



Efficiency measurement and cross-country differences among schools: A robust conditional nonparametric analysis[☆]

Jose M. Cordero^{a,*}, Cristina Polo^a, Daniel Santín^b, Rosa Simancas^a

^a Department of Economics, Faculty of Economics and Business, University of Extremadura, Avd. Elvas s/n, 06006 Badajoz, Spain

^b Department of Applied Economics VI, Faculty of Economics and Business, Complutense University of Madrid, Campus de Somosaguas, 28223, Pozuelo de Alarcón, Madrid, Spain



ARTICLE INFO

JEL codes:

I21
H75
C14

Keywords:

Education
Efficiency
Nonparametric
Cross-country comparison

ABSTRACT

Analyzing the efficiency of educational systems is one of the main focuses of the policy debate to promote national competitiveness and future economic growth. In this paper, we assess the performance of secondary schools from 36 countries (26 OECD countries and 10 partners) participating in PISA 2012. For this purpose, we apply a robust conditional nonparametric approach that allows us to incorporate the effect of contextual factors at both school and country level in the estimation of efficiency measures. Our results suggest that there is a greater heterogeneity across countries than across schools. Particularly, we find that differences in efficiency estimates are mainly explained by economic indicators and cultural values. In contrast, some factors previously identified as potential determinants of student achievement, like the existence of tracking or central examinations, do not seem to significantly affect the efficiency of secondary schools.

1. Introduction

Thanks to the participation of an extensive group of countries in international large-scale assessments like PISA (*Programme for International Student Assessment*) or TIMSS (*Trends in International Mathematics and Science Study*), researchers have access to rich and extensive cross-national databases that they can use to assess education system performance internationally (Gustafsson, 2008; Kamens, 2009). Researchers can use this information to analyze differences in achievement between and within countries and investigate why and how some schools and teachers are more effective than others in promoting student learning or assess the impact of skills on economic and social outcomes (Creemers and Kyriakides, 2008; Hanushek and Woessmann, 2011). Likewise, international comparisons are especially useful for evaluating the effects of some institutional features of education systems that cannot be estimated without access to data on different countries (Hanushek and Woessmann, 2014; Strietholt et al., 2014). Conclusions and results from these analyses provide valuable decision-making guidelines for policy makers to reorient the national education system based on what is currently working in other countries.

Most cross-country studies analyze educational effectiveness, i.e. they

estimate an educational production function by means of an equation linking resource inputs with educational outcomes after controlling for various contextual characteristics to investigate the main factors influencing educational attainment (see Hanushek, 1979; Todd and Wolpin, 2003). However, resource utilization is also a key matter of concern in science and technology management (Teddlie and Reynolds, 2000). Indeed, education system efficiency is now a hot topic among educational stakeholders because of the size of public spending on education and the shortage of resources raised from taxation that most countries are now facing. In particular, policy makers and researchers alike are concerned with developing guidelines for educational institutions to encourage improvements in school outcomes given their school factors. As a result, the literature on school performance assessment is growing, although most empirical studies address schools from the same country or region.¹

This study proposes an international comparison of education production efficiency using cross-country data on secondary schools from different countries participating in PISA 2012. In this sense, it is worth mentioning that international comparisons are extremely challenging, since countries might differ significantly with regard to multiple cultural and institutional features as well as the education system structure. These differences may pose an obstacle to the comparison of schools operating

[☆] We would like to thank the editor and the two anonymous referees for their helpful comments to improve the quality of this paper.

* Corresponding author.

E-mail addresses: jmcordero@unex.es (J.M. Cordero), cristinapf@unex.es (C. Polo), dsantin@ccee.ucm.es (D. Santín), rsimancas@unex.es (R. Simancas).

¹ See Worthington (2001) for an early review of this literature and Johnes (2015) or De Witte and López-Torres (2017) for updated revisions.

in diverse frameworks. Some studies have addressed this problem by limiting the comparison group to similar countries (e.g. Bogetoft et al., 2015; Cordero et al., 2017b; Dufrechou, 2016). In this paper, however, our dataset includes a large sample of thirty-six heterogeneous countries, thus we have gone one step further and taken into account data about the diverse educational contexts in which schools are operating when estimating the efficiency measures of school performance. In this way, each unit can be benchmarked with other units from different countries provided that their operational environment is similar.

To do this, we adopt the robust conditional nonparametric approach developed by Daraio and Simar (2005, 2007a, 2007b). These authors extend the probabilistic formulation of the production process proposed by Cazals et al. (2002) to account for heterogeneous contextual factors without imposing the restrictive separability assumption required by traditional second-stage models in order to provide meaningful results.² Using this approach, we can also test the significance of the contextual factors included in the model at both school and country level. Then, by exploiting the relationship between the conditional and unconditional measures, we can investigate the direction of their effect (favorable or unfavorable) on the production process. Furthermore, we can obtain clean efficiency scores by applying the second-stage approach suggested by Badin et al. (2012) to eliminate the effects of contextual conditions.

The country-level contextual factors addressed in this study include variables representing the economic, cultural and social context. This is an interesting contribution because most comparative studies based on data from large-scale international assessments overlook these variables (Thät and Must, 2013; Zhao et al., 2008). In particular, we retrieve data from the World Bank's Indicators database about some key economic indicators and approximate each country's cultural background by collecting data from the respondents of the World Values Survey (WVS) with regard to qualities that children are encouraged to learn at home.³ To the best of our knowledge, this data source has seldom been used in previous comparative studies. Some exceptions are represented by Coco and Lagravinese (2014), who use this information to derive a measure of cronyism as a potential determinant of educational performance in their evaluation of OECD countries using PISA data, and Mendez (2015), who examines the role of the above qualities in explaining country differences with respect to student performance in PISA. In our case, these variables are included as contextual factors influencing the performance of schools from one country compared with schools from other countries.

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 that focus especially on efficiency measures. Section 3 describes the methodology applied in our empirical analysis. Section 4 explains the main characteristics of the data and variables selected for the empirical analysis. Section 5 discusses the main results compared with the existing literature. Finally, Section 6 outlines some concluding remarks.

2. Literature review

Since the publication of the pioneering work by Woessmann (2003) combining international student- and school-level microdata with several country-level indicators, multiple studies have adopted a cross-country approach to explore the main determinants of educational achievement from different perspectives (Ammernüller et al., 2005; Hanushek and

Woessmann, 2011; Le Donne, 2014). These studies mainly address the use of econometric techniques to identify significant causal relationships between student background, school-related variables and educational outcomes (typically represented by test scores).⁴

The above empirical studies usually focus on some specific school factors, such as the class size (West and Woessmann, 2006; Woessmann and West, 2006), instructional time (Rivkin and Schiman, 2015) or divergences in performance between public and private schools (Vandenberghe and Robin, 2004; West and Woessmann, 2010). Likewise, these approaches have also been employed in a growing body of literature analyzing the impact of specific institutional features of education systems on educational attainment (Braga et al., 2013). These features include the existence of central examinations, which has been identified as a factor associated with better student performance in international tests (Bol et al., 2014; Woessmann, 2003, 2005); the practice of tracking,⁵ which appears to have a negative impact on average student performance (Hanushek and Woessmann, 2006) and promote educational inequality (Bol et al., 2014; Brunello and Checchi, 2007); or the length of pre-schooling, which has a positive effect on student performance (Schuetz et al., 2008).

As already mentioned, however, none of the above studies take into account the possibility of there being an unexpected level of inefficiency in student, school or country performance (Levin, 1974). Thus, over the last few years, interest in applying frontier methods to data from large-scale international assessments to evaluate the efficiency of education systems in a cross-country framework has grown notably. This parallel branch of research is valuable for monitoring efficiency differences across countries and the determinants that influence education system performance. Among those works, the most common ones are those using cross-sectional data aggregated at a country level (Afonso and St Aubyn, 2006; Agasisti, 2014; Aristovnik and Obadic, 2014; Bogetoft et al., 2015; Clements, 2002; Coco and Lagravinese, 2014; Giambona et al., 2011; Gimenez et al., 2007, 2017; Thieme et al., 2012; Verhoeven et al., 2007). Nevertheless, we can also find studies comparing the performance of education systems in different countries using school level data. For instance, Sutherland et al. (2009) study the performance of schools from 30 OECD countries participating in PISA 2003; Agasisti and Zoido (2015) derive efficiency measures for more than 8600 schools in 30 countries using PISA 2012 data comparing efficiency scores and measures of equity; Aparicio et al. (2018b) assess schools operating in the 34 OECD countries participating in PISA 2012 and identify different levels of inefficiency for reading and mathematics. Finally, De Jorge and Santín (2010) and Deutsch et al. (2013) use student-level PISA data to estimate the efficiency of European Union and Latin American countries, respectively.

Most of the above studies use nonparametric techniques like DEA or FDH to estimate performance efficiency measures since they are flexible enough to adapt to the characteristics of public services provision,⁶ especially to their multi-input multi-output nature. Moreover, in some cases, a two-stage procedure is also applied to examine the potential influence of contextual variables on efficiency estimates (e.g. Afonso and St Aubyn, 2006; Agasisti, 2014; Agasisti and Zoido, 2015; Aparicio et al., 2018b; De Jorge and Santín, 2010; Verhoeven et al., 2007). The main problem with this procedure is that it assumes that environmental factors affect the shape of the distribution of inefficiencies (i.e. mean, variance, etc.) but not the attainable set or the estimated frontier. This is often

² Second-stage models rely on the often-unrealistic assumption that contextual variables only affect the shape of the distribution of inefficiencies, but not the attainable set or the estimated frontier (see Simar and Wilson, 2007, 2011 for details).

³ Our definition of culture is based on the idea suggested by Fernandez and Fogli (2009), i.e. the set of beliefs and preferences that condition individuals' actions, vary systematically across either socially or geographically defined groups and are transmitted to successive generations.

⁴ See Cordero et al. (2017a) for a detailed review of this literature.

⁵ This is a form of stratification where students are placed in different schools or classes based on observed past or expected future achievement. This process differs widely across countries in terms of the age at which the selection takes place, as well as in the degree of differentiation between different tracks (generally distinguishing between academic or vocational education).

⁶ Nevertheless, some several studies (e.g. Deutsch et al., 2013; Sutherland et al., 2009) use parametric methods.

unrealistic, since some school-related factors, as well as the social, economic and institutional characteristics of each country, can be expected to be influencing both educational outcomes and the resources employed. Although the separability between the input–output space and the space of external variables can be tested in advance using the statistical tools developed by Daraio et al. (2010, 2015, 2018), none of the above empirical studies examined whether this assumption holds before applying this method.

Whenever the two-stage procedure is found to be inappropriate, the alternative option is to use conditional measures of efficiency (Daraio and Simar, 2005; 2007a; 2007b). Using these measures the contextual factors can be included directly in the estimation of efficiency scores without testing for this restrictive separability condition. To the best of our knowledge, Cordero et al. (2017b) were the first researchers to apply this methodology to measure efficiency in the educational context. They analyzed the effect of several environmental school factors and country indicators on the performance of primary schools from 16 European countries participating in the Progress in International Reading Literacy Study (PIRLS) in 2011.

In this study, we assess the performance of secondary schools using PISA 2012. Using PISA, we can explore the influence of a wider range of contextual factors, including some specific institutional features that only can be examined at the secondary education level, such as the influence of tracking⁷ or central examinations.⁸ In addition, our sample includes a much larger number of countries (36), which implies more heterogeneity across education systems.

3. Methodology

The production technology that schools use to transform a set of inputs ($x \in \mathbb{R}_+^p$) into outputs ($y \in \mathbb{R}_+^q$) can be characterized by the attainable set of feasible combinations of inputs and outputs as

$$\mathcal{P} = \{(x, y) \in \mathbb{R}_+^{p+q} | x \text{ can produce } y\} \quad (1)$$

Besides, additional contextual or environmental factors $Z \in \mathbb{R}_+^k$ that potentially affect school performance. Following the probabilistic formulation introduced by Cazals et al. (2002) and Daraio and Simar (2005, 2007b), the production process can be described by the joint probability function $H_{XY|Z}$ that represents the probability of a unit operating at level (x, y) and facing the same environmental conditions ($Z = z$) being dominated:

$$H_{XY|Z}(x, y|z) = \Pr(X \leq x, Y \geq y | Z = z) = S_Y(y|x, z)F_X(x|z) \quad (2)$$

where $S_Y(y|x, z)$ represents the survival function of Y and $F_X(x|z)$ the cumulative distribution function of X . Unless we consider the existing heterogeneity among schools, we would be implicitly assuming that all the schools are operating within the most favorable environment. This is unlikely in many cases. As proposed by Haelermans and De Witte (2012) or De Witte and Kortelainen (2013), we explore the potential influence of some external school-level variables. However, we also account for potential specific country-level features that can affect the school performance. This conditional methodology assumes that both school- and country-level factors can have a bearing on the shape of the best practice frontier, i.e. the separability condition does not hold for this model (Badin et al., 2014).

The conditional output measure $\lambda(x, y|z)$ can be analogously defined as the Farrell efficiency score:

$$\lambda(x, y|z) = \sup\{\lambda > 0 | H_{XY|Z}(x, \lambda y|z) > 0\} = \sup\{\lambda > 0 | S_{Y|X,Z}(\lambda y|x, z) > 0\} \quad (3)$$

We can use a plug-in rule providing conditional estimators of the full frontier, such as the free disposal hull (FDH) or data envelopment analysis (DEA), to define the nonparametric estimators of the conditional frontier $\lambda(x, y|z)$.⁹ These estimators have been extensively applied in many empirical papers¹⁰, although they are very sensitive to outlying observations or extreme data points because they both fully envelop all sample observations. Partial frontiers can be used to avoid this problem. Partial frontiers are regarded as robust versions of the previous ones (Daouia and Gijbels, 2011). Specifically, in this paper we use order- m frontiers developed by Cazals et al. (2002) for estimating the conditional $\lambda(x, y|z)$ and the unconditional $\lambda(x, y)$ efficiency estimators. The use of this approach implies considering the expectation of the best practice among m peers randomly drawn from the population of units for which $X \leq x$ as a benchmark. We repeat this procedure B times and compute the final order- m efficiency measure as the simple mean ($\hat{\lambda}_{mi}$) of the resulting measures ($\hat{\lambda}_{mi}^1, \dots, \hat{\lambda}_{mi}^B$). The unconditional $\lambda(x, y)$ and conditional order- m output efficiency estimators (Daraio and Simar, 2007a) can be defined as follows

$$\hat{\lambda}_m(x, y) = \int_0^\infty [1 - (1 - S_{Y|X}(uy|X \leq x))]^m du \quad (4)$$

$$\hat{\lambda}_m(x, y|z) = \int_0^\infty [1 - (1 - S_{Y|X,Z}(uy|X \leq x, Z = z))]^m du \quad (5)$$

where the ρ -dimensional random variables x_1, \dots, x_m are drawn randomly and repeatedly from the conditional distribution of X given $y_i \geq y$. For acceptable values of m , the efficiency score values will be greater than one. This indicates that schools are inefficient, as outputs can be increased without changing input levels. Therefore, an evaluated school may not belong to the set of schools used to output its score. In this case, it will be located above the production frontier, resulting in a score of less than one ($\hat{\lambda} < 1$). This school can be labelled as superefficient, since the order- m frontier has higher output levels than the school under analysis (Daraio and Simar, 2007a). The proportion of superefficient observations varies depending on the choice of m , since the probability of a unit not belonging to the frontier is conditioned by the size of the drawn sample relative to the total sample size. The main advantage of partial frontier estimates and related efficiency scores is that they are less influenced by extreme values and hence are more robust to outliers. Additionally, they are unaffected by the well-known *curse of dimensionality* which is a feature of most nonparametric estimators, including the DEA and FDH.

The computation of conditional efficiency estimators requires adopting smoothing techniques for the exogenous variables in z (due to the equality constraint $Z = z$):

$$\hat{S}_{Y|X,Z,n}(y|x, z) = \frac{\sum_{i=1}^n I(x_i \leq x, y_i \geq y) K_h((z - z_i)/h)}{\sum_{i=1}^n I(x_i \leq x) K_h((z - z_i)/h)} \quad (6)$$

This approach relies on the estimation of a nonparametric kernel function $K(\cdot)$ to select the appropriate reference partners and a bandwidth parameter h using some bandwidth choice method. In this case, we adopt the data-driven selection method developed by Badin et al. (2010). This approach relies on the least squares cross-validation (LSCV) procedure developed by Hall et al. (2004) and Li and Racine (2007). The advantage of this approach is that it detects and smoothes out irrelevant factors by providing large bandwidth parameters. Note that, since our

⁷ None of the countries apply this strategy to children aged under 10 years (see Brunello and Checchi, 2007, for details).

⁸ Not many countries use curriculum-based external examinations in primary education (see Woessmann et al., 2009, for details).

⁹ See Daraio and Simar (2007a) for details.

¹⁰ See Emrouznejad and Yang (2018) for a recent literature review.

dataset includes some discrete variables, we could use discrete kernels, although the most preferred option is to use the standard continuous kernels proposed by Racine and Li (2004) and Li and Racine (2007) to smooth all components of Z (see Badin and Daraio, 2011 for details).¹¹

Following Badin et al. (2012), we can also investigate the impact of Z on the reachable frontier independently of the efficiency values by analyzing the ratios of conditional to unconditional order- m estimators:

$$\hat{Q}_m = \frac{\hat{\lambda}_m(x, y|z)}{\hat{\lambda}_m(x, y)} \quad (7)$$

As explained in Badin et al. (2012), we can explore the effects of conditional variables on the boundary (shift of the frontier in the output direction) by setting a large value of m . This provides a robust estimator of the full frontier. In contrast, for small values of m , we examine the behavior of the shift towards the center of the distribution of efficiencies. For instance, if $m = 1$, the order- m frontier is simply an average production function and the ratios analyze the shift of the mean distribution of inefficiencies.¹²

This ratio can be nonparametrically regressed on a variable of interest in Z . Graphically, the slopes of the smoothed regression lines provide an interpretation of both marginal effects. In an output-oriented conditional model, a regression line with a positive slope indicates that the variable in Z has a positive effect on efficiency, whereas a downward-sloping line denotes a negative effect. Moreover, we can also investigate the statistical significance of Z explaining the variations of the ratio by applying the nonparametric bootstrap procedure described in Racine (1997) and Li and Racine (2007) to explain the variations of the ratio. This procedure can be understood as the nonparametric equivalent of standard t-tests in ordinary least squares regression (De Witte and Kortelainen, 2013). As pointed out by Daraio and Simar (2014), this procedure may be hard to apply because the dependent variable (ratio of efficiency estimators) is not actually observed and has to be replaced by estimates. As a result, this approach is not generally consistent when using full frontier estimates like FDH or DEA because of the well-known *curse of dimensionality* (Kneip et al., 2015). Fortunately, this is not a serious limitation when using partial frontiers of order- m as in our case, since their rates of convergence do not depend on the number of inputs and outputs.

Finally, note that just analyzing the ranking of conditional efficiency measures might be misleading, since schools actually are operating facing different contextual conditions, thus the maximum output defined by the production frontier could be achieved easier under some specific conditions with respect to others. In order to obtain a pure managerial efficiency measure we have applied the second-stage approach suggested by Badin et al. (2012), i.e. we regress conditional efficiency measures on the relevant explanatory variables $\mu(Z) = E(\lambda(X, Y|Z)|Z = z)$ and $\sigma^2(Z) = V(\lambda(X, Y|Z)|Z = z)$ using the following flexible location-scale regression model:

$$\lambda(X, Y|Z = z) = \mu(z) + \sigma(z)\varepsilon \quad (8)$$

where $E(\varepsilon|Z = z) = 0$ and $V(\varepsilon|Z = z) = 1$. In Eq. (8) $\mu(z)$ indicates the average effect of z on the conditional efficiency and $\sigma(z)$ provides additional information on the dispersion of the efficiency distribution as a function of the contextual variables. The estimated residuals yielded by this procedure (ε) can be viewed as the part of the efficiency score remaining after eliminating its dependence on the contextual variables, that is, as a cleansed version of the conditional efficiency referred to as *pure managerial efficiency* (Badin et al., 2012). These values are

standardized (mean zero and variance one) and can be used to make comparisons among schools facing different operating conditions. A large value of ε will denote poor results, whereas a small (or even negative) value indicates that the school's managerial performance is good. In our empirical analysis, this model will be estimated nonparametrically.¹³

4. Data and variables

In order to estimate an educational worldwide production frontier, we use data about schools operating in different countries participating in PISA 2012. This dataset provides international comparative data about 15-year-old students' competences in three main areas: reading, mathematics and science. In addition to those achievement measures, the database contains a vast amount of information about students' background and the characteristics of the schools that they attend. This information is sourced from student and school principal questionnaire responses. In addition, as we also wanted to account for the economic and social characteristics of the countries in which schools are operating, as well as some key institutional features, we gathered data from several additional sources. Specifically, we collected data about economic indicators from the World Bank Open Data section, whereas social indicators regarding countries' cultural values were retrieved from the longitudinal aggregate data of the WVS covering the period from 1981 to 2014.¹⁴

Although there were 65 participating countries (all 34 OECD members and 31 partners) in PISA 2012, we were unable to include all these countries in our empirical study because, for many, information about institutional features or cultural values (some countries did not participate in WVS) were missing. As a result, our dataset comprises schools from 36 countries (26 OECD countries and 10 partners) representing more than 70% of the world economy. The total number of schools is 12,264 distributed across countries as reported in Table 1.

In line with previous literature (e.g. Charchye et al., 2010; De Witte and Kortelainen, 2013), we used the results achieved by students in the three competences evaluated in PISA (mathematics, reading and science) aggregated at school level as output indicators. Rather than a single measure of achievement, the PISA dataset provides five plausible values for each discipline (estimated by considering a plausible distribution according to the test results).¹⁵ For the sake of simplicity, we consider only one plausible value (the first one) for each subject (PVMATH, PVREAD and PVSCI) in our analysis because these values provide both unbiased point and sampling variance estimates.¹⁶

The selection of the variables to be included as inputs is a major challenge in empirical studies using data from international large-scale assessments since the list of potential indicators is usually very long. In this respect, most empirical studies designed to measure school efficiency usually include some measures of human and capital resources (e.g. Agasisti and Zoido, 2015; Aparicio et al., 2018a; Crespo-Cebada et al., 2014; Santín and Sicilia, 2015). In this case, we decided to use the inverse of the student-teacher ratio, i.e. the number of teachers per (a hundred) students (TEACHERS) as a proxy for human resources, since education is labor intensive, and the index of the quality of school resources

¹¹ This approach has been applied in other studies using robust conditional indicators (e.g. Cordero et al., 2015; Tzeremes, 2015; Verschelde and Rogge, 2012).

¹² Mastromarco and Simar (2017) use similar values of the parameter m to investigate the influence on the shape of the frontier and the distribution of efficiencies.

¹³ We use nonparametric local constant models and bandwidths determined by the usual LSCV (Pagan and Ullah, 1999).

¹⁴ This dataset includes data from six different waves of the WVS: 1981–84, 1990–94, 1995–98, 1999–2004, 2005–2009 and 2010–2014 (<http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp>).

¹⁵ See Wu (2005) for a detailed discussion about the role of plausible values in large-scale surveys.

¹⁶ The use of one plausible value or five plausible values does not really make a substantial difference in large samples (see OECD, 2009, p. 44 for details).

Table 1

Dataset composition: number of schools by country.

Country	Observations	Country	Observations
Australia	775	Netherlands	179
Canada	883	New Zealand	176
Chile	220	Norway	197
Colombia	352	Poland	184
Czech Republic	295	Romania	178
Estonia	206	Russian Federation	226
Finland	311	Singapore	172
France	226	Slovak Republic	231
Germany	230	Slovenia	330
United Kingdom	507	Spain	902
Hong Kong	148	Sweden	209
Indonesia	209	Switzerland	411
Israel	172	Thailand	239
Italy	1192	Tunisia	153
Japan	191	Turkey	169
Korea	156	Uruguay	180
Lithuania	215	United States	161
Latvia	210	TOTAL	12,264
Mexico	1469		

(SCMATEDU) as a proxy for capital resources.¹⁷ We also included an additional variable representing the average socioeconomic status (ESCS)¹⁸ of students attending a school as a proxy of the quality of the school's students.¹⁹

We have divided the variables representing the school operating environment into three main groups. The first one includes a mixed set of six indicators derived from school principal responses with respect to different school resources and factors representing the school environment previously identified in the literature as relevant determinants of student performance: an index available in the PISA database representing the ratio of computers for education to the number of students in the grade corresponding to 15-year-old students (COMPUTERS), an index of disciplinary climate in the classroom²⁰ (DISCLIM), the proportion of fully certified teachers with respect to the total number of teachers (PROPCERT), the number of students enrolled at the school (SCHSIZE), an index of the relative level of responsibility of school staff in allocating

resources²¹ (RESPRES), and a dummy variable representing whether the school has other competing schools in the same area (COMPETITION).

The second group includes five continuous country-level indicators of economic and cultural aspects. The economic variables are gross domestic product (GDP) per capita and public expenditure per student on secondary education as a percentage of GDP per capita for each country in the year 2012 (EXPEDUC). We can use GDP per capita to investigate whether efficiency varies with income level, whereas public expenditure per student should identify countries attaching more importance to education. With regard to each country's cultural values, our source of information is, as mentioned above, the WVS. In particular, we use information taken from responses to a set of questions about which qualities are most valued when raising a child. Respondents are given a list of ten qualities that children can be encouraged to learn at home and then asked to select the five qualities that they think are most important. Among them, we selected the responses for three aspects such as hard work (HARDWORK), responsibility (RESP) and perseverance (PERSEV). These three dimensions are directly related to the concept of “conscientiousness”, since this is the only personality traits dimension that is positively correlated with educational outcomes according to the literature (see Borghans and Schils, 2011; Heckman, 2011 for details).

Finally, the third group of variables includes four indicators representing relevant institutional features that do or do not exist in the country. These variables are used to test whether or not their influence on efficiency is consistent with previous existing evidence about their effect on academic outcomes. Specifically, we considered whether the country implements early tracking (TRACKING) (OECD, 2010, 2011) and central examinations (EXAMS) (Bol et al., 2014; Eurydice, 2004; Woessmann et al., 2009). In addition, we also include other average country-level indicators like the length of pre-primary education (PREPRIM) or the proportion of private schools (PRIVATE) provided by the UNESCO Institute for Statistics. Table 2 shows the main descriptive statistics of all the above variables classified in different categories.

5. Results

In this section, we report the efficiency scores estimated using the robust order- m model with an output orientation. We opted for an output orientation because schools strive to maximize student achievement and cannot easily reduce their inputs at least in the short term. First, we estimate the unconditional estimator defined in equation (4). In line with Daraio and Simar (2005), we determine the size of the partial frontier as the value of m for which the decrease in the number of super-efficient observations stabilizes. In our application, this corresponds to $m = 150$.²² As a result, each school is compared to 150 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. Table 3 reports the ranking of countries according to the average estimated efficiency scores of their schools together with the corresponding standard deviation (in brackets), thus identifying the degree of variability within each country.

The last row in Table 3 shows that the average overall efficiency value for all the schools included in the sample is 1.2787. This indicates that if all schools were to perform as efficiently as the top performers, the test scores could increase on average by almost 28%, although variability across countries is high. Note also that around 1% of schools have an efficiency score below one. These super-efficient schools perform better than the 150 schools against which they are benchmarked. With regard to the average scores of units operating in different education systems, we observe that mainly Asian countries (Hong Kong, Korea, Thailand, Singapore and Japan) occupy the top positions, whereas the worst performers are mostly Eastern European countries (Slovenia, Lithuania and

¹⁷ This index was created by PISA analysts from the responses given by school principals regarding several educational resources such as computers, educational software, calculators, books, audiovisual resources or laboratory equipment. Since the original values included some negative values, they were rescaled to assure that all values are positive.

¹⁸ This index provides a measure of family background, including the highest levels of parents' occupation, educational resources and cultural possessions at home. The original values of this variable have also been rescaled to assure that all values are positive.

¹⁹ Although socioeconomic status unquestionably plays a role in explaining student and school educational results (Haveman and Wolfe, 1995), some studies (e.g. Agasisti, 2014; Aparicio et al., 2017; Thieme et al., 2013) consider this variable as an input, whereas other authors suggest that it be included as a nondiscretionary input (see Liberati et al., 2017; Ruggiero and Vitaliano, 1999). In this paper, we have included this variable as a traditional input. Accordingly, schools are evaluated according to their ability to make the most of their inputs (see Camanho et al., 2009 for details).

²⁰ This index is based on the responses provided to five items: (i) Students don't listen to what the teacher says; (ii) There is noise and disorder; (iii) The teacher has to wait a long time for students to quiet down; (iv) Students cannot work well, and (v) Students don't start working for a long time after the lesson begins. The four response categories were “every lesson”, “most lessons”, “some lessons”, to “never or hardly ever” (OECD, 2014, p. 331).

²¹ This index was derived from six items: the school principals' report regarding who had considerable responsibility for tasks related to resource allocation (selecting teachers for hire, firing teachers, establishing teachers' starting salaries, determining teachers' salary increases, formulating the school budget and deciding on budget allocations within the school) (OECD, 2014, p. 310).

²² We also tested two alternative values of m ($m = 100$ and $m = 200$) and the findings were similar.

Table 2
Descriptive statistics of variables included in the analysis.

Variable	Type	Mean	SD	Min	Max
PVMATH	Output	473.74	72.79	130.06	753.08
PVREAD	Output	475.57	72.05	155.09	716.68
PVSCI	Output	481.64	71.83	92.41	728.87
TEACHERS	Input	8.75	6.28	0.10	93.11
SCMATEDU	Input	3.53	1.09	0.01	5.58
ESCS	Input	4.27	0.83	0.16	6.24
School factors					
COMPUTERS	Continuous	0.73	0.94	0.00	30.00
DISCLIM	Continuous	2.49	0.48	0.01	4.34
PROPCERT	Continuous	0.80	0.33	0.00	1.00
SCHSIZE	Continuous	715.94	587.54	45.00	7600
RESPRES	Continuous	0.74	0.98	0.01	3.52
COMPETITION	Dummy	0.74	0.44	0.00	1.00
Country features					
GDP pc	Continuous	31,264	13,049	4876	65,640
EXPEDUC	Continuous	22.58	6.52	9.60	37.41
HARDWORK	Continuous	0.44	0.21	0.09	0.89
RESP	Continuous	0.75	0.10	0.26	0.91
PERSEV	Continuous	0.40	0.10	0.06	0.61
TRACKING	Dummy	0.34	0.47	0.00	1.00
EXAMS	Dummy	0.69	0.46	0.00	1.00
PREPRIM	Ordered discrete	2.68	0.79	1.00	4.00
PRIVATE	Continuous	0.19	0.20	0.00	0.97

Table 3
Ranking of countries according to their average efficiency scores (unconditional model) and their results for mathematics (standard deviation in brackets).

Country	Efficiency (unconditional model)	Results for mathematics
Hong Kong	1.0920 (0.106)	558.94 (65.66)
Turkey	1.1145 (0.085)	438.12 (75.33)
Thailand	1.1616 (0.083)	441.04 (68.72)
Korea	1.1657 (0.104)	550.84 (63.54)
Singapore	1.1758 (0.092)	568.22 (64.09)
Mexico	1.1821 (0.089)	410.11 (48.48)
Colombia	1.2088 (0.079)	381.08 (48.55)
Japan	1.2123 (0.103)	533.82 (70.84)
Poland	1.2310 (0.076)	526.54 (57.47)
Estonia	1.2423 (0.064)	518.06 (40.15)
New Zealand	1.2615 (0.083)	496.18 (53.77)
Netherlands	1.2716 (0.129)	517.31 (76.10)
Indonesia	1.2744 (0.094)	373.72 (49.41)
Canada	1.2775 (0.076)	507.21 (45.78)
Chile	1.2802 (0.087)	431.76 (68.34)
Russian Federation	1.2805 (0.084)	476.18 (49.23)
United States	1.2833 (0.080)	480.44 (47.74)
Spain	1.2877 (0.067)	490.95 (46.17)
Germany	1.2885 (0.105)	506.51 (75.07)
Czech Republic	1.2886 (0.103)	501.20 (77.57)
Finland	1.2900 (0.074)	508.38 (47.79)
Latvia	1.2975 (0.065)	486.24 (46.43)
Switzerland	1.3065 (0.084)	514.04 (55.50)
United Kingdom	1.3099 (0.079)	487.98 (50.17)
Tunisia	1.3151 (0.111)	384.20 (55.74)
France	1.3184 (0.071)	487.49 (76.28)
Australia	1.3215 (0.082)	492.21 (59.30)
Slovak Republic	1.3264 (0.102)	473.04 (68.30)
Norway	1.3307 (0.067)	490.78 (41.08)
Italy	1.3417 (0.102)	476.15 (72.02)
Israel	1.3421 (0.112)	463.89 (70.78)
Sweden	1.3465 (0.077)	483.07 (48.46)
Romania	1.3562 (0.088)	443.42 (58.95)
Lithuania	1.3606 (0.073)	468.74 (54.46)
Uruguay	1.3906 (0.090)	402.88 (61.46)
Slovenia	1.4329 (0.105)	463.87 (73.38)
TOTAL	1.2787 (0.159)	473.74 (72.79)

Standard deviation in brackets.

Table 4
Descriptive statistics for the whole sample with alternative models.

	Unconditional model	Conditional model 1	Conditional model 2
Mean	1.2787	1.1243	1.0683
SD	0.1589	0.1318	0.1047
Min	0.9836	0.7201	0.7274
1 st Q	1.1412	1.0326	1.0000
Median	1.2325	1.1019	1.0496
3 rd Q	1.3297	1.1837	1.1201
Max	2.6984	2.3361	2.0868
Efficient units	94 (0.77%)	355 (2.89%)	527 (4.30%)
Super-efficient units	111 (0.91%)	1543 (12.58%)	2825 (23.03%)

Romania). If we compare the ranking of countries according to the unconditional efficiency levels with their ranking according to their average results in mathematics, we find considerable disparities²³. For instance, Turkey, Thailand, Mexico and Colombia are ranked among the top performers in terms of efficiency levels, even though their students achieve results that are well below average. Thus, the position of some countries in the ranking improves substantially when we account for the educational process inputs in the efficiency analysis.

The main problem with this initial assessment is that it is based on the unrealistic assumption that all the evaluated schools are operating in the same context. Therefore, the estimated scores may not properly represent the school's level of efficiency. In other words, the attainable set is not dependent on Z. However, this assumption is unlikely to hold since the perceived frontier might be different depending on each school's operating environment. Therefore, the next step in our empirical analysis is to take into account the existing heterogeneity among schools in our estimation. To do this, we estimated two alternative conditional efficiency measures according to Eq. (5). Model 1 accounts for the set of variables related to the school-level environment. Subsequently, we estimated a second conditional model (Model 2) with an additional set of country-level variables, including economic indicators, indices representing cultural values, and institutional characteristics.

Table 4 summarizes the main descriptive statistics (mean, median, standard deviation, maximum, quartile values and the number of efficient and super-efficient units) of estimates for both models compared to the values of the unconditional model. As expected, when the analysis includes information about contextual factors, the overall mean inefficiency for both models (1.1243 for Model 1 and 1.0683 for Model 2) decreases and the number of efficient and super-efficient units increases. In fact, all schools increase their efficiency under a 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 illustrated in Fig. 1 shows a high concentration of units at the higher efficiency levels for Model 2, whereas this concentration is not as pronounced for Model 1.

Nevertheless, the most interesting conclusions can be drawn by exploring the divergences in efficiency levels across countries. Table 5 reports the rankings of countries according to the average estimated scores of their schools with each model. One key finding is that there are no significant changes to classification with respect to the unconditional model when school contextual factors are incorporated in Model 1, although the within-country variability is greater in most cases. In fact the value of the Spearman correlation coefficient between those models shown in Table 6 is quite high (0.725). The only notable changes are that countries like Poland, Finland or Switzerland move up and the United States, Colombia or Indonesia move down in the ranking.

²³ We use test scores for mathematics in the comparison because this was the main competence assessed in PISA 2012. In any case, they are highly correlated with the other competences (science and reading).

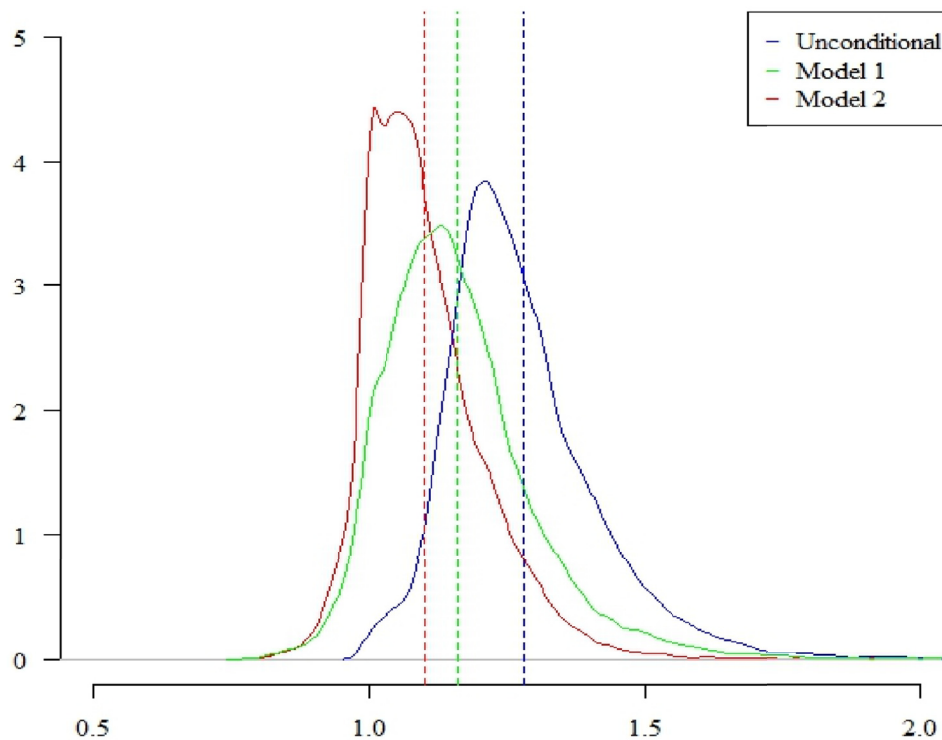


Fig. 1. Probability density of total efficiency scores with alternative models.

In contrast, the picture for Model 2 accounting for country features is quite different, since the Spearman correlation coefficient with the unconditional case now has a much lower value (0.508). A case in point is Uruguay, which moves up from the bottom position in the ranking according to Model 1 to second place according to Model 2. A similar result applies to other countries, such as Romania or Tunisia, which also improve when country-level factors are taken into account. In contrast, the consideration of country-level factors has a negative effect in terms of the ranking on other countries, like Mexico, Japan, Spain, the United States or the Czech Republic. Likewise, Hong Kong is ejected from the top position in the ranking and is now ranked in seventh place. The importance of those changes leads us to presume that the heterogeneity among different countries is more relevant than the heterogeneity among schools within the same country.

In order to examine the influence of the above external factors on efficiency estimates, we regress the ratio between the conditional and the unconditional efficiency scores on Z . As there are two different models (Model 1 and Model 2), we estimated a nonparametric regression for each model. Table 7 shows the p-values of the significance tests proposed by Li and Racine (2007) and Racine and Li (2004) after performing 1000 bootstrap samples. For Model 1, only two contextual school variables appear to have a significant effect on explaining the level of inefficiency, although their significance level is low (10%). This finding confirms our suspicion with regard to the little relevance of these variables as potential factors affecting the efficiency of schools. Actually, this result is confirmed by the lack of significance for most school variables in the estimation conducted for Model 2, in which we also account for heterogeneity across countries. However, we find that all the economic and cultural factors, as well as almost all the variables representing specific country features, have a significant influence on efficiency levels. Therefore, we can conclude that country factors appear to play a more significant role in explaining differences in school performance because there is more heterogeneity across countries than among schools within the same country.

The importance of economic indicators in explaining differences in efficiency performance across countries is consistent with some previous

results in the literature (Afonso and St Aubyn, 2006; Cordero et al., 2017b). Likewise, cultural or personality factors have also been considered as key determinants of educational performance in some recent cross-country studies (e.g. Conti et al., 2011; Méndez, 2015). Additionally, we found that cultural factors linked to personality traits (conscientiousness) are also significant variables affecting efficiency estimates. Based on this result, more attention should be paid to non-cognitive skills in order to boost efficiency in education. With regard to specific country-level educational features, the only factor that seems to have a significant influence on efficiency is the proportion of private schools in the country, whereas the length of pre-primary education and the existence of tracking and central examinations do not appear to affect school performance in terms of efficiency, even though, as explained in Section 2, they have been identified as important determinants of academic achievement in previous literature.

As described in Section 3, we investigate how the ratios \hat{Q}_m shown in Eq. (7) can provide information about the potential effects of contextual factors on shifts in the frontier and the distribution of inefficiencies. Fig. 2 shows the marginal effect of contextual factors with a significant effect. The left-hand side of Fig. 2 shows the effect on shifts in the frontier based on full frontier ratios (with a large value of m , in this case $m = 12,264$), while the right-hand side illustrates the effect inside the attainable sets, i.e. based on partial ratios with $m = 1$.

Starting with school variables, we observe different effects according to the plotted marginal views. On the one hand, we find that disciplinary climate in the classroom (Fig. 2a) has a definite favorable effect on the frontier, which is not so clear in the center of the distribution of the efficiencies (Fig. 2b). Likewise, the effect is rather blurred in both cases for the PROPCERT variable (Fig. 2c and d), since there are many variable values concentrated on the right-hand side of the graphs given the high proportion of fully certified teachers existing in most countries.

Regarding the two economic indicators (GDPpc and expenditure on secondary education), we find that both variables have slight negative effects on the frontier (Fig. 2e and g), whereas only GDPpc has a clear and negative effect on the distribution of efficiencies (Fig. 2f). For the variables representing cultural values associated with *conscientiousness*, we

Table 5

Average efficiency scores for each country with alternative models.

Unconditional model		Conditional model 1		Conditional model 2	
Country	Avg. score	Country	Avg. score	Country	Avg. score
Hong Kong	1.0920 (0.106)	Hong Kong	1.0403 (0.111)	New Zealand	1.0117 (0.098)
Turkey	1.1145 (0.085)	Poland	1.0542 (0.091)	Uruguay	1.0122 (0.063)
Thailand	1.1616 (0.082)	Turkey	1.0599 (0.103)	Chile	1.0125 (0.098)
Korea	1.1657 (0.104)	Mexico	1.0692 (0.118)	Romania	1.0142 (0.105)
Singapore	1.1758 (0.092)	Korea	1.0719 (0.121)	Poland	1.0181 (0.067)
Mexico	1.1821 (0.089)	Estonia	1.0853 (0.074)	Colombia	1.0209 (0.094)
Colombia	1.2088 (0.079)	Finland	1.0884 (0.102)	Hong Kong	1.0219 (0.084)
Japan	1.2123 (0.103)	New Zealand	1.1003 (0.105)	Tunisia	1.0221 (0.106)
Poland	1.2310 (0.076)	Switzerland	1.1011 (0.109)	Indonesia	1.0260 (0.106)
Estonia	1.2423 (0.064)	Germany	1.1015 (0.141)	Canada	1.0280 (0.073)
New Zealand	1.2615 (0.083)	Czech Rep.	1.1070 (0.136)	Netherlands	1.0311 (0.160)
Netherlands	1.2716 (0.129)	Latvia	1.1070 (0.111)	Turkey	1.0315 (0.087)
Indonesia	1.2744 (0.094)	Canada	1.1073 (0.103)	Switzerland	1.0347 (0.090)
Canada	1.2775 (0.076)	Norway	1.1098 (0.103)	Thailand	1.0394 (0.083)
Chile	1.2802 (0.087)	Singapore	1.1099 (0.103)	Germany	1.0429 (0.102)
Russian Fed.	1.2805 (0.084)	Chile	1.1111 (0.116)	Singapore	1.0471 (0.068)
United States	1.2833 (0.080)	Thailand	1.1128 (0.104)	UK	1.0491 (0.082)
Spain	1.2877 (0.067)	Japan	1.1136 (0.122)	Australia	1.0511 (0.105)
Germany	1.2885 (0.105)	Slovak Rep.	1.1184 (0.150)	Korea	1.0517 (0.083)
Czech Rep.	1.2886 (0.103)	Netherlands	1.1200 (0.172)	Finland	1.0639 (0.066)
Finland	1.2900 (0.074)	Sweden	1.1212 (0.112)	Mexico	1.0705 (0.076)
Latvia	1.2975 (0.064)	France	1.1252 (0.179)	Norway	1.0707 (0.079)
Switzerland	1.3065 (0.083)	Colombia	1.1301 (0.106)	Estonia	1.0720 (0.045)
UK	1.3099 (0.079)	Spain	1.1301 (0.089)	Slovak Rep.	1.0743 (0.109)
Tunisia	1.3151 (0.112)	Russian Fed.	1.1305 (0.104)	Latvia	1.0770 (0.073)
France	1.3184 (0.071)	Italy	1.1563 (0.154)	Czech Rep.	1.0789 (0.088)
Australia	1.3215 (0.082)	Australia	1.1650 (0.126)	Sweden	1.0835 (0.085)
Slovak Rep.	1.3264 (0.102)	Slovenia	1.1673 (0.162)	Slovenia	1.0897 (0.131)
Norway	1.3307 (0.067)	Indonesia	1.1677 (0.142)	Russian Fed.	1.0921 (0.067)
Italy	1.3417 (0.102)	Tunisia	1.1736 (0.180)	France	1.0929 (0.108)
Israel	1.3421 (0.112)	United States	1.1756 (0.143)	Japan	1.1027 (0.088)
Sweden	1.3465 (0.077)	Israel	1.1837 (0.182)	Spain	1.1190 (0.058)
Romania	1.3562 (0.088)	UK	1.1888 (0.107)	Italy	1.1283 (0.106)
Lithuania	1.3606 (0.073)	Romania	1.1898 (0.125)	Israel	1.1284 (0.143)
Uruguay	1.3906 (0.090)	Lithuania	1.1930 (0.082)	United States	1.1428 (0.129)
Slovenia		Uruguay		Lithuania	

Table 5 (continued)

Unconditional model		Conditional model 1		Conditional model 2	
Country	Avg. score	Country	Avg. score	Country	Avg. score
	1.4329 (0.106)		1.2186 (0.109)		1.1524 (0.076)
TOTAL	1.2787 (0.159)	TOTAL	1.1243 (0.132)	TOTAL	1.0683 (0.105)

Table 6

Spearman correlation coefficients.

	Unconditional Model	Conditional Model 1	Conditional Model 2
Unconditional Model	1.000		
Conditional Model 1	0.725	1.000	
Conditional Model 2	0.508	0.688	1.000

Table 7

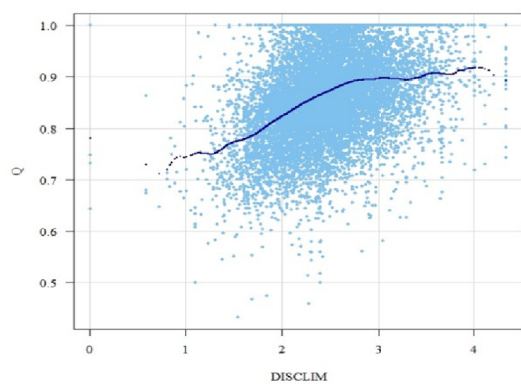
Influence of different factors on educational performance (Estimation of nonparametric significance tests).

School variables	Model 1	Model 2
	p-value	p-value
COMPUTERS	0.99	0.65
DISCLIM	0.05 ^b	0.00 ^a
PROPCERT	0.06 ^b	0.06 ^b
SCHSIZE	0.99	0.30
RESPRES	0.99	0.49
COMPETITION	0.99	0.58
Country variables		p-value
GDP pc		0.00 ^a
EXPEDUC		0.00 ^a
HARDWORK		0.00 ^a
RESP		0.00 ^a
PERSEV		0.00 ^a
TRACKING		0.42
EXAMS		0.68
PREPRIM		0.52
PRIVATE		0.00 ^a

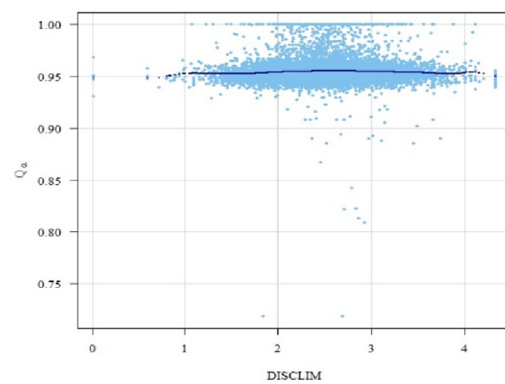
^a Denotes statistical significance at 1%.^b Denotes statistical significance at 10%.

observe very different behaviors. Thus, HARDWORK has a positive effect on the efficient frontier concentrated around higher values, but its effect on the distribution of efficiencies is unclear (Fig. 2i and j). The plots representing the effect of responsibility (RESP) and perseverance (PERSEV) are quite similar, since they have a negative effect on the frontier (Fig. 2k and m), especially for lower values, but an indefinite effect on the distribution of efficiencies (Fig. 2l and n). Finally, we observe that the proportion of private schools (PRIVATE) has a small negative effect on the frontier (Fig. 2o), whereas there is no clear pattern for the distribution of efficiencies (Fig. 2p).

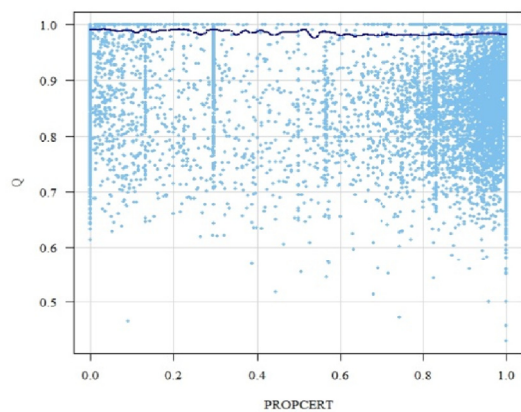
Finally, with the aim of providing some additional insights into the effects of exogenous variables on conditional efficiency scores, we also present the results of the second-stage analysis using the flexible local constant estimator described in Eq. (8). Specifically, the estimated location ($\hat{\mu}$) and scale ($\hat{\sigma}$) effects of the proposed regression are reported in Fig. 3, where the left-hand panels show the estimated location and the right-hand panels represent the scale. As a whole, the school variable (DISCLIM) appears to have a positive effect on location (Fig. 3a), while it is homogeneously distributed with regard to the scale, whereas PROPCERT presents a high concentration on higher values, but its effect is not clear, confirming that its influence on efficiency levels is relatively limited. The distribution for the two economic indicators has an inverted



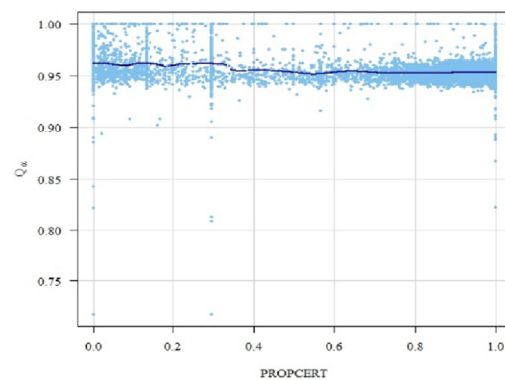
(a)



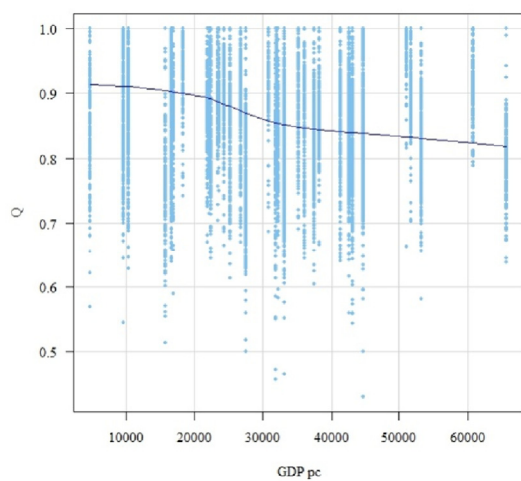
(b)



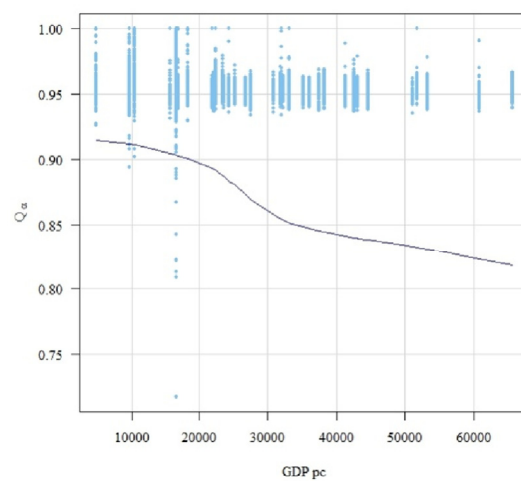
(c)



(d)

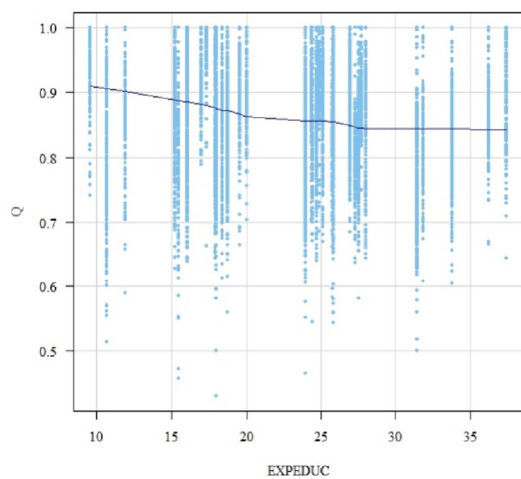


(e)

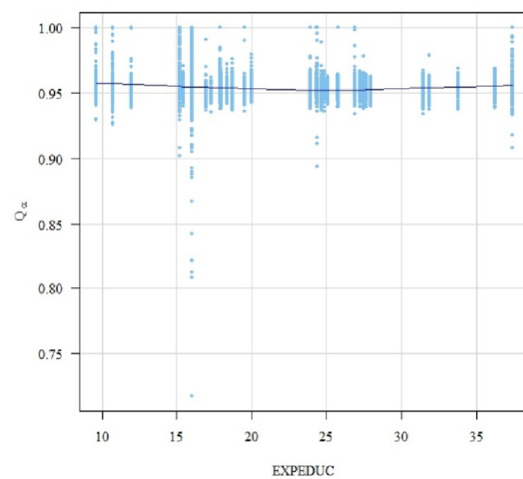


(f)

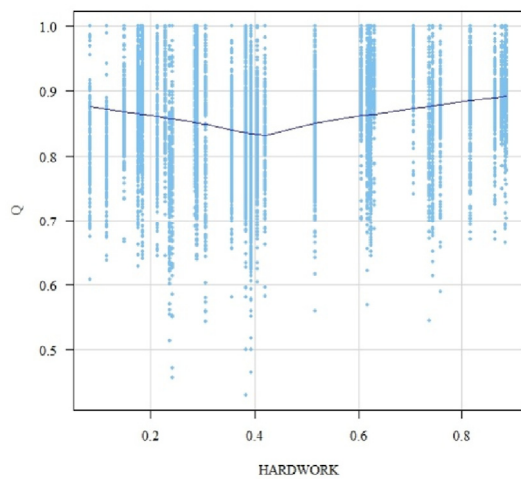
Fig. 2. Scatterplots of ratios vs. Z_s .



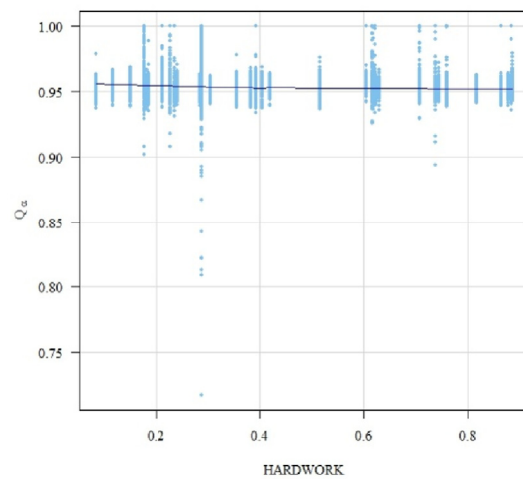
(g)



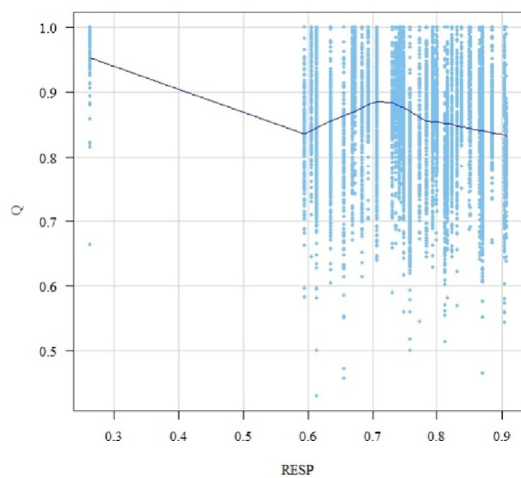
(h)



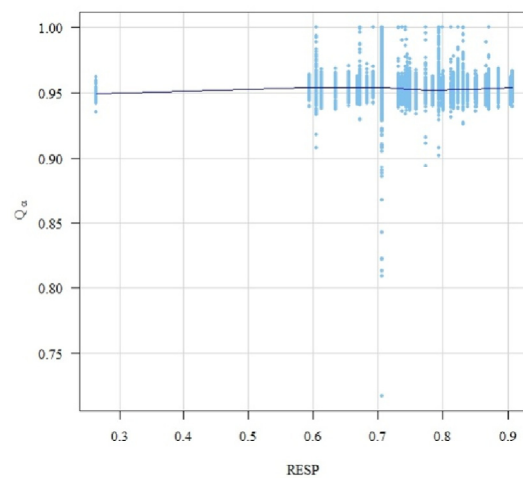
(i)



(j)



(k)



(l)

Fig. 2. (continued).

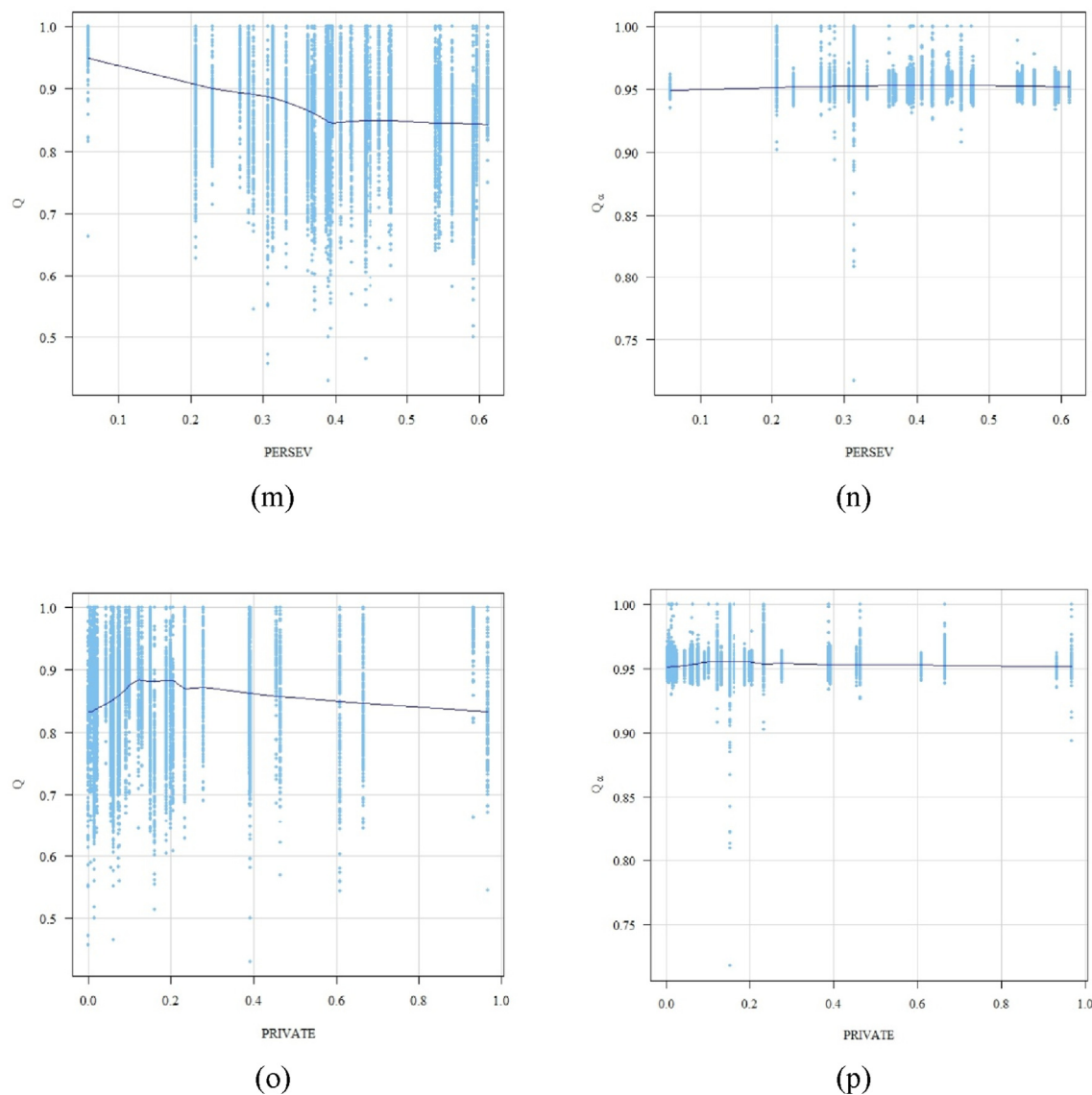


Fig. 2. (continued).

u-shape (Fig. 3e–h). Therefore, the effect might be favorable for lower, and unfavorable for higher values. However, the effect might not always be negative as the previous analysis suggested. The effect for HARDWORK (3i, 3j) appears to be slightly positive, while RESP (Fig. 3k and l) and PERSEV (Fig. 3m and n) have an inverted u-shape. Finally, the effect appears to be unfavorable for the PRIVATE variable (Fig. 3o and p), which confirms the result reported above.

Fig. 4 shows the empirical distribution of the residuals ($\hat{\varepsilon}_i$), i.e. *pure managerial efficiency*. The histogram represents a truncated normal distribution that is more highly concentrated below zero (where observations are more efficient than the average in the output orientation). *Pure managerial efficiency* can be regarded as an indicator that provides better results for ranking units according to their efficiency levels, since it is totally clean of the effects of contextual factors. Consequently, we can build another ranking of countries according to the average *pure efficiency* of their schools, representing scores “cleansed” of the heterogeneous conditions in their operating context (Table 8). In this hypothetical situation where efficiency is now fully comparable across schools and countries, we identify Korea, the Czech Republic, Germany, the Slovak Republic and the United States as the best performers, i.e. having the lowest (or even negative) mean values for the residuals, whereas France, Uruguay, Romania, Lithuania and Latvia have the highest residual values

and are, therefore, the least efficient.

6. Concluding remarks

The analysis of the efficiency of educational systems and their schools is one of the most interesting topics in the policy and academic debate for two reasons. On one hand, the academic results obtained by high schools as measured in standardized tests is a common way of measuring quality across different schools and countries. The main conclusion of this literature shows that a greater overall quality in the educational system of a country has a strong and stable association with greater economic growth rates (Hanushek and Kimko, 2000). On the other hand, in the last decades most countries have made a huge financial effort in public education although there is no straight positive correlation between higher per capita public expenditures on education and higher academic outcomes (Hanushek, 2003). For these reasons benchmarking schools and analyzing their efficiency worldwide is one the most promising tools to learn from best managerial practices. Moreover, this analysis may help policy makers to discarding educational policies that do not work and reallocating this public expenditure in more promising alternatives.

In this paper, we applied some of the most recent nonparametric methods to assess the performance of secondary schools operating in 36

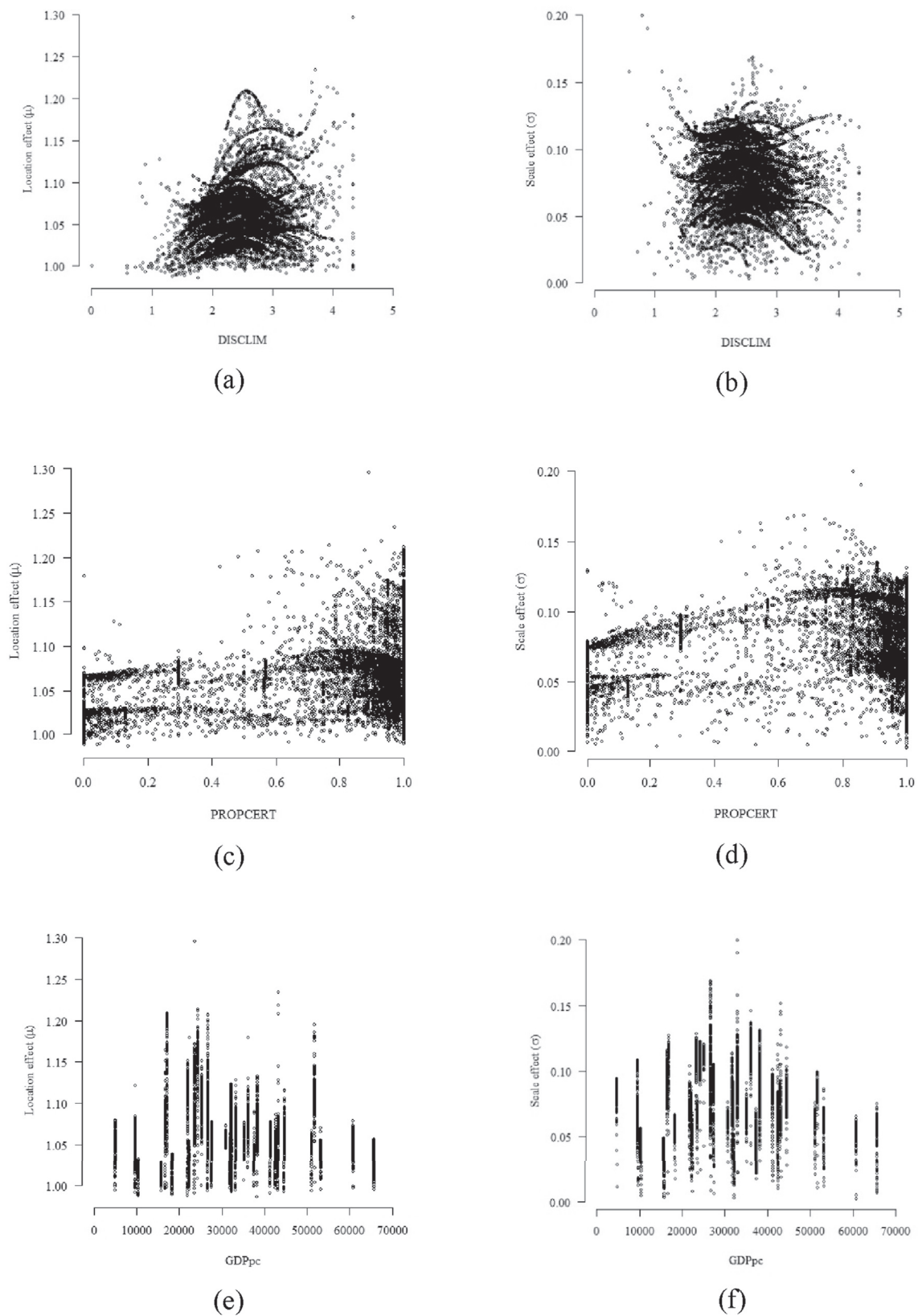
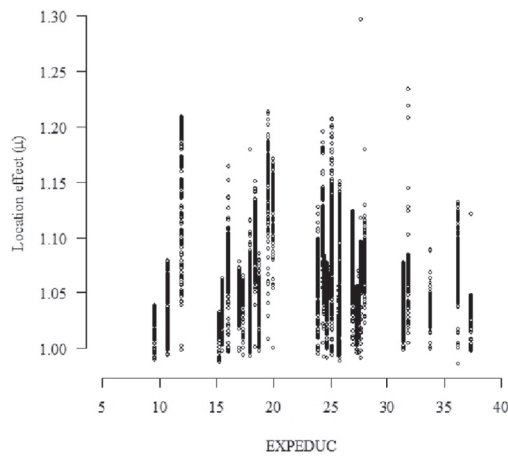


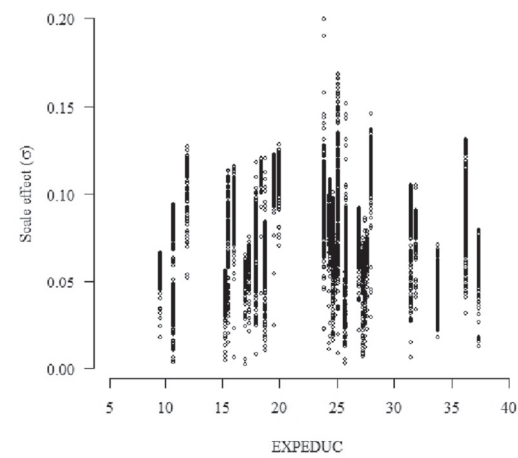
Fig. 3. Scatterplots of “location” (μ) and “scale” (σ) effects vs. exogenous variables.

different countries (26 OECD countries and 10 partners) using data from PISA 2012. Specifically, we employed the robust conditional model developed by [Daraia and Simar \(2005, 2007a, 2007b\)](#), which accounts for the existing heterogeneity among schools, as well as for potential

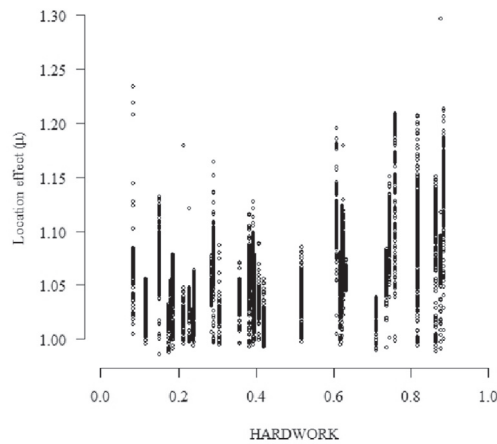
specific country-level features that can affect school performance. The main advantage of this approach, which has been little used in previous educational literature, is that our estimated frontier incorporates the effect of contextual variables at school and country level. Therefore, we



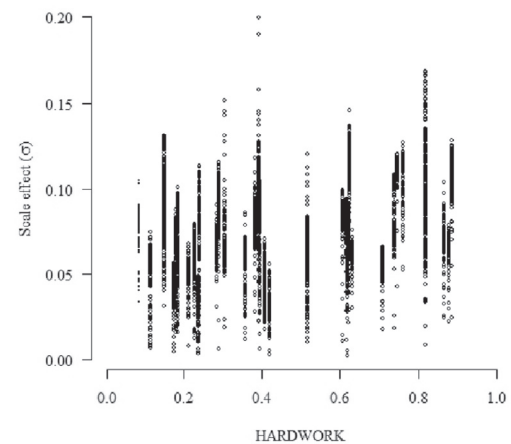
(g)



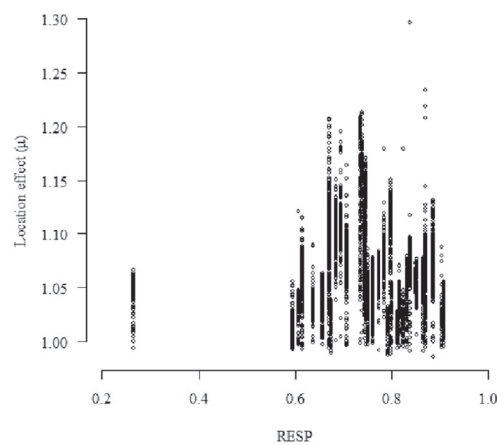
(h)



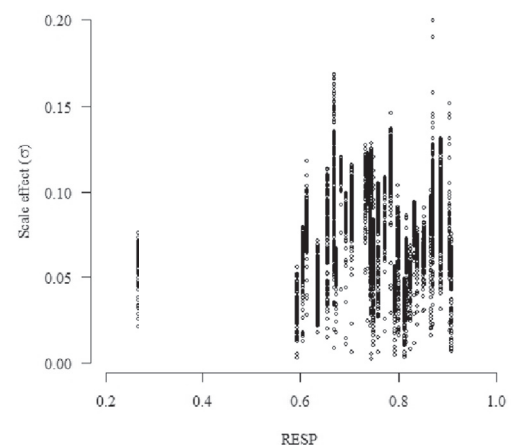
(i)



(j)



(k)



(l)

Fig. 3. (continued).

can explore the influence of contextual factors without assuming the usual separability condition implicit in most previous studies using the well-known second-stage approach.

Our results reveal several interesting issues. First, the ranking of countries according to school performance, represented by students' PISA test scores, changes substantially when the efficiency analysis takes

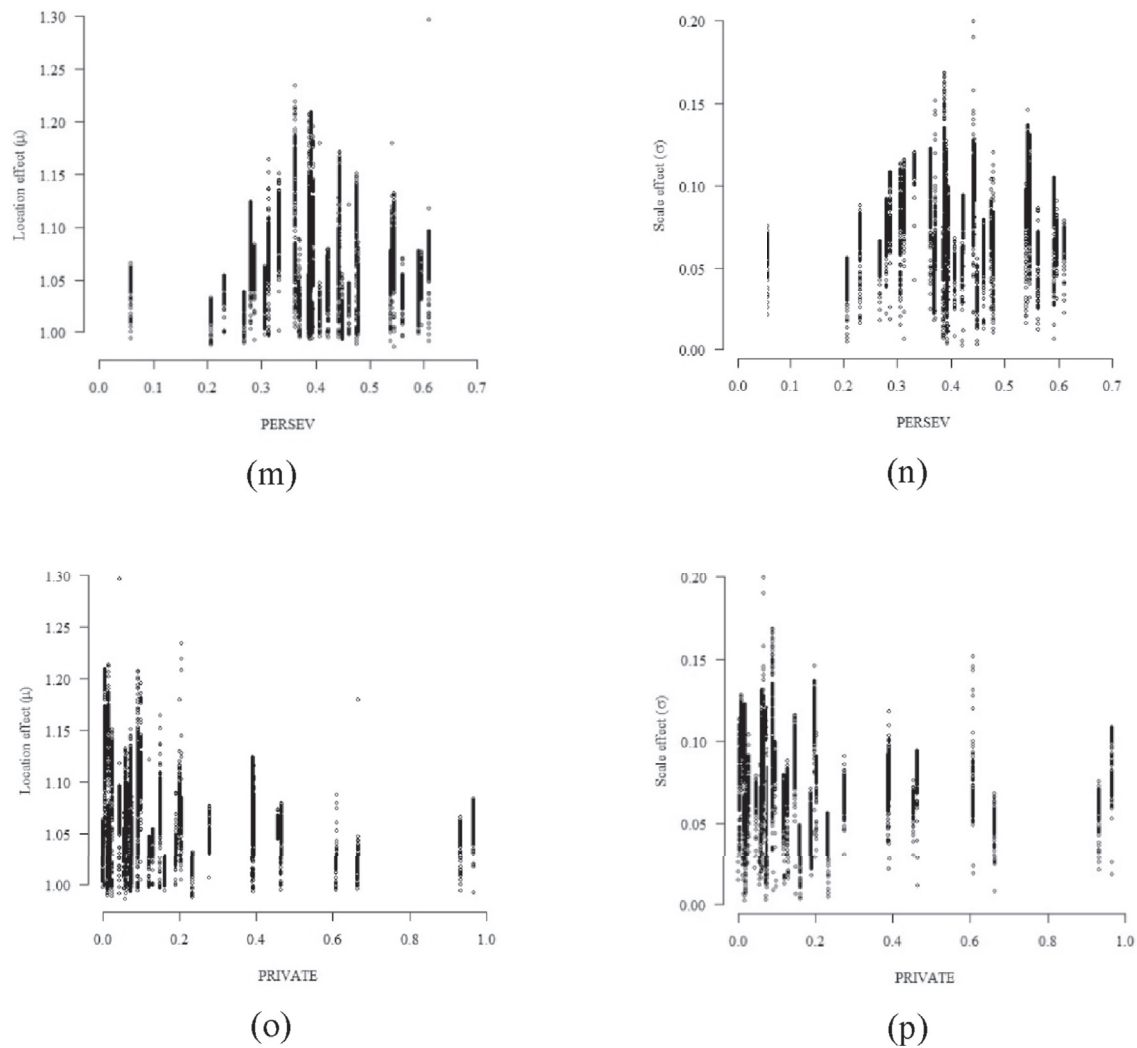
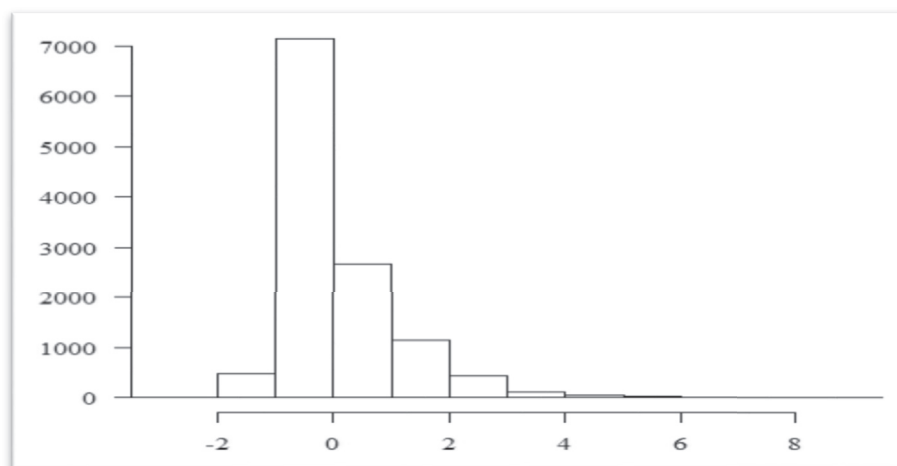


Fig. 3. (continued).

Fig. 4. Histogram of second-stage errors (*pure efficiency*).

into account the educational process inputs. In particular, some countries with poor results in terms of student proficiency (e.g. Turkey, Thailand, Mexico or Colombia) are ranked among the top performers according to their efficiency levels. Second, our findings suggest that cross-country heterogeneity is more relevant than among schools, since the country

ranking is hardly affected by the consideration of school contextual factors, whereas economic and cultural country variables have a greater impact on results. Third, we should highlight that, although not much studied in the efficiency literature as yet, non-cognitive cultural values have a significant impact as contextual factors affecting school

Table 8Ranking of countries according to *pure efficiency*.

Pure efficiency			
Country	Avg. score	Country	Avg. score
Korea	−0.0568	Chile	0.0143
Czech Rep.	−0.0494	Sweden	0.0154
Germany	−0.0473	Tunisia	0.0158
Slovak Rep.	−0.0393	Thailand	0.0208
United States	−0.0374	United Kingdom	0.0216
Estonia	−0.0300	Indonesia	0.0223
New Zealand	−0.0250	Singapore	0.0273
Colombia	−0.0178	Spain	0.0273
Poland	−0.0162	Italy	0.0298
Norway	−0.0130	Japan	0.0317
Russian Fed.	−0.0085	Hong Kong	0.0322
Canada	−0.0073	Netherlands	0.0477
Slovenia	0.0016	Turkey	0.0652
Switzerland	0.0039	Latvia	0.0657
Israel	0.0045	Lithuania	0.0744
Mexico	0.0055	Romania	0.0811
Australia	0.0132	Uruguay	0.0945
Finland	0.0137	France	0.1522
TOTAL		0.0148	

performance and explaining differences in efficiency estimates across countries. Nevertheless, some factors that have been previously identified as potential determinants of student achievement, like the use of tracking or central examinations, do not appear to affect the performance of secondary schools in terms of efficiency.

These findings provide some interesting insights for the analysis of school efficiency using a cross-country approach. However, more research will be needed in future to further explore the results discussed here. In particular, note that the results of the approach used in this paper cannot be interpreted causally, since it 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 respect, more research combining quasi-experimental methodologies with efficiency measurement techniques is still needed within this framework (see Santín and Sicilia, 2017, 2018) to exploit exogenous educational policies variations. Likewise, the proposed analysis could be replicated using student-level data. Actually, these data have been used in some studies to assess the performance of students in a specific country (e.g. Perelman and Santín, 2011; De Witte and Kortelainen, 2013). Note, however, that kernel bandwidth estimation is computationally intensive. Therefore, it can take a significant amount of time to complete all the estimations given the huge number of individual observations available in large-scale international databases. Finally, another potential future development of this line of research is to consider non-radial measures of performance in order to explore the possibility of students (or schools) focusing their efforts on some rather than other educational outcomes as suggested by Aparicio et al. (2018b).

Acknowledgements:

The authors thank the financial support from the Spanish Ministry for Economy and Competitiveness (*Ministerio de Economía, Industria y Competitividad*) through grant ECO2017-83759-P and Junta de Extremadura through grant IB16171.

References

- Afonso, A., St Aubyn, M., 2006. Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs. *Econ. Modell.* 23 (3), 476–491.
- Agasisti, T., 2014. The efficiency of public spending on education: an empirical comparison of EU countries. *Eur. J. Educ.* 49 (4), 543–557.

- Agasisti, T., Zoido, P., 2015. The Efficiency of Secondary Schools in an International Perspective: Preliminary Results from PISA 2012. OECD Education Working Papers, (No. 117). OECD Publishing.
- Ammermüller, A., Heijke, H., Wößmann, L., 2005. Schooling quality in Eastern Europe: educational production during transition. *Econ. Educ. Rev.* 24 (5), 579–599.
- Aparicio, J., Cordero, J.M., Pastor, J.T., 2017a. The determination of the least distance to the strongly efficient frontier in data envelopment analysis oriented models: modelling and computational aspects. *Omega* 71, 1–10.
- Aparicio, J., López-Torres, L., Santín, D., 2018a. Economic crisis and public education. A productivity analysis using a Hicks-Moorsteen index. *Econ. Modell.* 71, 34–44.
- Aparicio, J., Cordero, J.M., Gonzalez, M., Lopez-Espin, J.J., 2018b. Using non-radial DEA to assess school efficiency in a cross-country perspective: an empirical analysis of OECD countries. *Omega*. <https://doi.org/10.1016/j.omega.2017.07.004> (forthcoming).
- Aristovnik, A., Obadić, A., 2014. Measuring relative efficiency of secondary education in selected EU and OECD countries: the case of Slovenia and Croatia. *Technol. Econ. Dev. Econ.* 20 (3), 419–433.
- Badin, L., Daraio, C., 2011. Explaining efficiency in nonparametric frontier models. Recent developments in statistical inference. In: Van Keilegom, I., Wilson, P.W. (Eds.), *Exploring Research Frontiers in Contemporary Statistics and Econometrics*. Springer-Verlag, Berlin, pp. 151–175.
- Badin, L., Daraio, C., Simar, L., 2010. Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. *Eur. J. Oper. Res.* 201, 633–640.
- Badin, L., Daraio, C., Simar, L., 2012. How to measure the impact of environmental factors in a nonparametric production model? *Eur. J. Oper. Res.* 223, 818–833.
- Badin, L., Daraio, C., Simar, L., 2014. Explaining inefficiency in nonparametric production models: the state of the art. *Ann. Oper. Res.* 214 (1), 5–30.
- Bogetoft, P., Heinesen, E., Tranæs, T., 2015. The efficiency of educational production: a comparison of the Nordic countries with other OECD countries. *Econ. Modell.* 50, 310–321.
- Bol, T., Witschge, J., Van de Werfhorst, H.G., Dronkers, J., 2014. Curricular tracking and central examinations: counterbalancing the impact of social background on student achievement in 36 countries. *Soc. Forces* 92 (4), 1545–1572.
- Borghans, L., Schils, T., 2011. The Leaning Tower of PISA: the Effect of Test Motivation on Scores in the International Student Assessment. Mimeo. <http://www.sole-jole.org/13260.pdf>.
- Braga, M., Checchi, D., Meschi, E., 2013. Educational policies in a long-run perspective. *Econ. Pol.* 28 (73), 45–100.
- Brunello, G., Checchi, D., 2007. Does school tracking affect equality of opportunity? New international evidence. *Econ. Pol.* 52, 781–861.
- Camanho, A.S., Portela, M.C., Vaz, C.B., 2009. Efficiency analysis accounting for internal and external non-discretionary factors. *Comput. Oper. Res.* 36 (5), 1591–1601.
- Cazals, C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. *J. Econom.* 106, 1–25.
- Cherchye, L., De Witte, K., Ooghe, E., Nicaise, I., 2010. Efficiency and equity in private and public education: a nonparametric comparison. *Eur. J. Oper. Res.* 202 (2), 563–573.
- Clements, B., 2002. How efficient is education spending in Europe? *Eur. Rev. Econ. Finance* 1 (1), 3–26.
- Coco, G., Lagravinese, R., 2014. Cronyism and education performance. *Econ. Modell.* 38, 443–450.
- Conti, G., Heckman, J.J., Urzúa, S., 2011. Early Endowments, Education and Health. Human Capital and Economic Opportunity: a Global Working Group. Working Paper 2011–001.
- Cordero, J.M., Alonso-Morán, E., Nuño-Solinis, R., Orueta, J.F., Arce, R.S., 2015. Efficiency assessment of primary care providers: a conditional nonparametric approach. *Eur. J. Oper. Res.* 240 (1), 235–244.
- Cordero, J.M., Cristóbal, V., Santín, D., 2017a. Causal inference on education policies: a survey of empirical studies using PISA, TIMSS and PIRLS. *J. Econ. Surv.* <https://doi.org/10.1111/joes.12217> forthcoming.
- Cordero, J.M., Santín, D., Simancas, R., 2017b. Assessing European primary school performance through a conditional nonparametric model. *J. Oper. Res. Soc.* 68 (4), 364–376.
- Creemers, B., Kyriakides, L., 2008. *The Dynamics of Educational Effectiveness*. Routledge, London.
- Crespo-Cebada, E., Pedraja-Chaparro, F., Santín, D., 2014. Does school ownership matter? An unbiased efficiency comparison for regions of Spain. *J. Prod. Anal.* 41 (1), 153–172.
- Daouia, A., Gijbels, I., 2011. Robustness and inference in nonparametric partial frontier modeling. *J. Econom.* 161 (2), 147–165.
- Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *J. Prod. Anal.* 24 (1), 93–121.
- Daraio, C., Simar, L., 2007a. *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodologies and Applications*. Springer, New York.
- Daraio, C., Simar, L., 2007b. Conditional nonparametric frontier models for convex and non-convex technologies: a unifying approach. *J. Prod. Anal.* 28, 13–32.
- Daraio, C., Simar, L., 2014. Directional distances and their robust versions: computational and testing issues. *Eur. J. Oper. Res.* 237 (1), 358–369.
- Daraio, C., Simar, L., Wilson, P.W., 2010. Testing whether Two-stage Estimation Is Meaningful in Non-parametric Models of Production. ISBA Discussion Papers, no1031.
- Daraio, C., Simar, L., Wilson, P.W., 2015. Testing the "separability" Condition in Two-stage Nonparametric Models of Production. LEM Working Paper Series 2015/21.
- Daraio, C., Simar, L., Wilson, P.W., 2018. Central limit theorems for conditional efficiency measures and tests of the "separability" condition in nonparametric, two-stage models of production. *Econom. J.* <https://doi.org/10.1111/ectj.12103> forthcoming.

- De Jorge, J., Santin, D., 2010. Determinantes de la eficiencia educativa en la Unión Europea. *Hacienda Publica Espanola* 193, 131–155.
- De Witte, K., Kortelainen, M., 2013. What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Appl. Econ.* 45 (17), 2401–2412.
- De Witte, K., López-Torres, L., 2017. Efficiency in education: a review of literature and a way forward. *J. Oper. Res. Soc.* 68 (4), 339–363.
- Deutsch, J., Dumas, A., Siber, J., 2013. Estimating an educational production function for five countries of Latin America on the basis of the PISA data. *Econ. Educ. Rev.* 36, 245–262.
- Dufrechou, P.A., 2016. The efficiency of public education spending in Latin America: a comparison to high-income countries. *Int. J. Educ. Dev.* 49, 188–203.
- Emrouznejad, A., Yang, G.L., 2018. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Soc. Econ. Plann. Sci.* 61, 4–8.
- Eurydice, 2004. European Glossary on Education, Volume 1: Examinations, Qualifications and Titles. Eurydice, Brussels.
- Fernandez, R., Fogli, A., 2009. Culture: an empirical investigation of beliefs, work and fertility. *Am. Econ. J. Macroecon.* 1 (1), 146–177.
- Giambona, F., Vassallo, E., Vassiliadis, E., 2011. Educational systems efficiency in European Union countries. *Stud. Educ. Eval.* 37 (2), 108–122.
- Giménez, V., Prior, D., Thieme, C., 2007. Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. *J. Oper. Res. Soc.* 58 (8), 996–1007.
- Giménez, V., Thieme, C., Prior, D., Tortosa-Ausina, E., 2017. An international comparison of educational systems: a temporal analysis in presence of bad outputs. *J. Prod. Anal.* 47 (1), 83–101.
- Gustafsson, J.E., 2008. Effects of international comparative studies on educational quality on the quality of educational research. *Eur. Educ. Res. J.* 7 (1), 1–17.
- Haelermans, C., De Witte, K., 2012. The role of innovations in secondary school performance. Evidence from a conditional efficiency model. *Eur. J. Oper. Res.* 223 (2), 541–549.
- Hall, P., Racine, J.S., Li, Q., 2004. Cross-validation and the estimation of conditional probability densities. *J. Am. Stat. Assoc.* 99 (468), 1015–1026.
- Hanushek, E.A., 1979. Conceptual and empirical issues in the estimation of educational production functions. *J. Hum. Resour.* 14, 351–388.
- Hanushek, E.A., 2003. The failure of input-based schooling policies. *Econ. J.* 113 (485), 64–98.
- Hanushek, E.A., Kimko, D.D., 2000. Schooling, labor-force quality, and the growth of nations. *Am. Econ. Rev.* 90 (5), 1184–1208.
- Hanushek, E.A., Woessmann, L., 2006. Does educational tracking affect performance and inequality? Differences-in-Differences evidence across countries. *Econ. J.* 116 (510), 63–76.
- Hanushek, E.A., Woessmann, L., 2011. The economics of international differences in educational achievement. In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), *Handbook of the Economics of Education*, vol. 3. North Holland, Amsterdam, pp. 89–200.
- Hanushek, E.A., Woessmann, L., 2014. Institutional structures of the education system and student achievement: a review of cross-country economic research. In: Strietholt, R., Bos, W., Gustafsson, J.E., Rosen, M. (Eds.), *Educational Policy Evaluation through International Comparative Assessments*. Waxmann Verlag, pp. 145–176.
- Haveman, R., Wolfe, B., 1995. The determinants of children's attainments: a review of methods and findings. *J. Econ. Lit.* 33 (4), 1829–1878.
- Heckman, J.J., 2011. Integrating Personality Psychology into Economics. Working Paper No. 17378. National Bureau of Economic Research.
- Johnes, J., 2015. Operational research in education. *Eur. J. Oper. Res.* 243 (3), 683–696.
- Kamens, D.H., 2009. Globalization and the growth of international educational testing and national assessment. *Comp. Educ. Rev.* 54 (1), 5–25.
- Kneip, A., Simar, L., Wilson, P.W., 2015. When bias kills the variance: central limit theorems for DEA and FDH efficiency scores. *Econom. Theor.* 31 (2), 394–422.
- Le Donné, N., 2014. European variations in socioeconomic inequalities in students' cognitive achievement: the role of educational policies. *Eur. Socio Rev.* 30 (3), 329–343.
- Levin, H., 1974. Measuring the efficiency in educational production. *Publ. Finance Q.* 2, 3–24.
- Li, Q., Racine, J., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton.
- Liberati, P., Lagravinese, R., Resce, G., 2017. How Does Economic Social and Cultural Status Affect the Efficiency of Educational Attainments? a Comparative Analysis on PISA Results. Working Paper No. 217. Department of Economics-University Roma Tre.
- Mastromarco, C., Simar, L., 2017. Globalization and productivity: a robust nonparametric world frontier analysis. *Econ. Modell.* 69, 134–149.
- Mendez, I., 2015. The effect of the intergenerational transmission of noncognitive skills on student performance. *Econ. Educ. Rev.* 46, 78–97.
- OECD, 2009. *PISA Data Analysis Manual*, SPSS Second Edition. OECD Publishing, Paris.
- OECD, 2010. *PISA 2009 at a Glance*. OECD Publishing, Paris. <https://doi.org/10.1787/9789264095298-en>.
- OECD, 2011. *Education at a Glance 2011: OECD Indicators*. OECD Publishing, Paris. <https://doi.org/10.1787/eag-2011-en>.
- OECD, 2014. *PISA 2012 Technical Report*. OECD Publishing, Paris.
- Pagan, A., Ullah, A., 1999. *Nonparametric Econometrics*. Cambridge university press.
- Perelman, S., Santin, D., 2011. Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. *Econ. J.* 19 (1), 29–49.
- Racine, J.S., 1997. Consistent significance testing for nonparametric regression. *J. Bus. Econ. Stat.* 15 (3), 369–378.
- Racine, J.S., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous data. *J. Econom.* 119 (1), 99–130.
- Rivkin, S.G., Schiman, J.C., 2015. Instruction time, classroom quality, and academic achievement. *Econ. J.* 125, 425–448.
- Ruggiero, J., Vitaliano, D.F., 1999. Assessing the efficiency of public schools using data envelopment analysis and frontier regression. *Contemp. Econ. Pol.* 17 (3), 321–331.
- Santin, D., Sicilia, G., 2015. Measuring the efficiency of public schools in Uruguay: main drivers and policy implications. *Latin Am. Econ. Rev.* 24 (1), 1–28.
- Santin, D., Sicilia, G., 2017. Dealing with endogeneity in data envelopment analysis applications. *Expert Syst. Appl.* 68, 173–184.
- Santin, D., Sicilia, G., 2018. Using DEA for measuring teachers' performance and the impact on students' outcomes: evidence for Spain. *J. Prod. Anal.* 49 (1), 1–15.
- Schuetz, G., Ursprung, H.W., Woessmann, L., 2008. Education policy and equality of opportunity. *Kyklos* 61 (2), 279–308.
- Simar, L., Wilson, P.W., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *J. Econom.* 136 (1), 31–64.
- Simar, L., Wilson, P.W., 2011. Two-stage DEA: caveat emptor. *J. Prod. Anal.* 36 (2), 205.
- Strietholt, R., Gustafsson, J.E., Rosén, M., Bos, W., 2014. Outcomes and causal inference in international comparative assessments. *Educ. Pol. Eval. Int. Comp. Assessments* 9.
- Sutherland, D., Price, R., Gonand, F., 2009. Improving public spending efficiency in primary and secondary education. *OECD J. Econ. Stud.* 2009 (1), 1–30.
- Täht, K., Must, O., 2013. Comparability of educational achievement and learning attitudes across nations. *Educ. Res. Eval.* 19 (1), 19–38.
- Teddlie, C., Reynolds, D., 2000. *The International Handbook of School Effectiveness Research*. Routledge, London.
- Thieme, C., Giménez, V., Prior, D., 2012. A comparative analysis of the efficiency of national education systems. *Asia Pac. Econ. Rev.* 13 (1), 1–15.
- Thieme, C., Prior, D., Tortosa-Ausina, E., 2013. A multilevel decomposition of school performance using robust nonparametric frontier techniques. *Econ. Educ. Rev.* 32, 104–121.
- Todd, P.E., Wolpin, K.I., 2003. On the specification and estimation of the production function for cognitive achievement. *Econ. J.* 113 (485), F3–F33.
- Tzeremes, N.G., 2015. Efficiency dynamics in Indian banking: a conditional directional distance approach. *Eur. J. Oper. Res.* 240, 807–818.
- Vandenbergh, V., Robin, S., 2004. Evaluating the effectiveness of private education across countries: a comparison of methods. *Lab. Econ.* 11 (4), 487–506.
- Verhoeven, M., Gunnarsson, V., Carcillo, S., 2007. Education and Health in 77 Countries: Achieving Better Outcomes with Less Spending (No. 2007-2263). International Monetary Fund.
- Verschele, M., Rogge, N., 2012. An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: evidence from a conditional Data Envelopment Analysis approach. *Eur. J. Oper. Res.* 223 (1), 214–225.
- West, M.R., Woessmann, L., 2006. Which school systems sort weaker students into smaller classes? International evidence. *Eur. J. Polit. Econ.* 22 (4), 944–968.
- West, M.R., Woessmann, L., 2010. Every Catholic child in a Catholic school: historical resistance to state schooling, contemporary private competition and student achievement across countries. *Econ. J.* 120 (546), 229–255.
- Woessmann, L., 2003. School resources, educational institutions and student performance: the international evidence. *Oxf. Bull. Econ. Stat.* 65 (2), 117–170.
- Woessmann, L., 2005. Educational production in Europe. *Econ. Pol.* 20 (43), 445–504.
- Woessmann, L., Luedemann, E., Schütz, G., West, M., 2009. *School Accountability, Autonomy, and Choice Around the World*. Edward Elgar, Cheltenham, UK.
- Woessmann, L., West, M., 2006. Class-size effects in school systems around the world: evidence from between-grade variation in TIMSS. *Eur. Econ. Rev.* 50 (3), 695–736.
- Worthington, A.C., 2001. An empirical survey of frontier efficiency measurement techniques in education. *Educ. Econ.* 9 (3), 245–268.
- Wu, M., 2005. The role of plausible values in large-scale surveys. *Stud. Educ. Eval.* 31 (2–3), 114–128.
- Zhao, Y., Zhang, G., Yang, W., Kirkland, D., Han, X., Zhang, J., 2008. A comparative study of educational research in China and the United States. *Asia Pac. J. Educ.* 28 (1), 1–17.