0623	0.0138	0.0070	0.0172	−0.0199	−0.0240	−0.0079	−0.0091

deviation, minimum, maximum and number of efficient and superefficient units) of estimates for both models using order-m with the same trimming value ($m = \sqrt[3]{N^2}$). As expected, when we include information about contextual factors in the conditional model, the overall mean inefficiency decreases and the number of efficient and superefficient units increases notably. Actually, the results reported here justify why we chose to apply the model proposed by Florens et al. [16] instead of the traditional conditional model proposed by Daraio and Simar [8]. When the number of exogenous variables is as high as in our case (10), their inclusion in the conditional model significantly reduces its discriminative power (approximately 18% of the units are located on or above the frontier and the average efficiency is quite high). In fact, the efficiency of all schools increases under the conditional model. This makes sense since the analysis accounting for additional variables has a smaller reference group, as it only includes schools with similar characteristics. Moreover, the shape of the estimated distributions reported in Fig. 2 shows a high concentration of units at the higher efficiency levels in the conditional model, whereas this concentration is not as pronounced in the unconditional model (without contextual factors).

Subsequently, we investigated how the ratios $\widehat{R}_m$ shown in Equation (15) can provide information about the potential effects of contextual factors on the shape of the frontier. To do this, we computed the partial order-m ratios for $m = 9{,}369$ (robust version of the full ratios). Fig. 3 illustrates these ratios for each continuous exogenous variable. In order to facilitate the visual interpretation of these marginal effects on efficiency levels, we added a nonparametric regression line of the different ratios on *Zs*. For the dummy variables, we summarized the global direction of the effect in Table 9.

(a) Effect of school size
(b) Effect of percentage of students who skipped classes
(c) Effect of the level of responsibility in allocating resources
(d) Effect of the percentage of repeaters
(e) Effect of the proportion of fully certified teachers



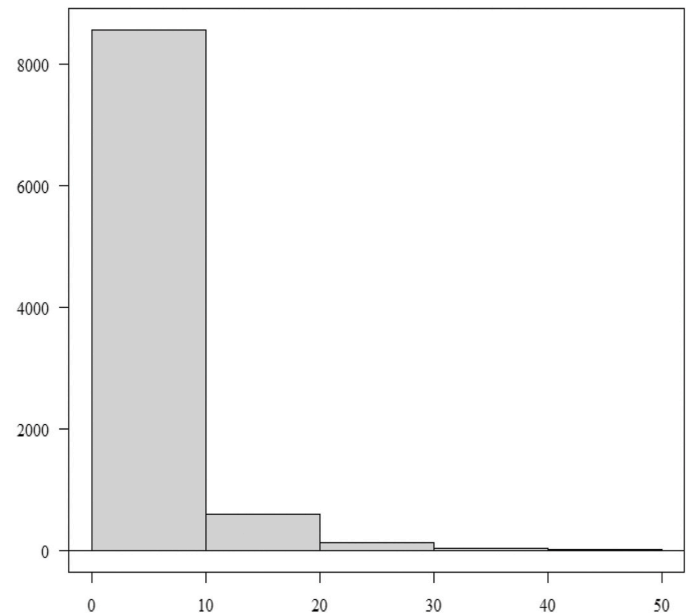


Fig. 1. Histograms of estimated pure inefficiencies relative to the full frontier $\widehat{\phi}$ (left panel) and the order-m frontier $\widehat{\phi}_m$ (right panel).

Table 6
Managerial efficiencies and results in science (average scores and standard deviations by country).

RK	Country	Managerial Efficiency	SD	RK	Country	PVSCIE	SD
1	Japan	0.6299	0.3416	1	Japan	537.43	63.80
2	Slovenia	0.6118	0.3210	2	Estonia	530.91	43.76
3	Norway	0.5947	0.3680	3	Finland	529.24	42.39
4	Korea	0.5889	0.3408	4	Korea	513.33	50.06
5	Canada	0.5348	0.3607	5	Canada	512.99	43.39
6	Israel	0.5341	0.3536	6	Netherlands	507.38	78.27
7	Austria	0.5165	0.3295	7	New Zealand	506.85	51.36
8	Greece	0.5096	0.3006	8	Poland	506.35	42.85
9	New Zealand	0.4990	0.3019	9	Australia	504.09	55.38
10	Poland	0.4814	0.3146	10	Germany	503.55	72.17
11	Finland	0.4791	0.2432	11	United Kingdom	502.21	49.65
12	Italy	0.4757	0.3101	12	Ireland	500.19	38.22
13	United States	0.4697	0.3840	13	Norway	499.02	35.33
14	Iceland	0.4696	0.3225	14	Switzerland	497.47	63.31
15	Switzerland	0.4696	0.3598	15	Sweden	496.60	51.33
16	Estonia	0.4671	0.3066	16	United States	493.10	48.74
17	United Kingdom	0.4521	0.3162	17	Belgium	492.85	73.51
18	Australia	0.4512	0.3275	18	Spain	492.22	35.95
19	France	0.4369	0.2910	19	Czech Republic	489.51	69.11
20	Sweden	0.4337	0.3044	20	Denmark	488.48	46.87
21	Denmark	0.4307	0.2840	21	Luxembourg	486.92	61.07
22	Turkey	0.4201	0.2514	22	France	486.44	75.13
23	Slovak Republic	0.3861	0.3246	23	Latvia	483.94	40.56
24	Czech Republic	0.3606	0.2787	24	Austria	483.46	70.12
25	Ireland	0.3537	0.2844	25	Portugal	481.76	53.36
26	Latvia	0.3480	0.2963	26	Italy	479.80	63.20
27	Belgium	0.3391	0.2923	27	Iceland	474.44	35.58
28	Spain	0.3319	0.2699	28	Slovenia	473.46	72.67
29	Luxembourg	0.3254	0.3126	29	Israel	463.15	68.16
30	Germany	0.3194	0.2265	30	Hungary	460.41	77.84
31	Hungary	0.3104	0.2274	31	Chile	451.73	68.50
32	Portugal	0.3095	0.2631	32	Slovak Republic	449.04	66.90
33	Mexico	0.3028	0.2295	33	Greece	447.67	64.02
34	Netherlands	0.2906	0.2584	34	Mexico	411.72	42.95
35	Chile	0.2777	0.2865	35	Turkey	410.02	58.09
	TOTAL	**0.4346**	**0.3024**		**TOTAL**	**487.08**	**56.39**

Note: Managerial efficiency represents the schools' average estimated pure efficiency order-*m* scores using the "whitened" inputs and outputs (cleaned from the effect of *Zs*) for each country. PVSCIE is the mean test score of schools in a country in the main subject evaluated in PISA 2015 (Science).

Fig. 3 shows different effects according to the plotted marginal views. School size appears to have a positive effect on the frontier (Fig. 3a), although it is only apparent for higher values. Likewise, the proportion of students who frequently skip classes has a clearly negative impact (Fig. 3b). For the index representing the responsibility of school staff in the allocation of resources, we found that it had a negative effect

Table 7
Proportion of efficient (and superefficient) units in each country.

RK	Country	%	RK	Country	%
1	Japan	31.82	19	United Kingdom	11.45
2	Korea	28.57	20	Sweden	11.39
3	Norway	27.07	21	Slovak Republic	10.34
4	Slovenia	23.12	22	Turkey	9.09
5	Israel	23.12	23	Ireland	8.38
6	Switzerland	18.50	24	Belgium	8.33
7	Greece	16.11	25	France	8.33
8	Poland	15.98	26	Spain	7.96
9	Luxembourg	15.91	27	Czech Republic	7.56
10	Canada	15.68	28	Denmark	7.21
11	Austria	14.87	29	Portugal	6.50
12	United States	14.69	30	Chile	5.73
13	Estonia	13.59	31	Finland	5.36
14	Australia	13.06	32	Netherlands	4.30
15	Iceland	12.10	33	Hungary	4.08
16	New Zealand	12.02	34	Mexico	4.00
17	Italy	11.60	35	Germany	3.82
18	Latvia	11.60		**TOTAL**	**12.66**

Table 8
Descriptive statistics of unconditional and conditional estimates.

	Unconditional model	Conditional model
Mean	1.1942	1.0857
SD	0.1309	0.0818
Min	0.9732	0.9676
Max	1.9609	1.1619
Efficient units	199 (2.12%)	466 (4.97%)
Superefficient units	216 (2.31%)	1,204 (12.85%)

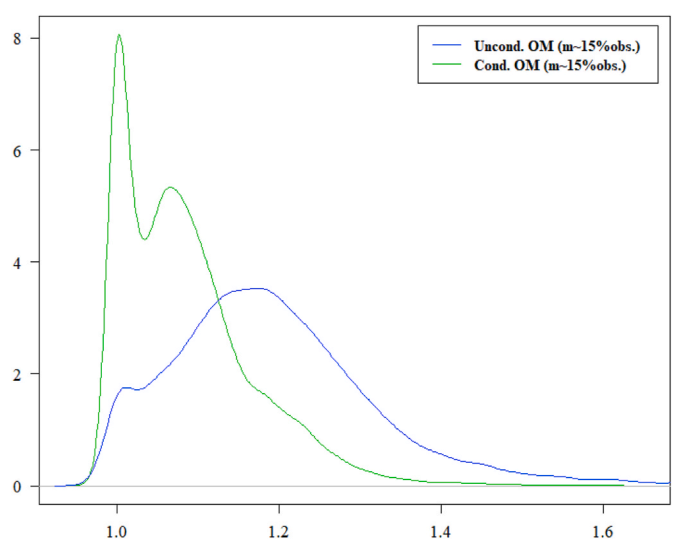


Fig. 2. Distributions of order-*m* scores (unconditional and conditional model).

for lower values, but then turned slightly positive as its values increased (Fig. 3c). These results are essentially consistent with previous evidence in empirical studies exploring the influence of these factors on efficiency measures of school performance (e.g., Ref. [23]. Nevertheless, we also found some unexpected results. For example, the variable reflecting the percentage of repeaters in the school (Fig. 3d) appeared to have no influence on efficiency, whereas the ratio of fully certified teachers was identified as having unfavourable effect (Fig. 3e).

With regard to the global impact of the dummy external variables on efficiency levels, the values reported in Table 9 indicate that location in a rural area has an unfavourable effect, as highlighted in previous studies based on PISA data [24]. For the remaining four variables (being a private school, having less than 20 students per class, having systems of accountability and using ability grouping in classes), the influence is positive, as was to be expected according to previous evidence found in the literature about the determinants of student achievement [53].

6. Conclusions

The analysis of the efficiency of educational systems based on the performance of their schools is one of the hottest topics in the field of economics of education for two main reasons. On the one hand, the academic results achieved by students attending secondary schools is widely accepted as a measure of the quality of education systems, which has a strong and stable association with greater economic growth rates [54]. On the other hand, most countries have made a huge financial effort to provide resources for education over the last decades, although there is no direct positive correlation between higher per capita public expenditure on education and higher academic outcomes [55]. For these reasons, benchmarking schools and analysing their efficiency worldwide is one the most promising tools for learning the best managerial practices. Moreover, this analysis may help policy makers to rule out educational policies that do not work and reallocate public expenditure to more promising alternatives.

In this paper, we have applied some of the most recent nonparametric methods to assess the performance of a sample of secondary schools from the 35 OECD countries participating in PISA 2015. Specifically, we used the flexible location-scale model suggested by Florens et al. [16]. Using this model, we can eliminate the dependence of production inputs/outputs on external factors (Z) by means of a nonparametric regression model. Therefore, we can estimate pure or managerial efficiencies uninfluenced by Z. The main advantage of this approach is that the estimated measures of performance are more reliable for producing rankings or benchmarks of countries or education systems in the hypothetical case that schools are all operating in the same environment.

Our results reveal several interesting issues. First, we found that, although there are some similarities between the classification of countries according to their school performance in terms of efficiency and the average results achieved in PISA, the consideration of inputs involved in the educational process, together with the corrections made to equalize the conditions under which they are operating, lead to some substantial changes in the ranking of countries. Thus, we find that there are nations with good results in PISA that are ranked relatively much lower according to their pure efficiency level (e.g. Netherlands, Germany or Finland), as well as countries with poor PISA results that are ranked among the top performers in terms of efficiency (e.g., Slovenia, Greece or Israel). Second, we notice that the influence of the contextual factors included in our model appear to be consistent with previous evidence in the literature on the determinants of efficiency and academic results. In particular, we observe a favourable effect for private ownership, small classes, school size, the use of ability grouping and the existence of accountability systems, and a negative impact of being located in a rural area and having a high proportion of students who skip classes.

Although these findings provide some insights for the analysis of school efficiency, more research is still needed to further explore the results discussed here. For instance, we should also explore the potential influence of cross-country heterogeneity by incorporating some additional contextual factors at country level, since some previous studies have suggested that such variables might have a bigger impact on efficiency measures than school environmental factors [19,25]. Likewise, we should note that the results of the approach used in this paper cannot be interpreted causally, since this would entail neglecting the potential presence of unobserved heterogeneity in school performance. For instance, we totally ignored the potential accumulative impact of inputs, since our results are based on cross-sectional data. In this sense, it is worth mentioning that some authors have recently developed methods

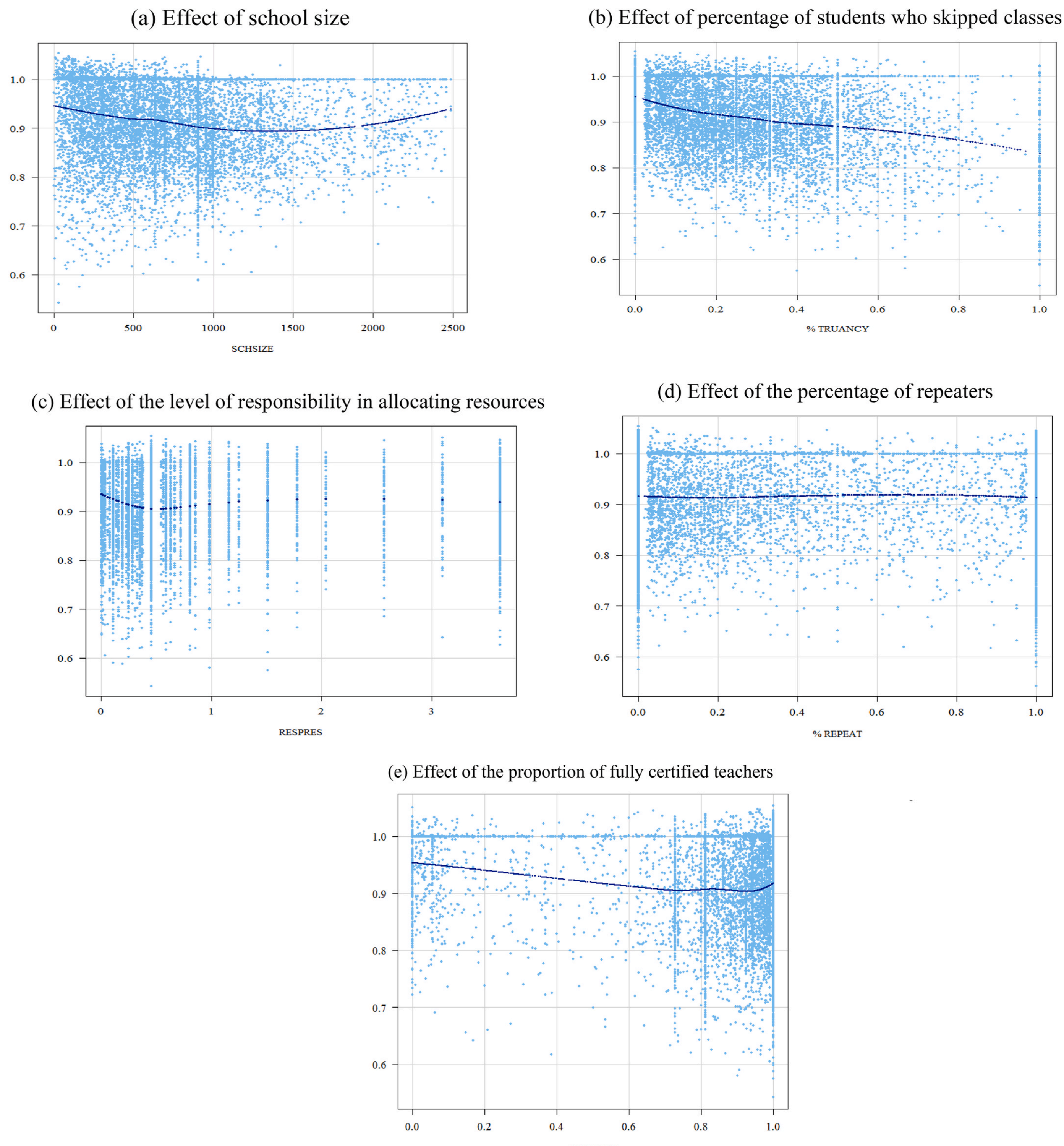


Fig. 3. Effect of continuous exogenous variables on efficiency estimates.

Table 9
Effect of dummy exogenous variables on efficiency estimates.

Variable	Rural	Private	Smclass	Account	Abgroup
Effect	Unfavourable	Favourable	Favourable	Favourable	Favourable

combining the use of impact evaluation techniques with production frontiers [56]. Therefore, a potential area for further research could be to apply these methodologies to derive more accurate estimates of performance that may lead to a causal interpretation of empirical results.

CRediT authorship contribution statement

Jose Manuel Cordero: Formal analysis, Writing - review & editing, The three authors have contributed to the design, analysis, discussion, writing and critical review of the paper, and give evidence of the originality and consent to publication. **Cristina Polo:** Formal analysis, Writing - review & editing, The three authors have contributed to the design, analysis, discussion, writing and critical review of the paper, and give evidence of the originality and consent to publication. **Rosa Simancas:** Formal analysis, Writing - review & editing, The three authors have contributed to the design, analysis, discussion, writing and critical review of the paper, and give evidence of the originality and consent to publication.

Acknowledgements

We are grateful to two anonymous reviewers for helpful comments on an earlier version of this manuscript. The authors also thank the financial support from the Spanish Ministry of Economy and Business (Ministerio de Economía y Empresa) through grant ECO2017-83759-P and Junta de Extremadura through grant IB16171.

References

[1] Gustafsson JE, Rosén M. Quality and credibility of international studies. In: Strietholt R, Bos W, Gustafsson JE, Rosén M, editors. Educational policy evaluation through international comparative assessments, Waxmann Verlag; 2014. p. 19–32.
[2] Woessmann L. School resources, educational institutions and student performance: the international evidence. Oxf Bull Econ Stat 2003;65(2):117–70.
[3] Ammermüller A, Heijke H, Woessmann L. Schooling quality in eastern europe: educational production during transition. Econ Educ Rev 2005;24(5):579–99.
[4] Hanushek EA, Woessmann L. Institutional structures of the education system and student achievement: a review of cross-country economic research. In: Strietholt R, Bos W, Gustafsson JE, Rosen M, editors. Educational policy evaluation through international comparative assessments. Waxmann Verlag; 2014. p. 145–76.
[5] Teddlie C, Reynolds D. The international handbook of school effectiveness research. London: Routledge; 2000.
[6] Dufrechou PA. The efficiency of public education spending in Latin America: a comparison to high-income countries. Int J Educ Dev 2016;49:188–203.
[7] Agasisti T, Zoido P. The efficiency of schools in developing countries analysed through PISA 2012 data. Soc Econ Plann Sci 2019:68. https://doi.org/10.1016/j.seps.2019.05.002. In press.
[8] Daraio C, Simar L. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. J Prod Anal 2005;24(1):93–121.
[9] Daraio C, Simar L. Advanced robust and nonparametric methods in efficiency analysis. Methodologies and Applications, Springer, New York. 2007.
[10] Daraio C, Simar L. Conditional nonparametric frontier models for convex and non-convex technologies: a unifying approach. J Prod Anal 2007;28:13–32.
[11] Cazals C, Florens JP, Simar L. Nonparametric frontier estimation: a robust approach. J Econom 2002;106:1–25.
[12] Daraio C, Simar L, Wilson PW. Central limit theorems for conditional efficiency measures and tests of the 'separability' condition in non-parametric, two-stage models of production. Econom J 2018;21(2):170–91.
[13] Badin L, Daraio C, Simar L. Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. Eur J Oper Res 2010;201(2):633–40.
[14] Badin L, Daraio C, Simar L. A bootstrap approach for bandwidth selection in estimating conditional efficiency measures. Eur J Oper Res 2019;277(2):784–97.
[15] Badin L, Daraio C, Simar L. How to measure the impact of environmental factors in a nonparametric production model? Eur J Oper Res 2012;223:818–33.
[16] Florens J, Simar L, van Keilegom I. Frontier estimation in nonparametric location-scale models. J Econom 2014;178:456–70.
[17] De Witte K, López-Torres L. Efficiency in education: a review of literature and a way forward. J Oper Res Soc 2017;68(4):339–63.
[18] De Jorge J, Santín D. Determinantes de la eficiencia educativa en la Unión Europea. Hacienda Publica Espanola 2010;193:131–55.
[19] Cordero JM, Santín D, Simancas R. Assessing European primary school performance through a conditional nonparametric model. J Oper Res Soc 2017;68(4):364–76.
[20] Deutsch J, Dumas A, Siber J. Estimating an educational production function for five countries of Latin America on the basis of the PISA data. Econ Educ Rev 2013;36:245–62.
[21] Wilson PW. Efficiency in education production among PISA countries with emphasis on transitioning economies. Department of Economics, University of Texas. 2005.
[22] Sutherland D, Price R, Gonand F. Improving public spending efficiency in primary and secondary education. OECD J: Econ Stud 2009;(1):1–30. 2009.
[23] Agasisti T, Zoido P. Comparing the efficiency of schools through international benchmarking: results from an empirical analysis of OECD PISA 2012 data. Educ Res 2018;47(6):352–62.
[24] Aparicio J, Cordero JM, González M, López-Espin JJ. Using non-radial DEA to assess school efficiency in a cross-country perspective: an empirical analysis of OECD countries. Omega 2018;79:9–20.
[25] Cordero JM, Polo C, Santín D, Simancas R. Efficiency measurement and cross-country differences among schools: a robust conditional nonparametric analysis. Econ Modell 2018;74:45–60.
[26] Afonso A, St Aubyn M. Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs. Econ Modell 2006;23(3):476–91.
[27] Clements B. How efficient is education spending in Europe? Eur Rev Econ Finc 2002;1(1):3–26.
[28] Giménez V, Prior D, Thieme C. Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. J Oper Res Soc 2007;58(8):996–1007.
[29] Verhoeven M, Gunnarsson V, Carcillo S. Education and health in G7 countries: achieving better outcomes with less spending, (No. 2007-2263). International Monetary Fund; 2007.
[30] Giambona F, Vassallo E, Vassiliadis E. Educational systems efficiency in European Union countries. Stud Educ Eval 2011;37(2):108–22.
[31] Thieme C, Giménez V, Prior D. A comparative analysis of the efficiency of national education systems. Asia Pac Educ Rev 2012;13(1):1–15.
[32] Agasisti T. The efficiency of public spending on education: an empirical comparison of EU countries. Eur J Educ 2014;49(4):543–57.
[33] Aristovnik A, Obadić A. Measuring relative efficiency of secondary education in selected EU and OECD countries: the case of Slovenia and Croatia. Technol Econ Dev Econ 2014;20(3):419–33.
[34] Coco G, Lagravinese R. Cronyism and education performance. Econ Modell 2014;38:443–50.
[35] Bogetoft P, Heinesen E, Tranæs T. The efficiency of educational production: a comparison of the Nordic countries with other OECD countries. Econ Modell 2015;50:310–21.
[36] Giménez V, Thieme C, Prior D, Tortosa-Ausina E. An international comparison of educational systems: a temporal analysis in presence of bad outputs. J Prod Anal 2017;47(1):83–101.
[37] Simar L, Wilson PW. Estimation and inference in two-stage, semi-parametric models of production processes. J Econom 2007;136(1):31–64.
[38] Simar L, Wilson PW. Two-stage DEA: caveat emptor. J Prod Anal 2011;36(2):205.
[39] O'Donnell C, Rao D, Battese G. Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empir Econ 2008;37(2):231–55.
[40] Fried HO, Schmidt SS, Yaisawarng S. Incorporating the operating environment into a nonparametric measure of technical efficiency. J Prod Anal 1999;12(3):249–67.
[41] Mastromarco C, Simar L. Globalization and productivity: a robust nonparametric world frontier analysis. Econ Modell 2018;69:134–49.
[42] Färe R, Grosskopf S. Theory and application of directional distance functions. J Prod Anal 2000;13(2):93–103.
[43] OECD. PISA 2015 assessment and analytical framework: science, reading, mathematic, financial literacy and collaborative problem solving. Paris: PISA, OECD Publishing; 2017. Revised edition.
[44] Willms JD, Smith T. A manual for conducting analyses with data from TIMSS and PISA. Report prepared for UNESCO Institute for Statistics. 2005.
[45] OECD. *PISA 2015 technical report.* PISA. Paris: OECD Publishing; 2016.
[46] Rasch G. (1960/1980). Probabilistic models for some intelligence and attainment tests. Danish Institute for Educational Research, Expanded edition. Copenhagen: The University of Chicago Press; 1980.
[47] Avvisati F, Keslair F. REPEST: stata module to run estimations with weighted replicate samples and plausible values. Statistical Software Components S457918, Boston College Department of Economics, revised 06 Jan 2020. 2014.
[48] Aparicio J, Cordero JM, Pastor JT. The determination of the least distance to the strongly efficient frontier in data envelopment analysis oriented models: modelling and computational aspects. Omega 2017;71:1–10.
[49] Crespo-Cebada E, Pedraja-Chaparro F, Santín D. Does school ownership matter? An unbiased efficiency comparison for regions of Spain. J Prod Anal 2014;41(1):153–72.
[50] Thieme C, Prior D, Tortosa-Ausina E. A multilevel decomposition of school performance using robust nonparametric frontier techniques. Econ Educ Rev 2013;32:104–21.
[51] Mastromarco C, Simar L. Cross-section dependence and latent heterogeneity to evaluate the impact of human capital on country performance. Discussion Paper UCL-Université Catholique de Louvain 2017. 2017/30.
[52] Tauchmann H. Partial frontier efficiency analysis. STATA J 2012;12(3):461–78.
[53] Cordero JM, Cristobal V, Santín D. Causal inference on education policies: a survey of empirical studies using PISA, TIMSS and PIRLS. J Econ Surv 2018;32(3):878–915.
[54] Hanushek EA, Kimko DD. Schooling, labor-force quality, and the growth of nations. Am Econ Rev 2000;90(5):1184–208.
[55] Hanushek EA. The failure of input-based schooling policies. Econ J 2003;113(485):64–98.

[56] Santín D, Sicilia G. Impact evaluation and frontier methods in education: a step forward. In: Johnes G, Johnes J, Agasisti T, López-Torres L, editors. (2017). *Handbook of Contemporary Education Economics*. London: Edward Elgar Publishing; 2017. p. 211–45.
[57] Roudaut N, Vanhems A. Explaining firms efficiency in the Ivorian manufacturing sector: a robust nonparametric approach. J Prod Anal 2012;37(2):155–69.
[58] Kourtzidis S, Tzeremes P, Tzeremes NG. Conditional time-dependent nonparametric estimators with an application to healthcare production function. J Appl Stat 2019;46(13):2481–90.
[59] Mastromarco C, Stastna L, Votapkova J. Efficiency of hospitals in the Czech Republic: conditional efficiency approach. J Prod Anal 2019;51(1):73–89.
[60] Tzeremes NG. Technological change, technological catch-up and export orientation: evidence from Latin American Countries. J Prod Anal 2019;52(1–3):85–100.
[61] Worthington AC. An empirical survey of frontier efficiency measurement techniques in education. Educ Econ 2001;9(3):245–68.
[62] Johnes J. Operational research in education. Eur J Oper Res 2015;243(3):683–96.
[63] Jeong S, Park B, Simar L. Nonparametric conditional efficiency measures: asymptotic properties. Ann Oper Res 2010;173:105–22.
[64] Wang FC, Hung WT, Shang JK. Measuring pure managerial efficiency of international tourist hotels in Taiwan. Serv Ind J 2006;26(1):59–71.
[65] Shyu J, Chiang T. Measuring the true managerial efficiency of bank branches in Taiwan: a three-stage DEA analysis. Expert Syst Appl 2012;39:11494–502.
[66] Cordero-Ferrera JM, Pedraja-Chaparro F, Salinas-Jiménez J. Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs. Appl Econ 2008;40(10):1323–39.
[67] Cazals C, Fève F, Florens JP, Simar L. Nonparametric instrumental variables estimation for efficiency frontier. J Econom 2016;190(2):349–59.
[68] Simar L, Vanhems A, Van Keilegom I. Unobserved heterogeneity and endogeneity in nonparametric frontier estimation. J Econom 2016;190(2):360–73.
[69] Simar L, Vanhems A. Probabilistic characterization of directional distances and their robust versions. J Econom 2012;166(2):342–54.
[70] Daraio C, Simar L. Directional distances and their robust versions: computational and testing issues. Eur J Oper Res 2014;237(1):358–69.
[71] Daraio C, Simar L. Efficiency and benchmarking with directional distances: a data-driven approach. J Oper Res Soc 2016;67(7):928–44.
[72] Daraio C, Simar L, Wilson PW. Fast and efficient computation of directional distance estimators. Annal Ops Res 2019. https://doi.org/10.1007/s10479-019-03163-9. forthcoming.
[73] Li Q, Racine JS. Nonparametric econometrics: theory and practice. Princeton: Princeton University Press; 2007.
[74] Hall P, Racine J, Li Q. Cross-validation and the estimation of conditional probability densities. J Am Stat Assoc 2004;99(468):1015–26.
[75] Li Q, Racine JS. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. J Bus Econ Stat 2008;26(4):423–34.
[76] Badin L, Daraio C. Explaining efficiency in nonparametric frontier models: recent developments in statistical inference. In: Exploring research frontiers in contemporary statistics and econometrics. Heidelberg: Physica; 2011. p. 151–75.
[77] Wu M. Measurement, sampling, and equating errors in large-scale assessments. Educ Meas 2010;29(4):15–27.
[78] Mislevy RJ, Beaton AE, Kaplan B, Sheehan KM. Estimating population characteristics from sparse matrix samples of item responses. J Educ Meas 1992;29(2):133–61.
[79] Wu M. The role of plausible values in large-scale surveys. Stud Educ Eval 2005;31(2–3):114–28.

José M. Cordero is a Full Professor in the Department of Economics of the University of Extremadura. He holds a PhD in Economics from the same university. His research interests are in the fields of efficiency and productivity analysis and economics of education. His research experience has been reflected in multiple scientific articles published in referred journals as well as in several books and book chapters. His publication outlets include *Omega, European Journal of Operational Research, Journal of Productivity Analysis, Socio Economic Planning Sciences, Computers and Operations Research, European Journal of Health Economics, Economic Modelling* and *Journal of Economic Surveys*, among others. He is currently the President of the AEDE (Spanish Economics of Education Association).

Cristina Polo is currently a Lecturer in the Department of Economics of the University of Extremadura. She holds a PhD in Economics from the same university. Her main research lines are focused on computational aspects related to efficiency measurement and its application in public services as diverse as education, health and municipal services. She has several publications in prestigious international journals like *Journal of Productivity Analysis, Operational Research, Economic Modelling* or *Applied Economics.*

Rosa Simancas is currently an Associate Professor in the Department of Economics of the University of Extremadura and Secretary of the AEDE (Spanish Economics of Education Association). Her main lines of research are economics of education and public policy evaluation. Her publication outlets include *Journal of the Operational Research Society, Survey Research Methods, Economic Modelling, Journal of Policy Modeling* and *Applied Economics*, among others. She has participated in several competitive research projects, including both at European and national level.