ISCTE IUL
Instituto Universitário de Lisboa

EFFICIENCY IN SCHOOL EDUCATION

A semi-parametric study of school efficiency in OECD countries

Gonçalo da Silva Lima

Dissertation submitted as partial requirement for the conferral of

Master in Economics

Supervisor:

Prof. Henrique Monteiro, Prof. Auxiliar, ISCTE Business School,
Departamento de Economia

October 2017

Page deliberately left blank.

Acknowledgements

To my late grandpa *A.*.

To my mum, my moral compass.

To my sister, who always put up with me, against all odds.

To my dad, who – even from afar – was always present.

To all my family and friends, that nurtured me and shaped me to be who I am today.

To Paris, that keeps inspiring me.

Gratitude to all work and university colleagues from whom I have learned immensely.

Special mention to my team, that allowed me to grow professionally and as a person.

To my supervisor, that always supported me through the most stressful moments of this dissertation.

And last, but not the least: to Emma. Thank you for all your encouragement and help!

All faults – either in my character or this dissertation – are attributable to me alone.

Abstract

Efficiency-driven reforms have become increasingly relevant in education policy, as education systems face tighter budget constraints. Educational authorities around the world often struggle to foster the best student outcomes out of the set of available school resources. This dissertation aims to contribute to this debate by using a semi-parametric approach to evaluate the efficiency across 34 OECD countries, using data from PISA 2015. The estimation of the education production possibilities frontier is made through a free disposal hull (FDH) method, a non-convex and non-parametric estimator, also extending the analysis to incorporate recently developed partial frontier methods (order-m and order-α). According to the different specifications, inefficient schools could have increased average student achievement between 9%-18%, for the same level of human and material resources, and given the socio-economic characteristics of their students. Differences in efficiency scores are also investigated. The results suggest that schools that enrol a larger number of students and where the principal can decide on budget allocations are more efficient. On the other hand, schools with high concentration of students from immigrant backgrounds and more repeaters are hindered in the provision of efficient allocations of school resources. Finally, no necessary trade-off is found between efficiency and equity in the provision of quality education.

Keywords: School efficiency, OECD, free disposal hull (FDH) estimator, partial frontiers analysis.

JEL Classification: I21, C14.

Resumo

Reformas tendo em vista aumentos de eficiência têm-se tornado crescentemente relevantes na definição de políticas educativas, especialmente no contexto de orçamentos educativos mais limitados. Neste sentido, responsáveis em diferentes sistemas educativos têm tentado saber como melhorar os resultados dos alunos, dados os recursos escolares disponíveis. Esta dissertação tem por objectivo contribuir para este debate, através de uma avaliação semi-paramétrica de eficiência escolar em 34 países da OCDE, recorrendo a dados do PISA 2015. Estimamos a fronteira de possibilidades de produção educativa através de *free disposal hull* (FDH), um estimador não-paramétrico e não-convexo. Também estendemos a análise para incorporar métodos de fronteiras parciais (order-m e order-α). De acordo com as diferentes especificações, as escolas ineficientes na amostra poderiam ter aumentado a qualidade de educação entre 9% e 18%, utilizando o mesmo nível de rescursos humanos e materiais, e tendo em conta as características socio-económicas dos seus alunos. A variação nos *scores* de eficiência é também investigada. Os resultados sugerem que escolas com um maior número de alunos e em que o diretor tem poder de decisão sobre a alocação do orçamento escolar são mais eficientes. Por outro lado, escolas com maior concentração de alunos de contextos familiares de imigração e com mais repetentes têm maior dificuldade em se aproximar da fronteira internacional de eficiência. Por fim, não há evidência de um *trade-off* necessário entre eficiência e equidade na provisão de educação de qualidade.

Palavas-chave: Eficiência escolar, OCDE, free disposal hull (FDH), análise de fronteiras parciais.

Classificação JEL: I21, C14.

Contents

List of Figures

List of Tables

1. Introduction

School education is a fundamental channel for enhancing individual and social well-being. The adequate development of cognitive and socio-emotional capabilities through quality school education leads to not only higher private financial returns in the future, but also economic growth, better health, improved nutrition and higher civic participation (OECD, 2012).

Reflecting this, primary and secondary education expenditures as a proportion of GDP per capita have been increasing among developed countries, in a long-term perspective (Wolff et al., 2014; Wolff, 2015). Nevertheless, the benefits from this increasing investment are not clear. In fact, wide-range international evidence has shown that higher expenditure in school education has no significant impacts on student performance, but are rather the context and the institutional arrangements moulding education systems and schools the main determinants of student success (Wößmann, 2016; Hanushek, 2006).

Despite the continuous increase in education expenditures, several countries have been facing tighter public budget constraints (OECD, 2013). Recent demographic developments have also been leading to a re-evaluation of human resource intensiveness in schools. Since most school resources in OECD countries are guaranteed by public funds, efficiency-driven reforms have become increasingly relevant in education policy.

From an economic perspective, the measurement of efficiency in education is especially interesting due to the nature of school activity. Despite the increasing tendency for the introduction of market-type mechanisms[1] in school education in the last decades (Levin, 2015) schools are far from operating in a competitive environment that would theoretically lead to economically efficient allocations of resources. Uncertainty about the educational process due to the intangibility of the inputs and outputs involved hampers the ability of school management to take efficiency-driven decisions. Moreover, schools generally enjoy a high degree of monopoly power over their catchment area, which relaxes the pressure for making the most out of the available resources (Levin et al., 1976, p. 158). This is

[1] Market-type mechanisms aim at facilitating the coordination between educational supply and demand, with the expectation that better outcomes and more efficient allocations can be generated. Typical examples of market-type mechanisms include free parental choice of schools, performance-based rewards and sanctions for schools and teachers, school management autonomy, or the promotion of school competition through increased accountability and benchmarking.

especially relevant in systems with legal binding constraints on school choice or schools located in remote locations.

International comparisons of efficiency in the use of school resources are useful for cross-country benchmarking. Furthermore, the identification of features of school systems correlated with efficient allocations enables further learning and assessment. However, and despite the increasing availability of international datasets in the area of school education, such as the ones provided by OECD's Programme for International Student Assessment (PISA) or the International Education Agency's (IEA) Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS), few studies have been providing international comparisons of school education efficiency (Witte and López-Torres, 2017).

In this dissertation, we attempt to contribute to the debate on school efficiency through a two-stage analysis, aimed at understanding how school efficiency varies across developed countries. First, we derive efficiency scores for 7 318 schools from 34 different OECD countries, using the most recent data from PISA. School efficiency is here measured as technical, rather than allocative efficiency. In this sense, we restrict our attention to the resources (e.g., human, physical) that are incorporated in the education process, but not taking into account its cost. This allows us to focus on the intensiveness in the use of this resources uncontaminated by differences in the prices of those resources. Additionally, we directly account for differences in socio-economic background across schools.

We start by presenting the distribution of school efficiency scores within and across countries. In order to compute these we employ three alternative methods. A free disposal hull (FDH) analysis –- a non-parametric technique for assessing the extent of inefficiencies within and across countries – works as our base model. But in order to control for outliers in the data we also use other recently developed non-parametric methods for efficiency analysis – order-m and order-α. Depending on the method, we find that schools could have increased student outcomes between 9 to 18%, if operating at the international production possibilities frontier. We also find that there is substantial heterogeneity on the characterisation of school efficiency across countries, and that this variation is mostly driven by differences in country-level factors.

In the second stage of the analysis, we attempt to understand what are the factors associated with differences in efficiency across schools. We use the derived efficiency scores as the dependent variable of a parametric regression model, to develop an exploratory analysis of the variables associated with efficiency. We find that larger schools, where the principal has greater autonomy in the allocation of the budget and where achievement data is posted publicly are more efficient. A higher concentration of immigrants and student ha have repeated at least one school year also hinders the ability to provide more efficient allocations. We also explore the relation between efficiency and equity. The results suggest that schools that have students from more diverse economic backgrounds and have less unequal student outcomes are also more efficient, controlling for all other factors. As in the literature, we find that the location of the school and its ownership are significant factors in its performance. In particular, schools located in communities with less than 15 000 inhabitants are relatively more efficient, controlling for all other factors. On the other hand, we find that private schools are generally further from the frontier, although the result does not hold across different models. Finally, schools with an average small class size are found to be less efficient in the preferred model specification.

This dissertation is organised in five further sections. We first start by shedding some light on the relevant literature for efficiency analysis in education. Section 2 presents and discusses the main concepts and methods in the literature. Section 3 discusses the main empirical results of international comparisons of school education efficiency employing similar or relevant approaches to the ones developed in our empirical study. Section 4 reflects on the methods used in our study and discusses their theoretical underpinnings. Section 5 presents the data and the variables used in more detail. Finally, section 6 presents and discusses the main results from our study.

2. The Measurement of Efficiency

The decade of 1950 witnessed the expansion of theoretical and empirical models laying the basis for efficiency analysis based on the estimation of production possibilities frontiers (Koopmans, 1951a; Debreu, 1951; Shephard, 1953; Farrell, 1957). From those models an extensive literature in production theory developed. Since then, the field blossomed with a myriad of different approaches: from parametric to non-parametric, employing statistical and non-statistical methods for estimating deterministic or stochastic production frontiers. This section aims at presenting and discussing the main conceptual and methodological developments that have empirical relevance for studying efficiency in school education.

2.1. Main concepts

Understanding efficiency starts by comprehending the nature of production processes. Any production process is composed of two types of basic elements: inputs and outputs. Inputs, or resources, are objects – tangible or intangible – which are utilised. The utilisation of those objects for production generates a second type of element – the output. The transformation of a given set of inputs into a given set of outputs is defined by a given function, describing the process of transformation by means of the existing technology, i.e., the available knowledge, codes and know-how to turn inputs into outputs. So, to what extent can such a process be deemed as efficient?

According to Koopmans (1951b, p. 60), following the principle of Pareto optimality, a production bundle – i.e., a pair of input and output sets – is efficient in two possible instances. On one hand, if an increase in a given output implies a decrease in at least one of the other outputs or a marginal increase in at least one of the inputs. On the other hand, if a marginal decrease in any of the inputs is made at the expense of an increase in one of the other inputs or in a reduction of at least one of the outputs. This notion of since been coined as Pareto-Koopmans efficiency.

In turn, Debreu (1951), following the notion of distance function in Shephard (1953), introduced the definition of 'coefficient of resource utilisation'. In order to gauge a measure of efficiency in a context of multiple inputs and outputs, the Shephard input distance function allows to compute relative efficiency by treating outputs as constants and by contracting the vector of inputs according to technological feasibility (Daraio and Simar, 2007b, p.16). The coefficient of resource utilisation is based on the reciprocal of the Shephard function and generalises the production process from the level of the firm to the level of the economy. It functions as a measure of deadweight loss in the economy due to suboptimal utilisation of resources, i.e, how much less resources can be used for the same level of satisfaction to the consumers. Based on these insights, Farrell (1957) alternatively defined a measure of *technical efficiency* also based on the reciprocal of the function first conceptualised in Shephard (1953). The Debreu-Farrell measure of technical efficiency can be translated as the minimum proportion by which a quantity of inputs used can be reduced for the same level of outputs (input-oriented approach), or the maximum proportion by which a set of outputs can be increased for the same level of inputs in production (output-oriented approach) (Daraio and Simar, 2007b, p. 14).

From the definitions, it follows that a given production bundle may be Debreu-Farrell efficient without being Pareto-Koopmans efficient. For the first measure, a firm can be operating efficiently while it being possible to reduce the level of at least one of the other inputs – the relevant factor for the efficiency characterisation is thus the equiproportionate reduction in *all* inputs. On the other hand, the optimality condition in the Pareto-Koopmans measure is stricter – efficient subsets are those for which no reduction in *any* of the inputs or expansion in any of the outputs is still possible.

Importantly, efficiency differs from productivity mainly due to its normative nature (Ray, 2004, p.15). While productivity can be roughly defined as the ratio between the sets of outputs and inputs used in the production process, inefficiencies are measured as the difference between the output-input ratio that *can be maximally attained*, given the production technology, and the *observed* output-input ratio (Lovell, 1993). Therefore, gains in productivity do not necessarily imply gains in technical efficiency[1]. It follows that a more productive production unit A is not necessarily more efficient than a less productive production unit B – the relevant comparison for efficiency evaluation is rather made through the projection of the bundles A and B on the production possibilities frontier of the production sector under consideration.

. Empirically, the question lies on how to specify the production possibilities frontier. Farrell (1957) had already pointed out the relative nature of the concept:

[1]This is only true for the particular case of technologies with constant returns to scale. For deeper insights on this, please refer to Ray (2004, pp. 14-22).

> 'Technical efficiency, then, is defined in relation to a given set of firms, in respect of a given set of factors measured in a specific way, and any change in these specifications will affect the measure' (Farrell, 1957, p. 260).

In this sense, productive efficiency of any firm can be empirically compared to the best observed practice in the set of firms considered.

Farrell (1957) also established the conceptual distinction between *technical* and *allocative* efficiency. Technical efficiency focuses on feasibility in production, i.e., the output that feasible to produce given the inputs and the technology. Allocative efficiency, in turn, takes into account the cost minimizing behaviour of the firm. In this case, the relevant variables for the assessment of efficiency are not only the quantities of input and output in production but also their prices. Thus, allocative efficiency measures the ability of a given production unit in choosing an optimal set of inputs with given prices. In economics, the focus is usually on the analysis of allocative efficiency in competitive markets. However, there is a well-defined set of instances in which prices might not provide the sufficient information for inferring the incentives for efficient production. For instance, the measurement of technical efficiency is more insightful when there are no available prices of inputs or outputs or when inputs are difficult to be separated in the production process. That is especially the case in the education sector, where input and output prices are rarely available or difficult to define with accuracy (Johnes, 2004, p. 635).

Other concepts of efficiency capture the differences in the scale of production. *Scale efficiency* has been defined based on different methods and assumptions, from constant returns to scale (CRS) to non-decreasing or non-increasing returns to scale (Banker et al., 1984). Assessments of scale efficiency enable to understand if differences in efficient production across firms can be attributed to inadequacies in their size. This is especially relevant from an economic point of view, as inefficiencies associated with the size of production might not be amendable in the short-run.

2.2. Main methods

The methods for efficiency measurement can be roughly separated in parametric or non-parametric, statistical or non-statistical and deterministic or stochastic.

Parametric methods assume that the production process can be described through a known production function, usuallt specified as a Cobb-Douglas or a more flexible translog production functions (Sutherland et al., 2009, p. 7). Parametric methods are valuable due to their statistical properties, stemming from the assumptions on the distribution of the error terms. The estimation procedure allows to easily build confidence intervals from the standard errors, while the parameter estimates enable to straightforwardly compute the marginal effects of given variables over efficiency estimates and their statistical significance. However, the estimates are sensitive to misspecification in the distribution of the errors (Lovell, 1993). Regression-based ordinary least squares (OLS) and other similar approaches have been used for estimating deterministic deviations from production frontiers in education(for a brief review of these methods see Johnes, 2004, pp. 625-28). However,the use of OLS for the estimation of production frontiers has obvious limitations. Since it is a regression to the mean technique, several points lie above the fitted line – and thus the production frontier. Alternative versions, such as Modified OLS (MOLS) or Corrected OLS (COLS) (Richmond, 1974; Greene, 1980b) shifting the frontier upwards, as well as maximum likelihood estimators (MLE) (Greene, 1980a) have been used in efficiency measurement in education to surpass these limitations (for empirical contributions in school education see, for instance, Jones and Zimmer, 2001; Häkkinen et al., 2003). Nevertheless, none of those methods is appropriate for efficiency assessment in the context of production processes with multiple outputs (Johnes, 2004, p. 642). Analyses of efficiency in education focusing on just one output might be argued to be less insightful, as schools pursue various goals through their activity, such as the development of the students' multiple cognitive skills, besides their socio-emotional capabilities.

Another limitation regards the deterministic nature of these methods. Parametric deterministic methods assume that the error term, i.e., the distance between the estimated frontier and each observation is due to inefficiency. To tackle this drawback, stochastic frontier analysis (SFA), as originally developed by Aigner et al. (1977), enabled the decomposition of the error term in the regression analysis into a component attributable to statistical noise and another attributed to inefficiency[2]. SFA methods, in particular, have been extensively used in the context of efficiency measurement in school education (e.g., Kang and Greene, 2002; Sutherland et al., 2009).

Non-parametric techniques, in turn, rely on a set of weaker assumptions regarding the functional form of the production process. In this case, the process by which inputs are turned into outputs is not set *a priori* – the importance of each factor in the production

[2]For more details on SFA please refer to Aigner et al. (1977, pp. 24-29).

process is thus fully inferred from the data through mathematical programming techniques. Notwithstanding the benefits from relaxing some of the usual assumptions, this bears some evident limitations for economists. Usual t-tests and marginal effects cannot be computed through a standard least square regression. Therefore, it is not possible to evaluate marginal costs or elasticities of substitution[3]. Nevertheless, these methods have gained further acceptance in the economics discipline through the development of tests based on non-parametric approaches to the estimation of the production frontier (see, per example, Afriat, 1972; Hanoch and Rothschild, 1972; Diewert and Parkan, 1983; Varian, 1984).

Data envelopment analysis (DEA) fits into this tradition of non-parametric analyses in production economics (Ray, 2004, p. 4). It has been widely applied as one of those mathematical programming techniques, also being the most used approach for empirical measurement of efficiency in education (Witte and López-Torres, 2017). DEA was first introduced by Charnes et al. (1978) building upon the Debreu-Farrell concept of technical efficiency – broadly applied to an hypothetical set of decision-making units (DMUs), rather than strictly focusing on productive firms with the objective of maximizing profit. Such generalization was especially important for the study of efficiency in public sector services, including those of school education. Furthermore, it introduced further flexibility to measure efficiency at different levels of the system. In that sense, not only schools but also education systems and students can be characterized as DMUs (e.g., Portela and Thanassoulis, 2001). At the school-level, DEA allows benchmarking against the set of efficient educational institutions and setting specific output targets for inefficient schools (Johnes, 2004). At the student-level, it further controls for differences in the acquired abilities of students and better assesses heterogeneity across schools and the true contribution of school management for performance. At the system-level, the method enables cross-country comparisons of resource use and waste.

Later, Banker et al. (1984) relaxed the CRS assumption and generalized the model for technologies with variable returns to scale (VRS), enabling to separate the standard measure of efficiency with the assumption of CRS into distinct measures of technical and scale efficiency. However, other axioms of the DEA model were kept, particularly the usual

[3]However, this does not necessarily mean that linear programming methods can be rendered as useless for economic analysis. For instance, already in 1958 Dorfman, Samuelson and Solow, stated in the foreword of their book that 'much of economic analysis is linear programming' (Dorfman et al., 1958, foreword).

assumption of convexity in production. Convex technologies imply that if two production plans are feasible then their linear combination is always feasible[4]. In fact, some authors have long been arguing that the convexity axiom might be violated in the presence of economies of scale or specialization and where there is indivisibility of inputs and outputs (Farrell, 1959). Mayston (2017) has recently extended the critique to the case of educational production functions.

As an alternative to the convex environment of DEA, the free disposable hull (FDH) method was introduced by Deprins et al. (1984), providing a more flexible understanding of the production frontier. The properties of the FDH estimator are also relevant for attaining accurate empirical measures. If the true educational production function is not convex, DEA is an inconsistent estimator of the frontier, while the FDH estimator remains consistent. On the other hand, if the production set is convex, the FDH estimator is still consistent, despite being biased (Daraio and Simar, 2007b, p. 33). Therefore, if no good reasons are presented for the production set to be convex, then FDH yields greater empirical validity. Nevertheless, this method only started to be systematically applied to the context of efficiency measurement in school education during the past decade (for an identification of these studies, see Witte and López-Torres, 2017).

A major limitation to DEA and FDH methods regards its generally deterministic nature. Efficiency estimates can be easily contaminated by statistical noise and are sensitive to measurement errors and outliers. Recent methodological developments have been able to overcome or at least limit these important drawbacks. Bootstrapping methods (Simar and Wilson, 1999) and other re-sampling techniques (Cazals et al., 2002; Daraio and Simar, 2005) allow for more robust efficiency estimates by incorporating statistical properties and stochasticity in non-parametric measures. These have been helping to bridge the gap between parametric and non-parametric literatures (for a review of stochastic non-parametric methods see, e.g., Olesen and Petersen, 2015).

Within this line of research, order-m estimators have been proposed to deal with the sensitivity of the FDH and DEA estimator to extreme values and measurement errors by allowing randomness in the estimation of efficiency. The order-m method computes partial frontiers by benchmarking each decision-making unit by expected best performance in random samples of m peers (Cazals et al., 2002)[5]. It is thus a partial frontier method

[4] Convexity will be further conceptualized and discussed in a forthcoming Annex

[5] For a detailed explanation of the method please refer to section 4.3, or Daraio and Simar (2007a, p.

by only considering sub-samples of the frontier to be estimated. Furthermore, order-m techniques allow for a statistical treatment of the results through the construction of standard errors based on bootstrapping. The most general estimator builds on the FDH set of assumptions about the production process but can also accommodate the further assumption of convex technologies, as in DEA (Cazals et al., 2002). Order-m methods have been recently applied in the context of efficiency in school education (as discussed in section 3.2).

On the other hand, order-α has been proposed as an alternative partial frontier analysis estimator (Aragon et al., 2005). While the sub-samples in order-m are drawn according to a number of m random observations, order-α follows a quantile-based approach. The efficiency of each decision-making unit is benchmarked with the minimum input consumption peers which are in the $(100 - \alpha)$ percentile of the distribution. Nevertheless, and to the best of my knowledge, there are no empirical studies using order-α estimators measuring school efficiency.

The main theoretical claims of this line of research are used to design the methodology or the empirical analysis. Its main contribution is to extend the literature of efficiency measurement in education using partial frontier analysis methods. Other contributions are made explicit in the next section, where a stronger focus is given to the empirical results of the studies using DEA, FDH and partial frontier analysis methods.

72).

3. Empirical Studies of School Efficiency

Non-parametric empirical studies of efficiency measurement in school education can be roughly characterized according to four features: *i)* the empirical production set, as characterized by the list of inputs and outputs chosen to be part of the educational production process; *ii)* the level of the school system at which efficiency is evaluated; *iii)* the list of variables the efficiency assessment is conditioned to; and *iv)* the concepts and methods used in the efficiency evaluation.

The inputs and outputs to include in the computation of the efficiency scores are especially important as the omission of relevant variables produces biased results (Bifulco and Bretschneider, 2001, p. 419). On the other hand, the inclusion of an extensive collection of inputs may lead to multicollinearity problems (Johnes, 2004, p. 655), meaning a principle of parsimony in choice is desirable.

Another issue relates to the level at which efficiency is measured. School-level efficiency has been the most common in the literature, while international comparisons of efficiency make use of both school-level and system-level efficiency scores (e.g., Cordero et al., 2017). This entails the estimation of two different frontiers – at the school- and at the system-level. Another strand of literature has been using student-level data for further separating between school management contribution and the individual characteristics of students for performance (as first developed in Portela and Thanassoulis, 2001). Notwithstanding, the use of student-level data for international non-parametric comparisons of efficiency has not been empirically assessed, mainly due to its computational cumbersomeness.

We here provide a selected overview of the main contributions for building our empirical strategy, focusing first on the outputs and inputs generally included in school education efficiency studies and then on the main insights from this type of literature.

3.1. Outputs and inputs used in empirical studies

There is no theoretical consensus about the set of measurable outputs of the school education process. Empirically, the most common method to assess educational quality relies on students' cognitive skills, usually measured through students' results in standardized achievement tests. The underlying assumptions are that the main objective of the school system is enhancing students' cognitive abilities and that these can be captured in the controlled environment of a test. International standardized tests like PISA, focused on assessing basic skills in reading, mathematics and science, or TIMSS and PIRLS, focused on covering a common core range of national curricula, allow to compare students' cognitive skills across countries in a sufficiently standardized manner. In fact, variation in cognitive skills is consistently correlated with variation of returns in the labour market (Hanushek and Luque, 2003). Furthermore, differences in test scores capture a significant part of the variation in high school completion and college continuation (Rivkin, 1995).

Recently, the inclusion of measures of non-cognitive skills, such as motivation or subjective perceptions of well-being, like sense of belonging to school, in cross-section international datasets has been increasing the body of literature using this type of outcomes of the educational process (e.g., Tramonte and Willms, 2010)[1].

On the other hand, the inputs of the educational production process used in the literature are varied and not consensually considered for the analyses. These are not only the physical, human and financial resources invested in education but also the contextual and institutional factors setting the constraints of school systems. Human resources often include teacher and student characteristics, such as teacher experience and student prior achievement (e.g., Cherchye et al., 2010), student-teacher ratios, as a measure of human resource intensity, and the hours at school or class (e.g., Afonso and St. Aubyn, 2006; Giménez et al., 2007). On the other hand, the number of computers per student (e.g., Agasisti and Zoido, 2015) or the quality of teaching materials and facilities (e.g., Giménez et al., 2007) are usually employed as proxies of physical and capital inputs. Financial resources typically include expenditure per student or teacher' salaries and are the most used school-related input variables in efficiency measurement studies in education (Witte and López-Torres, 2017). Measures of expenditure are relevant for the assessment of allocative efficiency as they consider the cost of the inputs in the schooling process. However, such type of studies present some important drawbacks for international comparisons of efficiency. Education expenditures across countries also vary due to disparities in unit labour costs, capital costs and other labour or financial market factors not directly related school systems' efficiency (Sutherland et al., 2009, p. 4). For instance, differences in teachers' salaries across countries may largely reflect differences in labour market rigidities rather

[1]For a survey on measures of non-cognitive skills refer to Farkas (2003). For a review on the importance of non-cognitive skills as outputs of the education system and their relation to successful transitions to the labour market see Bowles and Gintis (2002).

than substantial differences in teachers' quality or inadequacy in pay. Allocative efficiency studies may then attribute inefficiency of educational policy and school management to variations that are in fact exogenous to the school system.

Another relevant discussion for the choice of inputs to include in the analysis lies on the level of control educational authorities and school management have over the relevant variables. The distinction between discretionary and non-discretionary inputs and their inclusion in the analysis allows to qualify the conclusions according to the context in which schools operate. Not controlling for differences in the operational environment may lead to overestimated inefficiencies, as these can be significantly explained by differences in the operational environment (Johnes, 2004, pp. 656-57). Discretionary inputs denote the set of malleable conditions directly under the control of the school system or schools (Scheerens et al., 2011, p. 37). On the other hand, non-discretionary inputs account for environmental constraints, which despite not being controlled by the system affect its final outcomes. At the student-level, non-discretionary inputs are usually variables of socio-economic background, peer effects or innate ability. At the school-level, those include characteristics of the environment in which the school is integrated that are not controllable at least in the short-run, for example if it is situated in a rural or urban area, or if it is public or privately owned (Cordero-Ferrera et al., 2008, p. 1324). At the system-level, non-discretionary inputs are all the contextual factors that affect the allocation of resources but also the institutional factors that are not directly dependent on the education policy. Notwithstanding, the literature has been showing that the choice of inputs that are discretionary or non-discretionary depends on the nature of the assessment and largely reflects a methodological decision made by the analyst.

Recent empirical studies have been revealing the impact of contextual variables in efficiency scores. The family environment in which the child is raised has a significant impact on achievement, shapes motivation and determines aspirations. Not only the education of parents but also their involvement at home has been consistently shown to have a positive effect on educational efficiency. Associated with those factors, measures of socio-economic status and resources available at home are also positively correlated with student performance and school efficiency. Disadvantaged socio-economic backgrounds hamper student's academic success, while schools mainly composed by students from these backgrounds are generally less efficient due to the harshness of their operational environment (e.g., Witte and Kortelainen, 2013; Grosskopf et al., 2014; Agasisti and Zoido, 2015).

Aggregate contextual variables also have an important impact on the ability of educational institutions to translate school resources into student achievement. GDP per capita is generally correlated with efficiency differences across education systems, and has been extensively used as a relevant environmental variable (e.g., Afonso and St. Aubyn, 2006; Agasisti, 2014).

Other factors consistently correlated with lower levels of efficiency have been the proportion of students with immigrant status and from a non-native language background. Moreover, students with disabilities and additional educational needs are also more costly to educate (e.g., Grosskopf et al., 2014) hindering the capacity to attain more efficient allocations.

Finally, there has been little academic consensus regarding the effects of school ownership (private, public, charter) or school size (Witte and López-Torres, 2017). Larger schools can reduce costs through economies of scale but educational outcomes can also be negatively affected. The impacts of private or public ownership, on the other hand, seem to depend on which educational outcomes are considered and how contextual variables are included in the analysis. Although the average performance of students is higher in private schools, some studies find that efficiency scores can be higher in public schools, especially when controlling for the socio-economic background of the students (e.g., Agasisti, 2013).

Finally, efficiency measurement studies have been giving little insights regarding the contributions of specific teachers' characteristics to efficiency, since the evidence has been mixed and sometimes insignificant (Witte and López-Torres, 2017). More research and data seems to be needed at the class-room level for more robust results to be drawn regarding this type of inputs. Unfortunately, the PISA dataset used in my empirical inquiry does not provide data to robustly assess the importance of teaching factors. However, value-added models for measuring the impact of teachers on individual long-term outcomes have been providing clear evidence for the importance of teacher quality in future life prospects[2].

[2] For the golden standard of research in the area please refer to Chetty et al. (2014).

3.2. Main results of international efficiency comparisons

The evidence on the comparison of educational efficiency at an international level has been surprisingly scarce *vis-à-vis* the large number of efficiency measurement studies in education[3]. Despite the increasing availability of datasets with internationally comparable data, most of empirical analyses are still confined to national and sub-national contexts (for an extensive literature collection see Witte and López-Torres, 2017). Furthermore, not all of these studies are directed to assess the relevance of given institutional and funding arrangements to explain differences in efficiency.

Most international efficiency frontier studies use DEA to compute efficiency scores. Even so, there are considerable differences in the options for integrating non-discretionary inputs in the analyses. Most studies adopt a multi-stage approach, where efficiency scores are then used as the dependent of either a Tobit (Agasisti and Zoido, 2015; Afonso and St. Aubyn, 2006) or an OLS (Agasisti, 2014) regression. The regression analysis allows to identify the significance of the factors correlated with school efficiency across countries and to understand the influence of contextual and institutional factors on school efficiency. Moreover, international comparisons have been dealing with the deterministic nature of non-parametric models by incorporating bootstrapping techniques in the estimation procedure (e.g., Agasisti and Zoido, 2015; Cordero et al., 2017; Agasisti, 2014; Giménez et al., 2007; Afonso and St. Aubyn, 2006).

Afonso and St. Aubyn (2006) – using aggregate data from OECD countries – found that, on average, these school systems could have increased 15 year old's student achievement results at PISA 2003 by 11.6 percent, using the same level of resources. According to the authors, Finland, Korea and Sweden were the most technically efficient education systems, given the intensity of teachers used and the hours per year in school. The Finnish school system was also found to be operating at the efficiency frontier in an European comparison of countries, using PISA 2006 and 2009 data. In it, a 10 percent saving of school resources could still be possible at the European level (Agasisti, 2014). Nevertheless, when the efficiency scores were further corrected for differences in GDP per capita and parents' educational attainment other countries, such as Portugal or Australia, stood out as the most efficient school systems (Afonso and St. Aubyn, 2006).

[3]Table 9.1, in Annex A: Literature on School Efficiency, presents a summary of non-parametric international frontier studies focused on measuring efficiency in school education.

More recently, Agasisti and Zoido (2015) – using PISA 2012 data – found that, on average, schools across OECD countries could have increased mathematical and reading literacy by 27 percent using the same level of resources, had they been operating efficiently. Furthermore, classes of small average dimension were negatively associated with efficiency, stressing the effects of higher intensiveness of teaching resources. On the other hand, higher budget autonomy was found to positively impact school efficiency (Agasisti and Zoido, 2015). This goes in line with the results of other international comparison studies using SFA methods, where it is found that autonomy of decision-making at the school-level, besides higher school funding decentralization and benchmarking between schools yield the potential for improving efficiency (Sutherland et al., 2009).

Alternatively, Giménez et al. (2007) directly correct DEA scores for environmental factors, using a one-stage approach. The analysis – making use of TIMSS 1999 data – showed that environmental factors play a key role in explaining differences among the results. An average increase in academic outcomes of 10 percent could be obtained, with 6 percent attributable to environmental factors and 4 percent to inefficiency of the system itself.

Recent partial frontier analysis have been introducing additional layers of data re-sampling through order-m, based on FDH estimators. A conditional robust non-parametric comparison of sixteen European countries suggests that the achievement of students – as measured by the results at PIRLS 2011 – could have increased on average by 7 percent, if all schools would perform as efficiently as the best performers (Cordero et al., 2017). The authors took into account heterogeneity across countries and across schools and concluded that most of the differences in technical efficiency tend to be driven by country factors (60%), such as GDP per capita or expenditure in education, rather than specific characteristics of schools (40%).

Our empirical strategy will closely follow the one in Cordero et al. (2017) for the use of FDH and order-m techniques, and the one of Agasisti and Zoido (2015) regarding the use and treatment of PISA data, while taking into account the results and limitations of the other relevant studies.

4. Methodology

4.1. The education production function frontier

The measurement of efficiency implies the use of a production function, i.e., a relation between the input and output factors. However, the definition of an education production function is not straightforward as the process by which the educational inputs are turned into outputs is generally unknown[1]. Notwithstanding, we can characterize a general production technology set (Ψ) through a set of inputs $x = (x_1, x_2, ..., x_p) \in R_+^p$ and a set of outputs $y = (y_1, y_2, ..., y_q) \in R_+^q$ such that:

$$\Psi = \{(x, y) \in R_+^{p+q} | x \text{ can produce } y\} \tag{4.1}$$

For efficiency measurement, however, the object of interest is the production possibilities frontier (PPF), i.e., the set of points that represent the maximum level of output for each combination of inputs. In order to characterize the frontier, it is helpful to separately characterize its corresponding production possibilities set (PPS) by its input and output requirements (Lovell, 1993), here denoted $C(y)$ and $P(x)$, respectively, and whose definitions are given by Equations 4.2 and 4.3 below[2].

$$C(y) = \{x \in R_+^p | (x, y) \in \Psi\} \tag{4.2}$$

$$P(x) = \{y \in R_+^q | (x, y) \in \Psi\} \tag{4.3}$$

From an output-oriented perspective, the PPF can thus be characterized by its isoquant as:

$$IsoqP(x) = \{y | y \in P(x), \lambda y \notin P(x) \forall \lambda > 1\} \tag{4.4}$$

Which defines the set of production bundles for which, if no additional inputs are added, no equiproportional increase of all outputs is feasible. Technical efficiency – or *in*efficiency – is thus captured by the radial distance λ in Equation 4.4. Those production bundles which are in the PPS but not in its frontier are – through the definitions –

[1] Please refer to section 2 for details.

[2] Following the same notation as in Daraio and Simar (2007b).

inefficient. Analogously, λ can also be considered through the Shephard (1953) output distance function (S_O) which is defined as:

$$S_O(x,y) = \min\left\{\lambda \middle| \left(\frac{y}{\lambda}\right) \in P(x)\right\} \tag{4.5}$$

And from which the output-oriented Debreu-Farrel measure of technical efficiency (Debreu, 1951; Farrell, 1957; Charnes et al., 1978), $\lambda(x,y)$, can be derived as its reciprocal:

$$\lambda(x,y) = \frac{1}{S_O(x,y)} = \max\{\lambda|\lambda y \in P(x)\} \tag{4.6}$$

With $\lambda(x,y) \geq 1$. One can also straightforwardly establish the correspondence between $\lambda(x,y)$ and $IsoqP(x)$ as:

$$IsoqP(x) = \{y|\lambda(x,y) = 1\} \tag{4.7}$$

Meaning that the production bundles that define the frontier, and thus considered efficient, will be those in which $\lambda(x,y) = 1$[3].

As pointed out in the section 2.2, several methods – parametric and non-parametric – have been developed to provide reliable estimates of $\lambda(x,y)$ in empirical contexts. Non-parametric methods have been extensively used in the estimation of education production frontiers, mainly through data envelopment analysis (DEA) (as first proposed by Charnes et al., 1978). However, non-parametric methods are mostly atheoretical regarding the assumptions about the shape of the production function. This feature makes them attractive for the characterisation of productive processes with 'black-box' characteristics, i.e., those in which the way the different inputs combine to produce a given output is largely unknown. Such is the case of the production of student outcomes in schools, that far from being mechanical in nature, implies complex human relations hard to parametrize in an education production function.

Notwithstanding those concerns, minimum assumptions have to be established. Assuming school activity can be flexibly conceptualized as a production process, Ψ in Equation 4.1, effectively relates input factors such as students' characteristics, teachers or the school facilities to such outputs as cognitive and non-cognitive skills. A series of axioms

[3] A similar reasoning can be applied for the case in which the input requirement $C(y)$ given in Equation 4.2 is considered.

on the properties of the PPS further help to characterize it.

Axiom 1 (Feasibility)**.** *The production bundle (x, y) is feasible, i.e., the output set y can be produced from the input set x.*

Axiom 2 (No free lunch)**.** *$(x, y) \notin \Psi$ if $x = 0$ and $y \geq 0$, i.e., no outputs can be produced without any input.*

Axiom 3 (Free disposability of inputs and outputs)**.** *If a given production bundle (x_0, y_0) is feasible, then (x, y_0) is also feasible $\forall x \geq x_0$. Similarly, if a given production bundle (x_0, y_0) is feasible, then (x_0, y) is also feasible $\forall y \leq y_0$.*

The three axioms above define the theoretical production set as a hull – not necessarily convex. In fact, convexity is no necessary condition for the characterisation of the PPS and its corresponding PPF[4]. There are well-defined instances in which the convexity axiom can be violated and therefore lose its economic meaningfulness. A seminal paper of Farrell (1959) identified production processes characterised by economies of scale, economies of specialization or where there are indivisible inputs or outputs as potential circumstances in which the PPS may not be convex. Recently, Mayston (2017) pointed out the conditions in which convexity is not verified in educational production functions, even though more directed to the relationship between research and teaching in higher education.

While DEA analysis estimates a convex hull, and different models apply further restrictions to the nature of returns to scale, Deprins et al. (1984) alternatively developed an estimator that does not require a convex PPS. We further describe this method in the next sub-section.

4.2. The FDH estimator

The free disposal hull (FDH) estimator measures the distance between a given production bundle (x_0, y_0) to the production frontier. In empirical applications, it estimates the distance between the position of the production bundle in the PPS and its radial projection in the technology frontier of the sample of production bundles $\chi = \{(X_i, Y_i), i = 1, ..., n\}$. The empirical free disposal hull PPS can thus be defined as:

$$\hat{\Psi}_{FDH} = \{(x, y) \in R_+^{p+q} | y \leq Y_i; x \geq X_i; (X_i, Y_i) \in \chi\} \tag{4.8}$$

[4]Convexity in production is generally assumed, both due to its theoretical underpinnings and the greater tractability, when using parametric methods. Convexity here would imply that if two production bundles are feasible then their linear combination is also feasible. See, e.g., Jehle and Reny (2001).

Or, alternatively,

$$\hat{\Psi}_{FDH} = \{(x,y) \in R_{+}^{p+q} | y \leq \sum_{i=1}^{n} \gamma_i Y_i; x \geq \sum_{i=1}^{n} \gamma_i X_i; \\ \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \in \{0,1\}; i = 1, ..., n\} \tag{4.9}$$

Following Deprins et al. (1984), Daraio and Simar (2007b, p. 34) note that the FDH set is «the union of the all positive orthants in the inputs and of the negative orthants in the outputs whose origin coincides with the observed points in $(X_i, Y_i) \in \chi$», being the smallest free disposal set containing all the observed production bundles[5]. Figure 4.1 provides a stylized representation of a free disposal hull production set, compared to a convex PPS.

Similar to Equation 4.3, the estimated output requirement is given by:

$$\hat{P}(x)_{FDH} = \{y \in R_{+}^{q} | (x,y) \in \hat{\Psi}_{FDH}\} \tag{4.10}$$

And the estimated frontier by:

$$Isoq\hat{P}(x)_{FDH} = \{y | y \in \hat{P}(x)_{FDH}, \lambda y \notin \hat{P}(x)_{FDH} \forall \lambda > 1\} \tag{4.11}$$

The output oriented efficiency measure of a given production bundle (x_0, y_0) is thus given by the estimated λ_{FDH}:

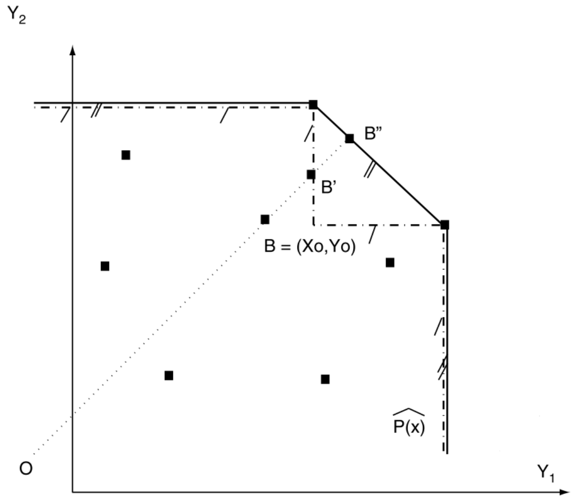
$$\begin{aligned} \hat{\lambda}_{FDH}(x_0, y_0) &= \max\{\lambda | \lambda y_0 \in \hat{P}_{FDH}(x_0)\} \\ &= \max\{\lambda | (x_0, y_0) \in \hat{\Psi}_{FDH}\} \end{aligned} \tag{4.12}$$

Therefore, from Equations 4.9 and 4.12, the efficiency scores FDH estimator is given

[5]The estimated FDH expression compares to the DEA convex counterpart, which in Banker et al. (1984) is defined as:

$$\hat{\Psi}_{DEA} = \{(x,y) \in R_{+}^{p+q} | y \leq \sum_{i=1}^{n} \gamma_i Y_i; x \geq \sum_{i=1}^{n} \gamma_i X_i; \\ \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0; i = 1, ..., n\}$$

The difference between the two PPS is given by the weights γ, for which in this case are not binary, being allowed to take values different than either 0 or 1, and endowing the PPF with its convex shape. However, the FDH estimator is a more general estimator of Ψ, as it does not require the convexity assumption.



Note: B' is the radial projection of the production bundle B in the FDH frontier and B'' is the same type of projection but in a convex DEA frontier. Source: Daraio and Simar (2007b, p. 36)

Figure 4.1: FDH and DEA estimation of $P(x)$ and $IsoqP(x)$ in a 2-output space.

by:

$$\hat{\lambda}_{FDH}(x_0, y_0) = \max\{\lambda | \lambda y_0 \leq \sum_{i=1}^{n} \gamma_i Y_i; x_0 \geq \sum_{i=1}^{n} \gamma_i X_i; \sum_{i=1}^{n} \gamma_i = 1; \quad \gamma_i \in \{0,1\}; i = 1, ..., n\} \tag{4.13}$$

$\hat{\lambda}_{FDH}$ is thus the solution to an integer linear program (Daraio and Simar, 2007b, p. 35). For the particular case of a sample χ of n schools, the efficiency of a given school is evaluated by the corresponding estimated score $\hat{\lambda}_{FDH}$ computed taking into account the full sample of schools' inputs and outputs $(X_i, Y_i) \in \chi$.

In practical terms, the estimator determines the set of observed production bundles in $\hat{\Psi}$ that weakly dominates the production bundle (x_0, y_0), i.e., that uses at most the same input or produces at least the same output. In the case of a 1-input 1-output production process, the efficiency score is computed with respect to the following weakly dominating set:

$$D_0 = \{i | (X_i, Y_i) \in \chi, X_i \leq x_0, Y_i \geq y_0\} \tag{4.14}$$

The efficiency score is thus computed, through an output oriented perspective, as:

$$\hat{\lambda}_{FDH}(x_0, y_0) = \max_{i \in D_0} \left(\frac{Y_i}{y_0} \right) \tag{4.15}$$

Which for the case of multiple outputs is computed through a max-min approach given by:

$$\hat{\lambda}_{FDH}(x_0, y_0) = \max_{i \in D_0} \left\{ \min_{j=1,\dots,q} \left(\frac{Y_i^j}{y_0^j} \right) \right\} \tag{4.16}$$

In which Y_i^j is the quantity of output j produced by the dominating school i, and y_0^j is the quantity of output j delivered by the school under efficiency evaluation, with $Y_i, y_0^j \in R_+^q$.

The technical efficiency scores of our empirical study are estimated by the FDH method, through Equation 4.16. As in Cordero et al. (2017), it is assumed that schools have the objective to maximize the cognitive abilities of their students and are not able to easily adequate their inputs in the short term. Therefore an output-oriented perspective is taken, where efficiencies are obtained through an expansion of output rather than a reduction of input.

4.3. Controlling for super-efficiency

An important limitation of FDH – and DEA – estimation relates to its deterministic nature. Statistical inference based on envelopment techniques is sensitive to extreme values or outliers (Cazals et al., 2002, p. 3). In particular, it is possible for the efficiency scores to be significantly affected by the existence of outlier 'super-efficient' production bundles, i.e., observations with abnormally high output quantities or abnormally low input quantities due, for instance, to measurement errors. The existence of a significant number of super-efficient observations in the sample biases the estimation – a production bundle (x_0, y_0) that would otherwise be efficient reveals an underestimated efficiency score if the dominating set D_0 exclusively contains units which are super-efficient in the above sense.

Parametric methods have been dealing with this problem mainly through maximum likelihood estimation of stochastic frontiers, dividing the error term in inefficiency and random noise (Aigner et al., 1977; Greene, 1980a). However, these require a parametric specification of the production function. On the other hand, non-parametric methods have been tackling this drawback mainly through partial frontier analysis, making use of a statistical version of Ψ.

Our empirical study also controls for super-efficiency by relying on additional partial frontier analysis models, namely order-m and order-α, which are now succinctly described.

4.3.1. *Order-m estimator*

The order-m estimation method was first introduced by Cazals et al. (2002) and, unlike the standard FDH estimator, does not consider the full frontier as the benchmark for the efficiency measurement. The technique consists in enveloping only a subsample of $m \geq 1$ observations that are randomly drawn with replacement from the set of observations with at least the same level of output.

Since the sub-samples are drawn with replacement, and given the statistical properties of the method, it is possible that a school with production bundle (x, y) has an efficiency score lower than the estimated partial frontiers' mean, which implies that $\hat{\lambda}_m(x, y) < 1$, being super-efficient.

The expected order-m frontier can thus be defined as «*the expected value of the maximum of m random variables $Y^1, ..., Y^m$ drawn from the conditional distribution function of Y given that $X \leq x$*» (Daraio and Simar, 2007a, p. 81). For the case with multiple outputs, and for a given number of inputs x and m i.i.d. random variables Y_i, with $i = 1, ...m$, the PPS can be defined for any y as:

$$\tilde{\lambda}_m(x, y) = \max_{i=1,\dots,m} \left\{ \min_{j=1,\dots,q} \left(\frac{Y_i^j}{y^j} \right) \right\} \tag{4.17}$$

The expected order-m output efficiency measure, $\lambda_m(x, y)$ is thus defined, for all x in which the distribution function of X is not zero, as:

$$\begin{aligned} \lambda_m(x, y) &= E\left(\tilde{\lambda}_m(x, y|X \leq x)\right) = \\ &= E\left[\max_{i=1,\dots,m} \left\{ \min_{j=1,\dots,q} \left(\frac{Y_i^j}{y^j} \right) \right\} \middle| X \leq x \right] \end{aligned} \tag{4.18}$$

Increasing the number m of observations randomly drawn from the sample of n schools for partial frontier analysis approximates the efficiency scores from those of the full frontier, as the $\lim_{m \to \infty} \lambda_m(x, y) = \lambda_{FDH}(x, y)$. Since the value of the estimator is not bounded, there will be schools which are super-efficient ($\hat{\lambda}_m(x, y) < 1$), for finite values of m[6]. Therefore, in empirical applications, the number of super-efficient production units is dependent on the number m the analyst chooses to benchmark each production unit

[6] The order-m estimator is also $\sqrt{n}$-consistent, assymptotically unbiased and assymptotically normally distributed. However, for a study of the assymptotic properties of the order-m estimator please refer to Cazals et al. (2002).

with.

The method is then repeated B times, thus benchmarking each observation to B randomly drawn partial frontiers. The order-m efficiency measure is finally computed as the simple mean of the estimated distances. According to the algorithm in Daraio and Simar (2007a, pp. 82-83) the order-m estimation can be summarized in four steps, namely:

1. From the set of schools that produce at least as much of any output as school i, a sample of m schools is drawn randomly with replacement.
2. Pseudo FDH efficiency scores, $\tilde{\lambda}^b_m(x, y)$, are computed through Equation 4.17.
3. Steps 1 and 2 are repeated B times, for $b = 1, ..., B$, with B large.
4. Order-m is computed as the average $\hat{\lambda}_m(x, y) \approx \frac{1}{B}\sum_{b=1}^{B} \tilde{\lambda}^b_m(x, y)$.

The choice of B is then a matter of accuracy. The greater the B the more accurate will be the approximation, which comes however, at the expense of a larger computation time (Tauchmann, 2011, p. 4).

In economic terms, the interpretation of the estimates should be straightforward. A given school with an estimated order-m efficiency score of $1/z$ produces z times the estimated maximum attainable output from a set of m other schools drawn randomly from the empirical sample.

4.3.2. *Order-alpha estimator*

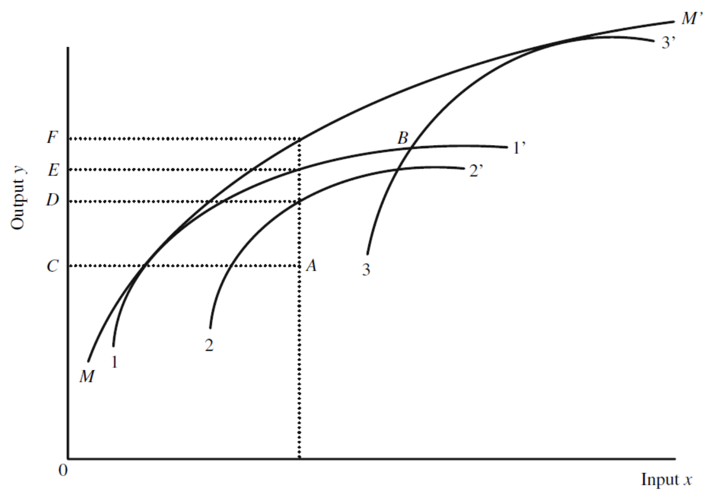
Similarly to order-m, Aragon et al. (2005) developed an estimator based on re-sampling techniques, however following a quantile-based approach. According to the order-α method a given school with production plan (x, y) is evaluated against the frontier defined by the output level exceeded by $(1 - \alpha) \times 100\%$ of the schools that use x or lower levels of each input. Following the notation in Tauchmann (2011, p. 4) the output-oriented order-α estimator can be written as:

$$\hat{\lambda}_\alpha(x, y) = P_{(1-\alpha)} \left\{ \min_{j=1,\dots,q} \left(\frac{Y_i^j}{y^j} \right) \Big| X \leq x \right\} \tag{4.19}$$

Where $P(1 - \alpha)$ denotes the $\alpha \times 100$th percentile of schools where $X \leq x$. For $\alpha = 1$ the estimator coincides with FDH, while for $\alpha < 1$ some observations will be classified as super-efficient. If $\hat{\lambda}_\alpha(x, y) = 1$ the school is efficient at the $\alpha \times 100\%$ level in the sense that it is dominated by other schools using the same or less input than x with only a probability of $1 - \alpha$[7].

4.4. Metafrontier approach

In international analysis of school efficiency – as the one developed here – data are generally hierarchical. Schools can be both evaluated in comparison with other schools operating in the same country or with other schools operating in different education systems. Extending the approach followed in Portela and Thanassoulis (2001) and Thanassoulis and Portela (2002), and in the spirit of O'Donnell et al. (2008), two different types of production frontiers can be estimated: *i)* k local frontiers for the sub-samples of n_k schools in each kth country in the full sample; and *ii)* an international best practice frontier for the whole sample of n schools. It is assumed that schools in the same country operate under a similar set of institutional rules than schools from different countries, being therefore submitted to more similar constraints. Therefore, the distance of a given school to its respective local frontier can be conceptualized as the part of the overall efficiency score that can be attributable to school efficiency (ScE), or the country-level efficiency score, while the distance from the local frontier to the international metafrontier can be interpreted as the country effect (CE) (Cordero et al., 2017, p. 367). Figure 4.2 provides an illustration of a metafrontier enveloping 3 local frontiers in a 1-input 1-output scenario.



Source: O'Donnell et al. (2008, p. 236)

Figure 4.2: Metafrontier illustration with 3 local frontiers

[7]The order-α estimator has the same statistical properties of the order-m estimator, i.e., it is assymptotically unbiased, normally distributed and $\sqrt{n}$-consistent (Aragon et al., 2005). Nevertheless, Daouia and Simar (2007) show that the order-α method is more robust to extreme values.

1-B-3' represents the non-convex metafrontier enveloping the local frontiers, if it is assumed that these are all the k separable groups. Otherwise, the metafrontier can be theoretically conceived as the *M-M'* frontier, as there may be other feasible production bundles (O'Donnell et al., 2008, p. 235).

Taking an output-oriented perspective, the efficiency scores resulting from benchmarking the inefficient production bundle denoted by A and with output C in Figure 4.2 against its corresponding local frontier *2-2'* and the international frontier *1-B-3'* can be computed as:

$$\lambda_A^{GE} = \frac{0E}{0C} \tag{4.20}$$

Where λ_A^{GE} denotes the global efficiency score of A and $0E$ and $0C$ the distances between the origin and coordinates C and E, respectively. On the other hand,

$$\lambda_A^{ScE} = \frac{0D}{0C} \tag{4.21}$$

Which is the ratio between the potential output defined by the local frontier and the actual output C.

The country effect of unit A (λ_A^{CE}) is finally given by:

$$\lambda_A^{CE} = \frac{0E}{0D} = \frac{0E/0C}{0D/0C} = \frac{\lambda_i^{GE}}{\lambda_i^{ScE}} \tag{4.22}$$

From Equation 4.22, and generalizing, global efficiency can be easily derived as $\lambda_i^{GE} = \lambda_i^{ScE} \times \lambda_i^{CE}$.

4.5. Explaining the variation in school efficiency

After accounting for the decomposition of the global efficiency scores in within-country school efficiency and country effects the relevance of other variables in explaining the distribution of efficiency scores is ensued. Educational activity is influenced by institutional and contextual factors that hinder or catalyse efficient allocations of school resources. The introduction of environmental variables, $z = (z_1, ..., z_h) \in R^h$, in the analysis enables to investigate what is the impact of given factors affecting the organization of school activity on their efficiency in providing quality education. Given the nature of the data used in this study three separate subsets of variables are considered: school characteristics (z^{schl}), student characteristics (z^{std}) and school-level policies (z^{pol}).

The second-stage analysis was performed using parametric methods, namely by regressing the estimated efficiency scores on the sets of covariates, such that:

$$\hat{\lambda}_i^{GE} = z_i\beta + \epsilon_i, i = 1, ..., n \tag{4.23}$$

Where $\hat{\lambda}_i^{GE}$ are the global efficiency scores estimated in the first stage, β is a vector of coefficients determining the marginal linear impact of the environmental variables on the efficiency scores and ϵ_i is the error term.

However, several approaches have been discussed in the literature on efficiency estimation[8]. The evaluation of the impact of z on the variation of efficiency can be summarized in three different families. A one-stage approach, where z is assumed to shape the PPS and where the environmental variables are included in the estimation of the FDH frontier, such that $\Psi \subset R_+^p \times R_+^q \times R^h$ (Banker and Morey, 1986a,b). In this case, variables which affect efficiency favourably are considered as inputs of the production process and as outputs if the effect is assumed to be unfavourable. The main limitations of this approach relate to the added postulates about the technology, namely the free-disposability of the extended FDH set and the assumption about the sign of the effect of the environmental variable on efficiency (Daraio and Simar, 2007a, p. 98)[9]. A third limitation relates to quantify the effect of z on the change in efficiency, i.e., what would be the marginal effect on efficiency were the environmental conditions more favourable.

A second family of methods has been developed by Daraio and Simar (2005) and also considers a fully non-parametric approach, while differing in the way environmental factors are considered. In this case, the statistical joint distribution of the input and output variables is assumed to be conditioned on z and the effect of the favourable or unfavourable effect of the environmental variables can then be assessed using appropriate non-parametric methods, namely through kernel density estimation. Despite its statistical robustness, this approach also holds some drawbacks. First, it implies additional discretion by the analyst in the choice of the appropriate bandwidth for the kernel function. Second, despite being possible to assess the statistical significance of z on efficiency

[8] Please refer to Daraio and Simar (2007a, pp. 95-100) for a review of the main methods.

[9] Nevertheless, this method is partially employed as part of the empirical strategy, as a model which includes the average socio-economic background of students – a variable not directly controlable by the school managemet – as an input of the production process. In this sense, the estimated efficiency scores attempt to capture the best a given school can do given the socio-economic characteristics of its student population.

and its negative or positive sign, it is not possible to obtain the marginal impacts on the efficiency scores[10].

Finally, a multiple-stage approach with parametric estimation in the second stage of the analysis was the methodological choice for this study. In this case, the first-stage estimated efficiency scores are regressed in the set of environmental variables through adequate limited dependent variable model (e.g., tobit and truncated model), as their distribution has a lower bound in 1 (for the full frontier output-oriented case). Nevertheless, the approach followed has important limitations. Simar and Wilson (2007) note that the $\hat{\lambda}_i$'s are serially correlated with the error terms in the second stage in an unknown way, and are not independent and identically distributed (i.i.d.)[11], meaning that standard estimation is invalid, such as in Equation 4.23. Moreover, the technique implies a separability condition between $\Psi \subset R^p_+ \times R^q_+$ and $z \in R^h$. Therefore, the z are not assumed to affect the position of the PPF but rather the distance between the inefficient units from the frontier (Daraio and Simar, 2007a, p. 100). In this sense, the second-stage regression aims at finding the significant factors that make inefficient units closer to the frontier.

In order to overcome these concerns, we follow Simar and Wilson (2007, pp. 41-42), where a truncated regression with bootstrapped coefficients and standard errors is proposed. The method can be summarized in the following steps[12]:

1. From the sample of $\chi = \{(X_i, Y_i), i = 1, ..., n\}$ estimate $\hat{\lambda}_i^{GE}$ $\forall i = 1, ..., n$ through Equation 4.13.
2. Use the method of maximum likelihood estimation to obtain an estimate $\hat{\beta}$ of β and $\hat{\sigma_\epsilon}$ of σ_ϵ (the standard errors) in a truncated regression of $\hat{\lambda}_i$ on z_i in Equation 4.23 using the $r < n$ observations where $\hat{\lambda}_i > 1$.
3. Loop over steps (3.1.-3.3.) L times to collect a set of bootstrapped estimates $\Upsilon = \{(\hat{\beta}^*, \hat{\sigma_\epsilon}^*)_b\}_{b=1}^{L}$

   3.1. For each $i = 1, ..., r$ draw ϵ_i from a left truncated normal distribution

[10] Cordero et al. (2017) opt for this method to assess the effect of environmental variables on estimated order-m efficinecy scores.

[11] As the $\hat{\lambda}_i$ are calculated based on the relative position of points in the input-ouput space they are necessarily not mutually independent.

[12] In particular, Algorithm #1, from the two presented in Simar and Wilson (2007), is the one used due to computational reasons. Algortithm #2 differs from the first by also employing bootstranpping techniques in the first-stage efficiency estimation. However, FDH esimtation, but especially order-m is computationally cumbersome for large samples (using the current estimation technology). As an example, one of the order-m international frontier estimations took over 12 hours to complete. While bootstrapping helps to correct for eventual bias of the estimator, computational time would increase exponentially. Due to the large size of the sample it was chosen to just use bootstrapping techiques in the second stage of the analysis, and thus Algorithm #1 instead of Algorithm #2. For practical computation of the coefficients the *simarwilson*.ado routine was employed, using Stata 13.1.

3.2. For each $i = 1, ..., r$, compute $\lambda_i^* = z_i\hat{\beta} + \epsilon_i$

3.3. Use the maximum likelihood method to estimate the truncated regression of λ_i^* on z_i, with estimates $(\hat{\beta}^*, \hat{\sigma_\epsilon}^*)$

4. Use the bootstrap estimates in Υ and the original estimates $(\hat{\beta}, \hat{\sigma_\epsilon})$ to construct estimated confidence intervals for each element of β and the σ_ϵ's

In order to increase the accuracy of the estimates, we will perform $L = 2000$ bootstrap repetitions in the preferred specifications, as suggested by the authors (Simar and Wilson, 2007, p. 44). For the estimation process, the relevant subsets of covariates $(z^{schl}, z^{std}, z^{pol})$ were sequentially introduced as blocks using a backward iterative procedure to only keep the variables significant at a 10% significance level and maximize the explanatory power of the model.

Finally, the analysis was further extended to study the effects in the efficiency scores computed taking into account each k country frontiers $(\lambda_{k,i}^{ScE})$, which enabled to evaluate if the same effects would still verify for within-country variations of efficiency.

5. Data and Variables

5.1. PISA dataset and sample

The empirical analysis relies on the 2015 Programme for International Student Assessment (PISA 2015), conducted by the Organization for Economic Co-operation and Development (OECD)[1]. PISA, conducted every three years since 2000, are conceived to obtain internationally comparable data on student achievement in three broad topics: Maths, Science and Reading[2]. The assessment is based on a standardized achievement test targeted to 15 year-old students irrespective of the grade or type of school attended[3]. The test aims to evaluate different cognitive dimensions – e.g., in the case of science, content, procedural and epistemic knowledge are considered[4].

In order to draw a representative sample of the 15 year-old students population in each country, a two-stage sampling procedure is used. In a first stage, schools are randomly selected based on the number of 15 year-old attending. In a second stage, a random sample of the students in each of the initially chosen schools is selected[5].

Student performance in each subject is measured through a psychometric scale with mean of 500 test-score points and standard deviation of 100 across countries belonging to the OECD group[6]. Each student is tested for a broad array of topics in the concerned subject with differing levels of difficulty. As the participants only answer to a representative portion of the complete test, a set of ten *plausible values* for the complete test score is drawn for each student based on a given distribution of performance[7].

Besides the data on achievement a large set of characterising variables is collected.

[1]For a complete description of the results see OECD (2016a).

[2]The 2015 edition also included testing in Financial Literacy and Problem Solving, besides indicators of self-percepted student well-being.

[3]However, PISA tests are not taken by 15 year-old students which are enroled below grade 7. The schools attended include those with general but also vocational curriculums.

[4]For a more complete description of PISA please refer to (OECD, 2016a, pp. 25-32).

[5]For a more complete understanding of the sampling procedures refer to OECD (2016b, Chp.4), available at `http://www.oecd.org/pisa/data/2015-technical-report/`.

[6]Not only OECD countries participate in the assessment. The international impact of the study has been increasing the group of participating countries which now also includes lower- and upper-middle income countries. In the 2015 edition 37 other partner economies also participated in the assessment. The performance results are scaled to have a mean of 500 and a standard deviation of 100 in the first year in which the subject assessment is made

[7]The plausible values can be interepreted as the ability range for each student, being randomly obtained from the distribution function of the test results.

Students and their families are asked to provide a series of information on individual characteristics and social and economic background. A questionnaire is also filled out by the principal of each of the selected schools about the available resources, the school's characteristics and its institutional setting.

Several steps were taken in the construction of the dataset. First, the student's PISA scores were combined with the information of the questionnaire on their individual characteristics and socio-economic background, through a student unique identifier common to both datasets. Given the profusion of information, only a set of relevant variables for the study was selected, including composite variables. Second, the dataset with school-level information was merged with the dataset with student-level variables through a school unique identifier. Finally – as the empirical analysis would be performed at the school-level – school-level means on performance and the other student background variables were computed for each of the schools. In order to retain the heterogeneity of the student population composition of each school, an extended collection of indicators were computed, including the standard deviation of PISA test scores in the school, and a characterization of the student performance and socio-economic background distributions through several percentiles and inequality measures, such as S90/S10 or S75/S25. As the psychometric scale of the PISA test-scores is adjusted to student-level analysis, not school-level, care should be taken in interpreting the results.

Despite the general reliability of the data, extensive quality checks were performed, namely for the existence of implausible values in key-variables. Since a large set of explanatory variables is considered in the second stage-regression, the existence of missing values was also common across both the student- and school-level datasets. As dropping all students or schools with missing values would result in significant sample reduction, a careful analysis of its pattern was performed. Only students or schools with a number of missing values above previously defined thresholds of reliability were dropped. Therefore, all observations for which there were more than five missing values in the key variables were dropped. In general, the choice of the core variables to keep for the second-stage empirical analysis was driven by the pattern of missing values, where variables with a high share of the latter would be more carefully considered for inclusion. The sample reduction was always followed by an analysis on the similarity of the distribution of the variables, both in the pooled sample and by country. All the observations with missing values in the inputs or outputs chosen were kept out of the final sample. Also, all schools with no computers were also not included. Restricting to OECD countries, 8.35% of the individuals were dropped due to missing values, a total of 23 885 students. In the case of the school-level dataset, 19.39% of the schools had to be dropped from the original sample, 1 274 of them due to lack of information in student-teacher ratios. The final dataset includes 7318 schools from 34 OECD countries[8].

The most important limitation of PISA data for the study of school efficiency relates to its cross-sectional nature. As there are no data tracking the achievement of each individual through the years, it is not possible to more accurately disentangle the added-value of each school for its students' cognitive skills from their innate ability or the influence of other schools in their educational path. Further, given the high percentage of missing values in the input and output variables, and given that no data imputation procedures were used, the sample set of schools is most likely not fully representative of the country. Therefore, the results must be weighed by these considerations.

5.2. Data description

Table 5.1 summarizes the number of schools by each country in the sample. Restricting the efficiency estimation to this group of countries enables to compare education systems relatively more similar in the rates of 15 year-olds participating in formal schooling and level of investment in education. Even so, there are significant differences in the schools' operational environment, the regulatory frameworks and the performance of students[9]. The number of schools included in the analysis also varies substantially across countries. As for the case of the original PISA dataset, the number of schools per education system is not exclusively dependent on the size of the country but rather the variability the school selection may capture according to the sampling procedure[10].

The choice of the inputs and outputs to compute the efficiency scores followed the variables usually considered in the literature on school efficiency[11], while being constrained

[8]Luxembourg was not considered in the final sample as the number of schools for the efficiency estimation was too small.

[9]Other authors have been also referring the relevant differences in cultural practices and attitudes across countries. For instance, Cordero et al. (2017) use a set of self-percepted measures of responsibility and motivation, collected by the World Bank, in order to control for differences in students' attitudes across countries. The PISA dataset contains simple and index variables of student motivation, but the high share of missing values makes their inclusion in the analysis less reliable.

[10]Please refer to OECD (2016b, Chp.4), available at `http://www.oecd.org/pisa/data/2015-technical-report/`.

[11]Please refer to section 3.2.

Country	Number of schools	Country	Number of schools
Australia	607	Iceland	197
Austria	226	Israel	167
Belgium	219	Italy	173
Canada	584	Japan	209
Switzerland	196	Korea	86
Chile	304	Latvia	133
Czech Republic	235	Mexico	170
Germany	156	Netherlands	154
Denmark	152	Norway	161
Spain	208	New Zealand	263
Estonia	167	Poland	268
Finland	191	Portugal	176
France	211	Slovakia	179
United Kingdom	85	Slovenia	190
Greece	148	Sweden	138
Hungary	141	Turkey	367
Ireland	313	USA	144
		Total	**7318**

Table 5.1. Sample number of schools by country

Variable	Obs.	Mean	Std. Dev.	Min	Max
Outputs					
Math scores (*pv1math*)	7318	490.86	58.02	229.48	672.67
Science scores (*pv1scie*)	7318	495.07	60.94	257.03	677.53
Inputs					
Teacher-student ratio (*tsratio*)	7318	0.09	0.07	0.01	1
Computer-student ratio (*comp*)	7318	0.90	0.91	0.003	34.70
ESCS (*escs*)	7318	6.17	1	1	8.60

Table 5.2. Summary statistics of inputs and outputs

by data availability. Due to a parsimony principle, a limited number of inputs and outputs is used in the first-stage estimation as suggested in Johnes (2004, p.655).

Table 5.2 presents summary statistics for the variables considered in the baseline first-stage estimations. As outputs, *pv1math* and *pv1scie* are the average PISA test-scores in Mathematics and Science of the students in each school. The averages are computed using *plausible value* 1, from the 10 available[12]. Alternative models using other plausible values as outputs – and the test-scores in reading – are also considered for robustness checks to the results. In the initial baseline model, the inputs for the efficiency estimation include the teacher-student ratio (*tsratio*) and the computer-student ratio (*comp*) at each school, as proxies of human and material resources intensiveness, respectively. Other specifications of the educational production function consider the family background of students through the average of the economic, social and cultural status of the students in the school (*escs*).

[12]Please refer to Annex B: Summary Statistics by Country for summary statistics by each country in the sample. Due to the use of only one of the plausible values for computing the averages, the results should not be strictly compared to the final PISA results and rankings presented in OECD (2016a).

ESCS is a composite score including the influence of multiple dimensions, namely parental education, parents' professional occupation and household possessions, as a proxy for wealth. The values are computed through principal factor analysis and scaled to the average OECD student to have a score of 0[13]. However, for the computation of the efficiency scores, the variable was re-scaled to only assume positive values and a standard deviation of 1.

The averages of Math and Science scores are close to the OECD student-level average of 490 and 493, respectively (OECD, 2016a, p. 44). There is considerable variation across countries (Table 10.1,Annex B: Summary Statistics by Country). The standard deviations are, however, below the OECD student-level results, as average school performance varies less than individual student achievement. Importantly, the standard deviations of the scores are also heterogeneous in a country-by-country basis. For instance, both Japan and Estonia have, on average, relatively higher-performing schools. However, an analysis of their standard deviations leads to conclude that the distribution is more concentrated in the latter than in the former. Such variations in the distribution of the scores by country help justifying the relevance of investigating the variability in performance and efficiency within each country.

Regarding the inputs, schools in the sample have an average of around 10.6 students per teacher and 0.9 computers per student. As for the case of outputs, average inputs vary considerably across the countries in the sample (Table 10.2Annex B: Summary Statistics by Country). While Mexican schools have, on average, the lowest amount of teachers per student, central European countries, such as Slovenia, Hungary and Poland present the highest average human resource intensiveness. In particular, the standard deviation in Slovenia is almost three times the standard deviation of the full sample. The average economic, social and cultural status of student populations also varies substantially across and within countries. Student populations in Turkey and Mexico are considerably disadvantaged in relation to the rest of the sample. On the other hand, the socio-economic composition of students in Iceland and Norway simultaneously register relatively low variability across schools and the highest mean values in this indicator.

The variables for the second-stage regression are separated in three types: students' characteristics (z^{stud}), schools' characteristics (z^{schl}) and school-level policies or prac-

[13] For methodological details please refer to OECD (2016b, Chp. 16), available at `http://www.oecd.org/pisa/data/2015-technical-report/`.

tices (z^{pol}). Annex C: Descriptive Statistics presents summary statistics for each of the variables. Given the exploratory nature of the empirical analysis, an extensive list of variables is listed, aiming to reflect both school averages and within-school variation, through standard deviations.

The set of students' characteristics aims to understand what are the factors related to the student population composition distancing relatively more inefficient schools from the ones defining the frontier. As a very comprehensive measure of students' socio-economic background was already included as an input of the educational production process, other (at most weakly correlated) variables were considered for analysis. These include the school standard deviation of students' parents education (*sd_pared*) and wealth (*sd_wealth*), the proportion of females at school (*prop_fem_school*), a composite index of the students' cultural possessions (*cultposs*) and its respective within-school standard deviation (*sd_cultposs*). The index of cultural possessions, is computed, similarly to ESCS, through principal factor analysis. The index is computed based on answers to questions related to the availability of household items related to culture (such as books or paintings). The standard deviation of the economic, social and cultural status variable (*sd_escs*) was also computed for each school as a proxy of the diversity of the enrolled students. Other variables – often considered in the literature – were also included as student characteristics. In particular, we separately assess the effect of having first generation and second generation immigrants (*prop_immig_1* and *prop_immig_2*). On average, 5.5% of students taking the PISA test in the sample schools are immigrants, while 6.5% are second-generation immigrants. Similarly, we have computed the proportion of students that have repeated at least 1 year in their school path. On average, 10% of the students have repeated at least a year in the average sample school. Across countries it varies from as high as 26% in Spain to 0% in the case of Norway or Japan, which do not apply year retention policies.

The set of school characteristics was mainly derived from the questionnaires to the principals of the schools participating in PISA 2015. It includes dummies indicating if the school is private (including publicly dependent and independent) (*private*), if it is located in a rural area or small town[14] (*rural*). Similar variables have been verified in the literature on education as being significantly correlated with student performance and school

[14]The measure for rural location includes the pre-established categories rural (less than 3 000 inhabitants) and small town (less than 15 0000 inhabitants).

efficiency[15]. In fact, running simple regressions having the test-scores in mathematics as dependent variable and each of the three dummies as covariates shows that rural schools and schools with small classes – irrespective of their location – perform significantly worse, while private schools perform better than public ones[16]. In the sample, 15.6% percent of the schools are private while 31.9% are located in small towns or the rural side (Annex C: Descriptive Statistics). Nevertheless, the school offer profile differs markedly across education systems. While about a peak 61% of Chilean schools in the sample are private, more than 95% of schools from countries like Norway, Estonia, Greece or Poland are public[17]. Other school characterising variables were the total number of enrolled students (*enrol*), and composite indices of quality of educational leadership (*lead*), adequacy of educational staff (*staffshort*) and the adequacy of educational materials at school (*edushort*)[18]. Finally, standard deviations were computed for the performance at mathematics (*sd_pv1math*) and science (*sd_pv1scie*), besides within-school inequality measures of performance, as proxies for equity in educational achievement. As in most variables, the inter-decile ratios of student outcomes (*s90_10_pv1math* and s90_10_pv1scie) within schools are heterogeneous among countries. Average inequality in results within schools is the largest in Iceland (1.63) and Sweden (1.62),significantly above the sample average of 1.5. This implies that, on average, students in the 90th percentile of the achievement distribution within each school in those countries have results about 60% higher than those in the 10th percentile.

Finally, the set of school-level policies and practices includes variables taken from the school principals' questionnaire. Continuous variables considered are a composite index of school autonomy (*schaut*), derived from the responsibility of school leadership in firing or hiring teachers, or setting the curriculum. A set of dummies were also derived for the second-stage regression, such as if student achievement data are posted publicly (*achv_public*), if school leader decides about budget allocation within the school (*bud-*

[15]Please refer to the Literautre Review section.

[16]Results are available at request. On average, rural schools perform 14 test-score points below urban and schools with small classes perform between 17 and 29 test-score points lower – for a 95% confidence level. Private schools score between 14 and 22 points higher than the average public school. However, when the results are controlled for the socio-economic background of students the performance differences between private and public schools become insignificant.

[17]Summary statistics of the second-stage regression variables by country are available at request.

[18]The composite indices were all computed through principal factor analysis based on answers from the school principals. Please refer to Chapter 16 of the forthcoming PISA 2015 technical report for methodological details.

get_alloc) or if teachers have autonomy to define the courses' content (*curricu_teach*). Furthermore, variables on the responsibility for determining which courses are offered (*curric_offer_p*, *curric_offer_t*) are also included in the analysis. School-level policies in terms of students' grouping according to ability are also assessed, namely in different classes (*ability_out*) or within the same class (*ability_in*), or if its classes have, on average, 15 or less students (*small_class*). Finally, it is investigated what is the impact of having external inspection as a way to monitor teachers' practices (*teachevl_insp*) in the explanation of efficiency differences across schools. As in the case of schools' characteristics, school-level policies variables also vary considerably across countries – which is jointly analysed with the results of the empirical analysis.

6. Results

6.1. Analysis of efficiency distribution

6.1.1. *Raw efficiency*

Figure 6.1 presents the global and country-level efficiency scores computed through FDH, using average achievement at Maths and Science as outputs and teacher-student and computer-student ratios as inputs[1]. The results are presented by country and ordered by average global efficiency. According to the method, schools in the international sample could have, on average, expanded their output by about 25.2%[2] given the same level of inputs[3]. Nevertheless, potential efficiency gains vary across countries. While Greek inefficient schools could have, on average, increased output levels by 32.4% holding inputs constant, inefficient Japanese schools could have had gains of 20.3% were they operating at the frontier.

As addressed in section 4.4, global efficiency measures the distance between each inefficient school and an enveloping international frontier. However, given institutional differences across education systems, schools benchmarked as inefficient at the international level can be otherwise operating at the technological frontier of their own country. Figure 6.1 presents country-means for global efficiency scores – i.e., the distance of each school relative to the international PPF – and country-level efficiency scores – i.e., the distance of each school relative to its respective national efficiency frontier. When national PPFs are considered, Iceland is the country where schools are, on average, operating the closest to the national frontier (1.064). This result clearly contrasts with the position of Iceland in the global efficiency ranking, meaning that the difference to the enveloping international frontier is mostly explained by the country effect (1.304) – i.e., the distance

[1]Table 12.1, in Annex D: Summary of Efficiency Scores provides more detailed summary statistics by country.

[2]Given the mean global efficiency of 1.337. The figure can be easily computed as $1-1/1.337 = 0.252$. When only inefficient schools are taken into account the glob+al efficiency average is 1.3406, meaning a potential expansion of 25.4%

[3]These results are in line with those of Agasisti and Zoido (2015), where inefficient schools could have increased the output by 27%. However, comparability is limited. Agasisti and Zoido (2015) use a different set of countries, with data from the PISA 2012 survey. Furthermore, the authors use DEA – assuming the convexity of the educational production function. Finally, the computation of the efficiency scores reported by the authors include an additional input, namely the average socio-economic background of the students of each school, which will only be introduced in other specifications of our empirical specification.

between the international and international frontier. The Finish system, in turn, has schools simultaneously operating relatively close to the national frontier and the least disperse distribution of country-level efficiency scores, as measured by their standard deviation (0.056). On the other hand, the Hungarian school system presents the worst results in terms of bringing inefficient schools close to their national frontier (1.313), also having the highest standard deviation (0.222). Across the full sample of countries, schools could have expanded their output by 17% if operating at the national PPF (18.3% if only considering inefficient ones)[4].

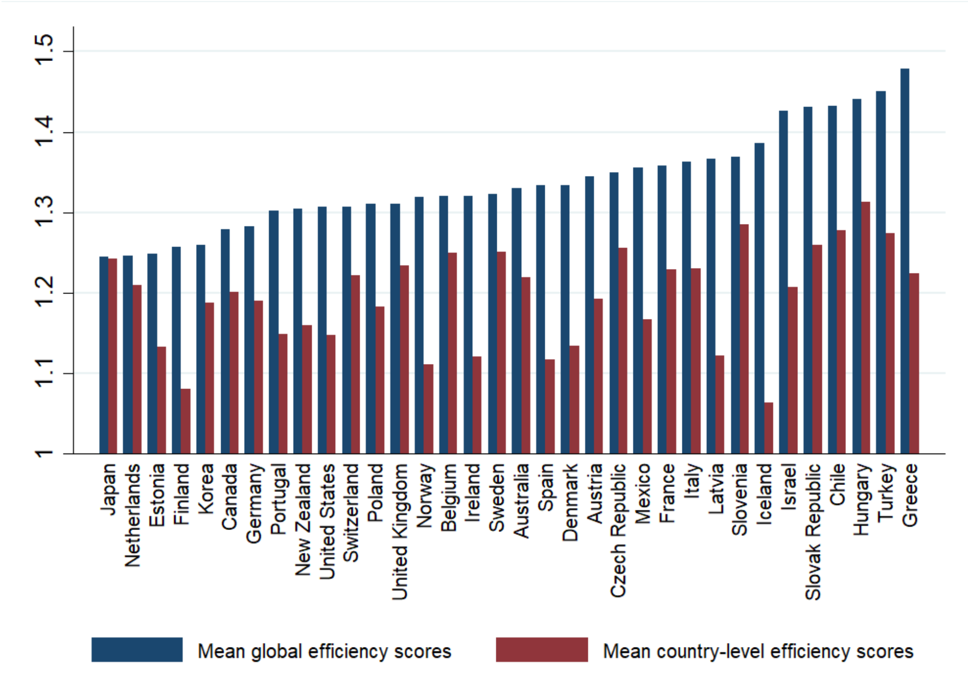


Figure 6.1: Mean global and country-level efficiency scores by country (raw efficiency)

In the first specification of the model, school effects are, on average, larger than country effects. Therefore, according to the metafrontier approach, the results seem to indicate that differences across schools within countries contribute more to global inefficiency, than differences across national PPFs. However, these results are dependent on a specific understanding of efficiency. In particular the computation of the efficiency scores does not account for the characteristics of students in each school. In the next section we will try to extend the results by providing an alternative specification, where average socio-economic background of the students in each school is taken into account in the concept of school efficiency.

[4]Table 12.1 in Annex D: Summary of Efficiency Scores presents the full decomposition of efficiency scores by country, including their standard deviation.

6.1.2. *Accounting for socioeconomic background*

Students' socio-economic background has been consistently shown to have a significant impact on school achievement (Witte and López-Torres, 2017) (see section 3 for a discussion). Therefore the initial model was extended to take the differences in the composition of the schools' student intake characteristics into account. Taking advantage of the PISA index for cultural, economic and social background (ESCS)[5], each schools' efficiency score was re-computed including its average ESCS as an additional input. Therefore, school efficiency is here conceptualised as the ability of each school to provide quality education given the available resources *and* the socio-economic characteristics of its students. Figure 6.2 depicts the new efficiency scores by country.

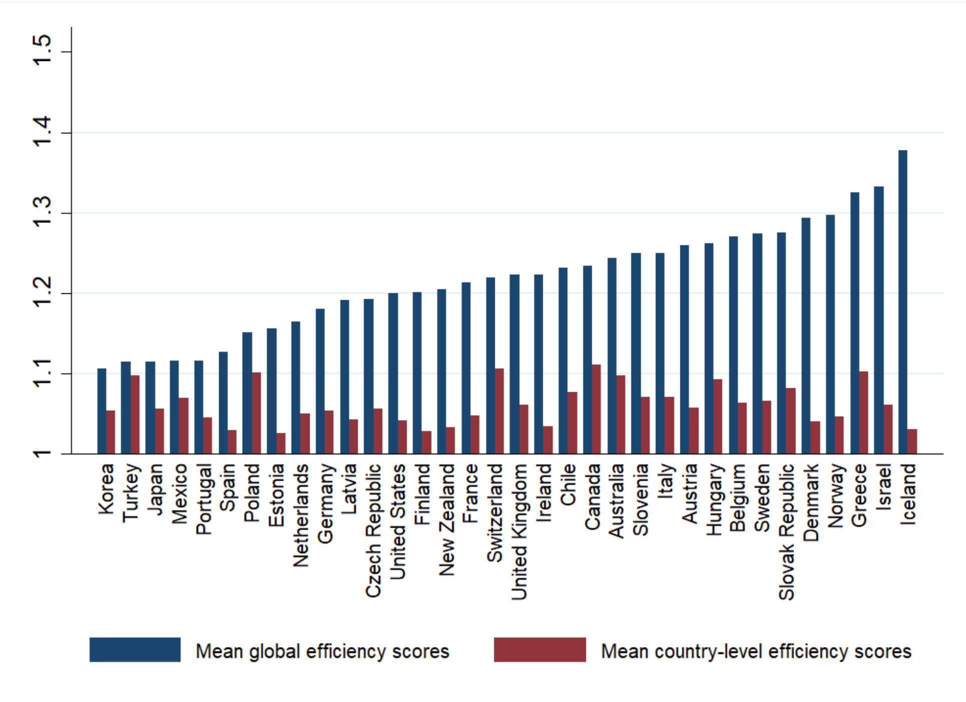


Figure 6.2: Mean global and country-level efficiency scores by country (after accounting for ESCS)

When technical efficiency takes into account average ESCS, most schools move closer towards national and international PPFs, as the definition of efficiency is now broader and cognisant of the social environment in which the school operates. This movement towards the international frontier depends on two factors. The first pertains to the introduction of an additional input. As discussed in section 3.1, efficiency scores depend on the number of inputs and outputs in the analysis. Due to the introduction of an additional input, no

[5]Check section 5.2 for a brief explanation of this index

school has lost on efficiency. The second factor relates to the negative effect of the average socio-economic background of schools. As Figure 6.3 depicts, the gains in efficiency were not uniform to all schools – schools where the ESCS was less favourable also had higher efficiency gains from the inclusion of that input in the analysis.

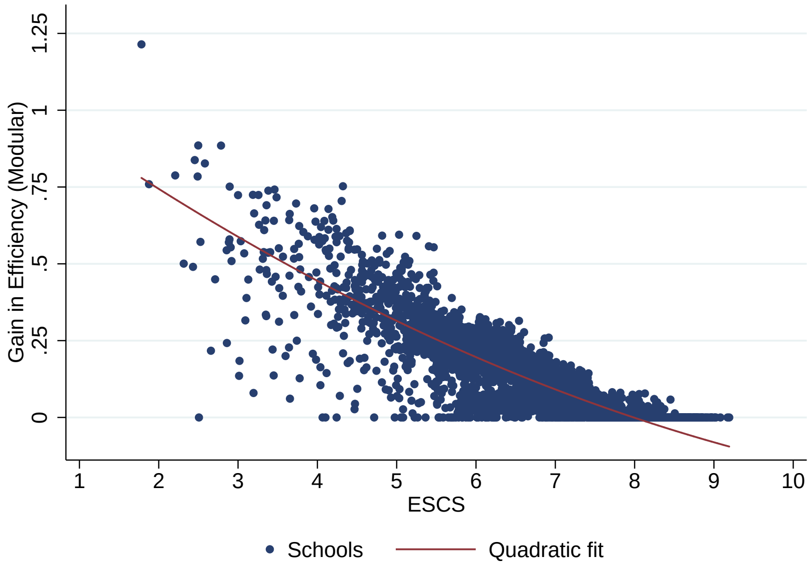


Figure 6.3: Effect of inlcuding ESCS in the efficiency analysis

In this specification of the model, schools could have – on average – expanded the output by 18.2% were they operating at the international PPF. The new evaluation corresponds to an average absolute gain in efficiency of 0.093 compared to the case in which average ESCS is not included in the model.

According to the specification, the estimated international efficiency frontier is defined by 3.3% of schools in the sample. Most of these are located in Mexico and Turkey (25% of their schools are in the efficiency frontier), where the contextual conditions for the operation of schools are harsher. But countries where the socio-economic background of students is relatively more favourable on average also contribute with several schools for the international efficiency frontier. Japan, Korea, Portugal, the Netherlands or Canada all have 10 or more schools defining the international PPF. In total, 23 different countries have schools which are globally efficient ($\lambda^{GE}_{FDH} = 1$) (Annex D: Summary of Efficiency Scores, Table 12.5).

The distribution of global efficiency scores is also heterogeneous within countries. As Figure 6.4 depicts, different types of distribution can be found across the countries, providing greater insight to the characteristics of school efficiency in each of them.

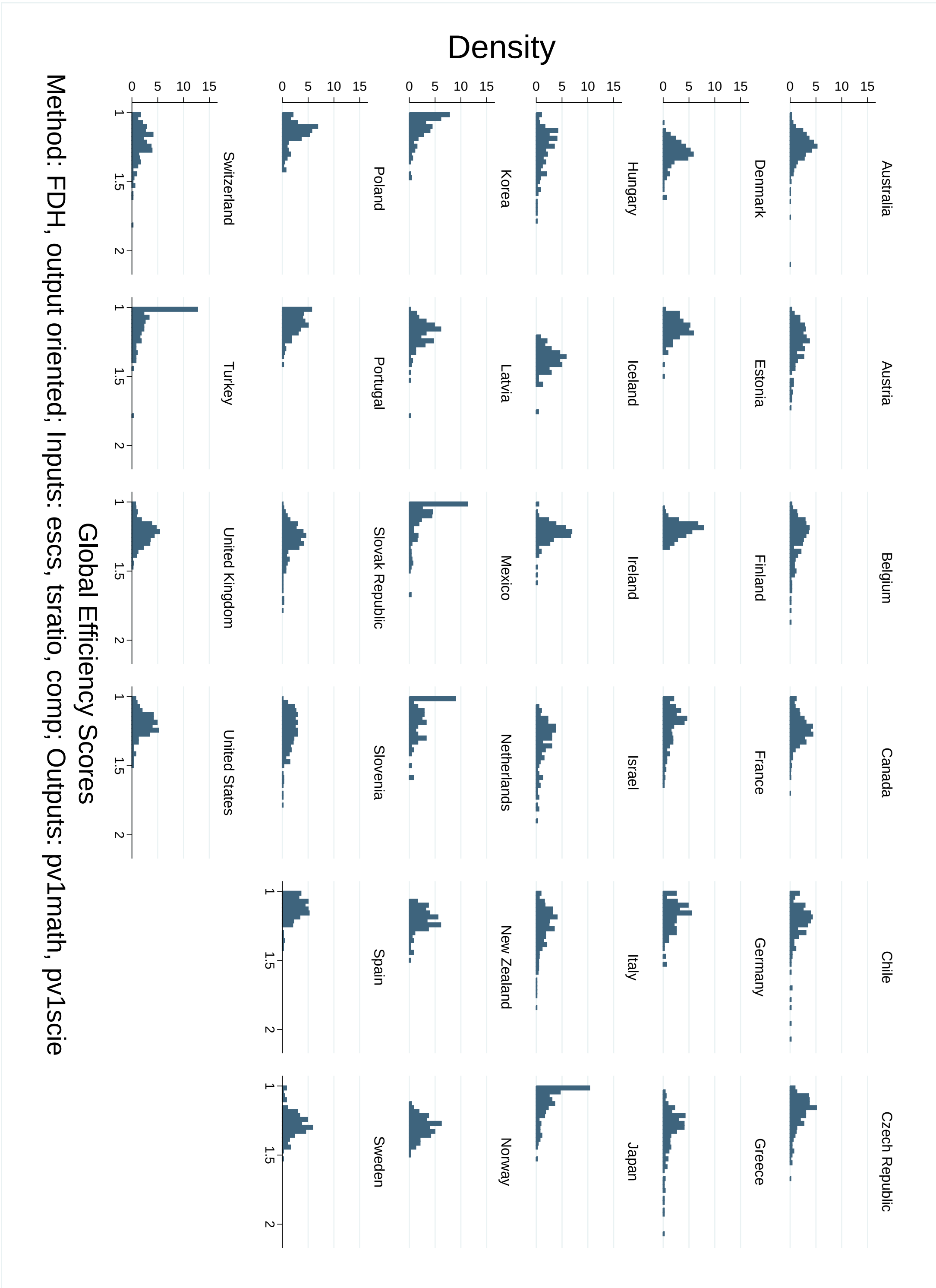


Figure 6.4: Distribution of global efficiency scores by country (after accounting for ESCS)

While in Australia, school efficiency resembles a normal-shaped distribution, in countries like Slovenia or Austria it is more uniform. In Finland, schools are more concentrated around the mean, as these have relatively more similar characteristics. In fact, the Finnish system does not seem to trade-off quality school education for more equitable results across the students. Such feature is especially noticeable when analysing the distribution of country-level efficiency scores, 66 schools (43.4 percent of Finish schools) in the sample define the national PPF. Although the least efficient school in Finland's sample could have still increased the student outcomes by 13.7 percent for the same level of resources and student characteristics, the average inefficient school could have increased student outcomes by 4.7 percent if operating at the national PPF. The Japanese system, on the other hand, while having several schools in the international efficiency frontier has a more scattered distribution of inefficient schools when only considering the school effect (λ^{ScE}). In this case, the most inefficient school in Japan could have still increased its student outcomes by 22.6 percent if operating at its national efficiency frontier.

In fact, focussing our attention in national PPFs, schools could have still increased their outputs, on average, by 6.4% holding inputs constant (9.4% only taking inefficient schools into account)[6]. Therefore, the inclusion of average ESCS resulted in larger gains in efficiency relative to the national PPFs than the international one.

According to the new specification of the model, country effects are larger than the efficiency scores computed through each country-level frontier (Annex D: Summary of Efficiency Scores, Table 12.2). Such result contrasts with the model without ESCS as an input. This implies that when the average socio-economic characteristics of students are taken into account, differences in school inefficiencies across countries are mostly driven by specific characteristics of the organisation of each education system rather than the organisation of schools.

Figure 6.5 depicts the distributions of country effects (λ^{CE}) and school efficiency scores (λ^{ScE}). In particular, 7.9% of the schools in the sample have $\lambda^{CE} = 1$, meaning that these have the same efficiency scores, irrespective if they are measured in relation to the national PPF or to the international one.

We now extend the analysis to the characteristics of those schools defining the international and national efficiency frontiers. Table 6.1 reports some summary statistics

[6] Table 12.2 in Annex D: Summary of Efficiency Scores presents the summary statistics of efficiency scores for all countries when taking ESCS into account.

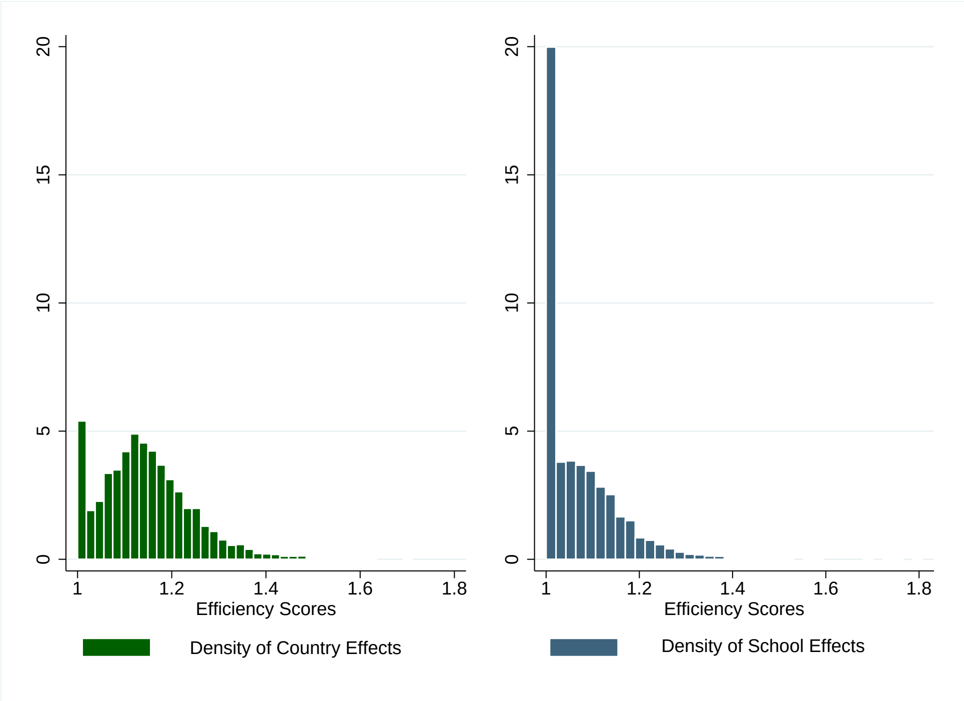


Figure 6.5: Distributions of country and school effects (after accounting for ESCS)

characterising the set of globally and nationally efficient schools. According to the model specification, efficient schools have student outcomes above average in both subjects, but can still vary from as low as 298 to as high as 678. The average student-teacher ratio of globally efficient schools is 17.5 student. However, when the national PPFs are taken as benchmark the number drops to about 12 students per teacher. Regarding the characteristics of the students' population (ESCS), the mean measures are below the average of the full sample (6.17), irrespective of whether the schools are globally or just nationally efficient.

Table 6.1 also reports additional schools' characteristics. Private schools are 1.09 times more likely to be found within the cohort of globally efficient schools than in the full sample (which is not the case for the set of nationally efficient schools). On the other hand, rural schools, defined as those located in communities with less than 15 000 inhabitants, as well as schools with small classes are less likely to have a score of $\lambda^{GE} = 1$. Section 6.2 provides a deeper analysis of the factors bringing inefficient schools closer to the frontier.

6.1.3. *Accounting for super-efficiency*

In order to circumvent the deterministic nature of the FDH estimator, the analysis was further controlled for the existence of super-efficient units, through partial frontier analysis

Efficient schools (international frontier; n = 246)			
Outputs / Inputs	*Mean*	*Min*	*Max*
Average Maths scores	515.36	298.29	672.67
Average Science Scores	520.20	302.01	677.53
Student-teachers ratio (1/tsratio)	17.48	4.62	100
Computers per student	0.38	0.003	3.50
ESCS	5.20	1.00	8.09
Other characteristics	*%*	*Likelihood*	
Private	17.07	1.09	
Rural	21.81	0.68	
Average class size <= 15 students	2.85	0.52	
Efficient schools (national frontiers; n = 2528)			
Outputs / Inputs	*Mean*	*Min*	*Max*
Average Maths scores	505.21	292.99	672.67
Average Science Scores	508.54	257.03	677.53
Student-teachers ratio (1/tsratio)	12.05	1	100
Computers per student	0.66	0.003	8.33
ESCS	6.01	1.00	8.60
Other characteristics	*%*	*Likelihood*	
Private	13.92	0.89	
Rural	29.83	0.93	
Average class size <= 15 students	4.27	0.78	

Note: The Likelihood indicator refers to the ratio between the percentage of efficient schools with a given feature and the percentage of all schools in the sample with that same feature.

Table 6.1. Summary characteristics of efficient schools (FDH)

methods[7]. The outputs considered were the average Math and Science scores in PISA. Inputs were, again, the teacher-student ratio, the computer-student ratio and the average ESCS of the students in each school.

For the order-m computation, we set the number of randomly picked schools in $m = 85$ as it corresponds to the size of the smallest national sample (i.e., Iceland), according to the procedure in Cordero et al. (2017). The number of replacement draws was set at $B = 200$. This implies that 85 schools with at least as much output as the evaluated school were randomly drawn with replacement 200 times and assigned a pseudo-efficiency score for each of the draws. The order-m efficiency score was then computed as the simple average of all the pseudo-efficiency scores for each school (see section 4.3 for a description of the method). From the application of the method, 4.7% of the schools were classified as super-efficient when benchmarked to international partial frontiers ($\lambda^{GE} < 1$), while only 1.2% are efficient ($\lambda^{GE} = 1$).

Figure 6.6 presents the average order-m efficiency scores, excluding all super-efficient schools, i.e., those which the method identified as outliers. The distribution across education systems is somewhat different than the one in Figure 6.2[8]. According to this

[7]See section 4.3 for a description of these methods.
[8]Table 12.3 in Annex D: Summary of Efficiency Scores presents the complete decomposition of order-m

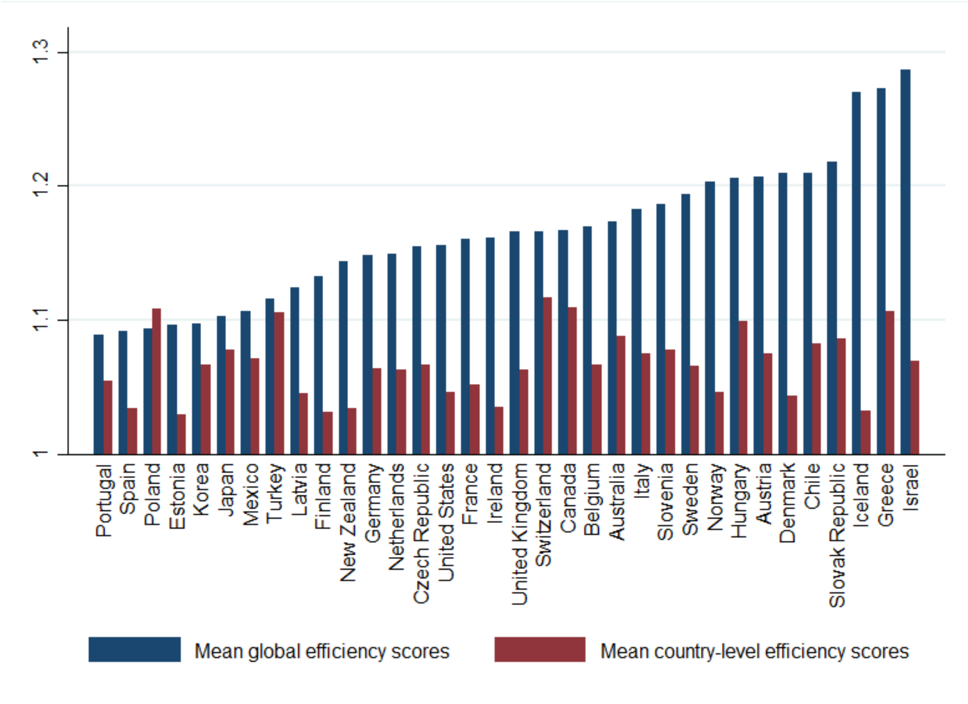


Figure 6.6: Mean global and country-level efficiency scores by country (Order-m; excluding super-efficient schools)

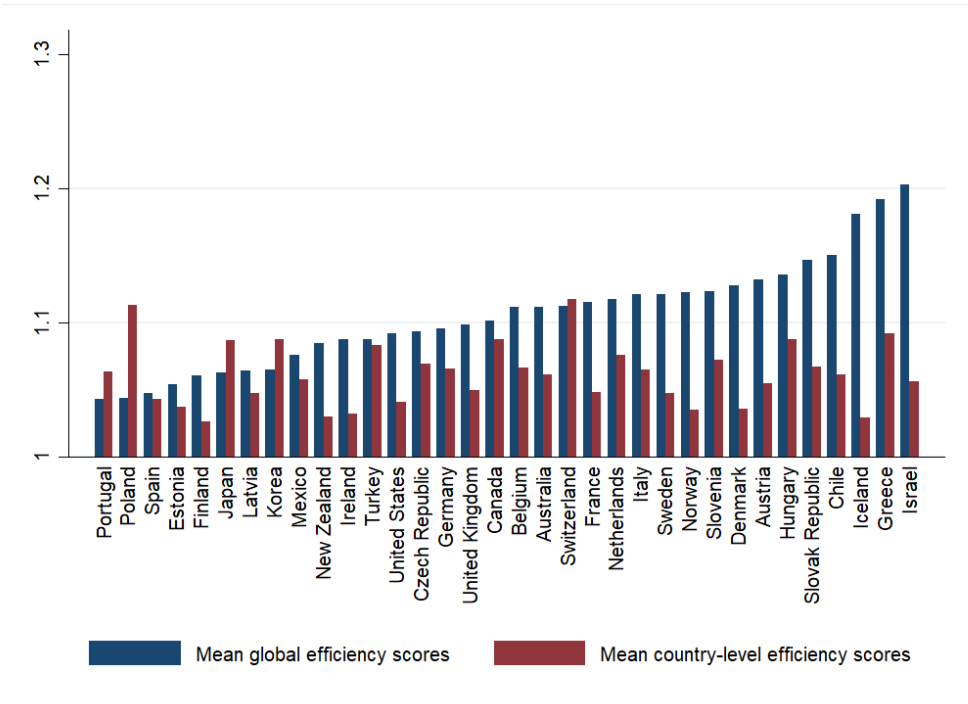


Figure 6.7: Mean global and country-level efficiency scores by country (Order-α; excluding super-efficient schools)

model, Portugal, Spain and Poland top the ranking of countries with the highest mean global efficiency. Poland, in particular, simultaneously has one of the lowest mean global efficiency scores, and one of the highest mean country-level efficiency scores (λ^{ScE}), with the latter surpassing the former. Countries like Iceland, Finland or Ireland, on the other hand, have schools closer to their own local efficiency frontiers.

Two possible interpretations can be provided for the average global efficiency score. The first follows the type of interpretation that has been provided. On average, inefficient schools across the international sample could have increased student outcomes by 14.1%, holding inputs constant. Nevertheless, this is only true to the extent in which the international PPF can be conceptualised as a stochastic frontier – in fact, the order-m technique does not draw a unique frontier, but rather several partial ones. A more accurate interpretation would be that the average inefficient school produces 85.9% of the estimated potential output of a repeatedly and randomly drawn set of 85 schools in the sample (see section 4.3). When only taking into account local frontiers, the number increases to 90.9% of the estimated potential output of a set of randomly selected schools within each country.

Efficiency scores were also alternatively computed through order-α (see section 4.3 for an explanation of the method). In fact, Aragon et al. (2005) and Daouia and Simar (2007) argue that the technique is more robust to outlier observations than order-m. However, to the best of our knowledge, it has not been applied to the study of international comparisons of school efficiency (see section 3), this being the first one using such method. The parameter α was set at 0.95, meaning that a school is classified as efficient if it is dominated by other schools using no more input with a probability of only 0.05. All the schools not enveloped by the efficiency frontier are thus termed as super-efficient, having a score smaller than 1.

Figure 6.7 depicts the mean efficiency scores by country, when applying order-α for the exclusion of super-efficient schools[9]. The ranking of countries remains relatively identical, with Portugal, Spain and Poland still topping the league table and Israel, Greece and Iceland on the bottom[10]. The average inefficient school produces 90.2% of the esti-

efficiency scores, including super-efficient schools.

[9]Table 12.4 in Annex D: Summary of Efficiency Scores reports the decomposition of the efficiency scores by country, including super-efficient schools.

[10]Spearman correlations between the efficiency scores computed through order-m and order-α reach 0.97, in the case of international frontiers, and 0.92 in the case of national frontiers. See Annex E: Correlations Tables for the entire set of correlations across models.

Efficient schools (international frontier; n = 169)			
Outputs / Inputs	*Mean*	*Min*	*Max*
Average Maths scores	468.89	298.29	591.10
Average Science Scores	472.51	302.01	604.60
Student-teachers ratio (1/tsratio)	18.23	5.37	100.00
Computers per student	0.41	0.003	3.50
ESCS	4.86	1	7.91
Other characteristics	*%*	*Likelihood*	
Private	11.24	0.72	
Rural	31.55	0.99	
Average class size <= 15 students	4.14	0.76	
Efficient schools (national frontiers; n = 2329)			
Outputs / Inputs	*Mean*	*Min*	*Max*
Average Maths scores	497.64	292.99	672.67
Average Science Scores	501.12	257.03	677.53
Student-teachers ratio (1/tsratio)	12.29	1.07	100.00
Computers per student	0.64	0.003	8.33
ESCS	5.95	1	8.38
Other characteristics	*%*	*Likelihood*	
Private	13.87	0.89	
Rural	29.33	0.92	
Average class size <= 15 students	4.51	0.83	

Note: The Likelihood indicator refers to the ratio between the percentage of efficient schools with a given feature and the percentage of all schools in the sample with that same feature.

Table 6.2. Summary statistics of efficient schools (Order-α;α=0.95)

mated maximum attainable output of those schools that are dominated by others only with a probability of 0.05. For the case of national PPFs, the average inefficient schools produce 92.1% of the schools efficient at the 95% level, despite the heterogeneity across countries. Interestingly, in contrast with FDH and order-m estimators, average school effects (1.046) loom larger than average country effects (1.031), under order-α (Table 12.4, Annex D: Summary of Efficiency Scores). However, when only non-super-efficient schools are considered (either at the country or international level; n=4570) country effects (1.06) overtake school ones (1.05).

Excluding super-efficient outliers, we can also compare the characteristics of efficient schools using partial frontier methods with those in Table 6.1. Table 6.2 reports their main features. Noticeably, student outcomes are comparatively smaller than for the case of FDH, as schools with relatively large student achievement were trimmed off as outliers. The average socio-economic background of efficient schools according to the method is also smaller than in the latter case. Restricting to national PPFs, average scores are below the ones computed through FDH, but more similar than for the case of the international frontier. Also, student-teacher ratios of the average efficient school in countries remains around 12. Finally, Table 6.2 shows that private and rural schools, as well as those with small classes are less likely to be part of either the international or national efficiency frontiers.

The combination of results from full and partial frontier models allowed us to assess similarities and differences in the interpretation of the international distribution of school efficiency. As both order-α and order-m models are derived from a statistical reformulation of the FDH estimator, it would not be surprising to verify strong correlations between the models. Figure 6.8 shows this positive linear relationship, both at the international and the national levels[11]. The strong correlations between models leads to conclude that, given the empirical data, efficiency scores are relatively robust to changes in the method for the evaluation of school efficiency. However, as verified by the analysis undertaken, the distribution of results is heterogeneous across full and partial frontier models. Especially when analysing national averages, the relative position of some of the countries differs across models. Furthermore, while the characterisation of the set of efficient schools did not change substantially in qualitative terms, quantitative changes were substantial. The relevance of these results should thus be weighted by some considerations. While the theoretical literature stresses the greater robustness of partial frontier models, the application of the estimators to empirical data is not without caveats. The trimming parameters (i.e., m and α), although applied according to best practice, are still set at the discretion of the analyst. Powerful for outlier detection, partial frontier models were here applied in complement to a full frontier approach. Nevertheless, as the number of identified super-efficient units varies with the choice of trimming parameters (Cazals et al., 2002; Aragon et al., 2005), quantitative interpretations are qualified according to these concerns.

6.2. Factors associated with school efficiency

We now turn to explore the environmental factors associated with variations in school efficiency. The dependent variable of the main regressions is the estimated FDH global efficiency score of each school ($\hat{\lambda}^{GE}_{FDH}$), accounting for the average socio-economic background of students. Choosing FDH as the preferred model for the estimation of the dependent variable allows us to retain a greater number of observations[12]. The key covariates considered for the analysis include school and student characteristics, as well as

[11] See Annex E: Correlations Tables for the entire set of correlations across models.

[12] Global efficiency scores estimated through order-m and order-α were also considered as the dependent variable, to perform robustness checks to the preferred econometric specification.

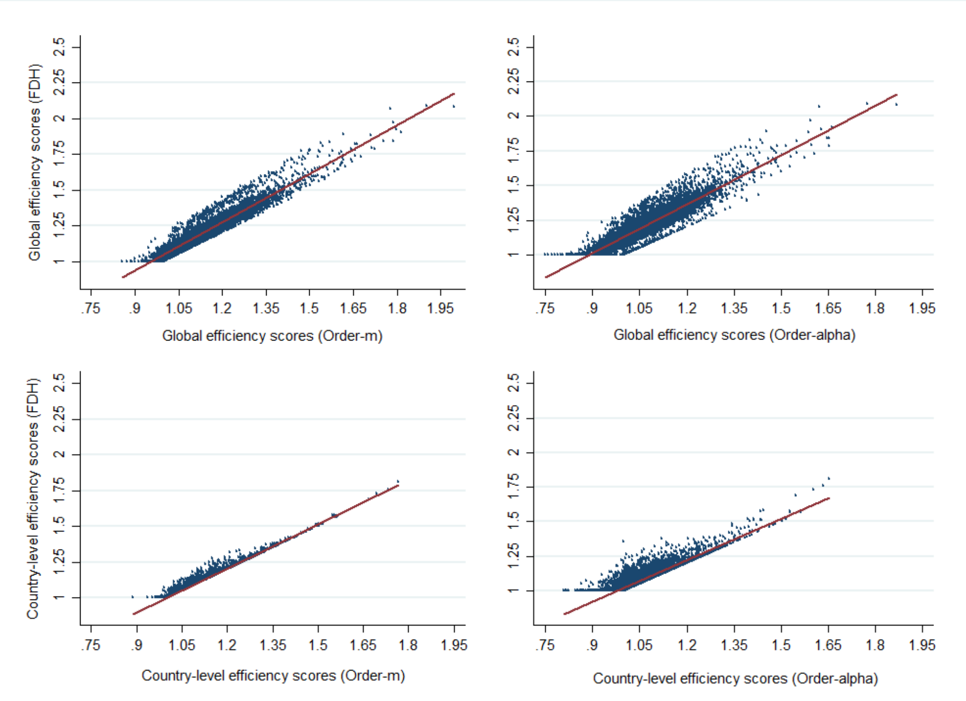


Figure 6.8: Relationship of FDH with order-m and order-α efficiency scores

identified practices at the school-level. In particular, we look into the potential effects of the number of students, school ownership, school location and within-school inequality of student outcomes as the set of school characterising variables. Student characteristics focus on the within-school inequality of socio-economic background and the gender composition of the enrolled body of students. School-level practices include the autonomy of the principal to decide on budget allocations within the school, publication of students' grades, as well as the type of regulations for grouping students in classes. We also used composite indices of school leadership and shortage of educational staff as school-level controls, where these were significant. However, due to it being based on self-reported assessments, we will not interpret its effects[13].

Given the nature of the dependent variable, we run a left truncated regression model, with maximum likelihood estimation and bootstrapped standard errors (see section 4.5)[14]. Following Simar and Wilson (2007), the estimation excludes efficient schools (3.3% of the sample), i.e., those where $\hat{\lambda}^{GE}_{FDH} = 1$. Therefore, the interpretation pertains to the factors

[13]Please refer to section 5.2 and Annex C: Descriptive Statistics for summary descriptions of each of these variables.

[14]The practical estimation was only possible due to the routine developed by H. Tauchmann (2017) SIMARWILSON: Stata module to perform Simar & Wilson efficiency analysis (version 2.2), which can be found in http://fmwww.bc.edu/repec/bocode/s/simarwilson.ado. All regressions were ran using Stata software, version 13.1.

associated with inefficient schools being closer or further from the international PPF. The environmental factors were sequentially added to the regression in blocks. Variables not significant at a 10% level of significance were then sequentially removed from the regression. For most specifications, each regression was repeated 500 times for bootstrapped standard errors. For the preferred specifications, the number of repetitions reached up to 2000.

6.2.1. *Schools' characteristics*

Table 6.3 reports the results for the study of the impact of school characteristics on explaining differences in relative inefficiency across schools. Column (A) of Table 6.3 presents an initial regression, also including potentially significant interactions across given variables. According to the specification, the number of students at school has a small (-9.03e-05) yet strongly significant effect in explaining the differences in global efficiency scores across inefficient schools. In fact, an increase of one standard deviation in student enrolment would imply an impact in the average school's efficiency score of -0.048[15]. As lower scores imply greater efficiency in output-oriented methods, the results suggest that larger schools are also more efficient in providing quality education, controlling for other factors. The marginal results help to support the claim that larger schools are able to garner economies of scale. However, such conclusion contributes to a literature where results are mostly mixed (section 3.1). In particular, the effect of enrolment holds in a context of large dispersion. In our sub-sample of inefficient schools the number of students varies from as low as 13 to as high as 7000, with a mean of about 695 students per school (below the full sample mean of 709). The qualitative interpretation also holds when including country dummies in the regression (Table 6.3, Column (B)), with only a slight quantitative change.

The specifications initially presented were also tested for the effects of school ownership in the level of efficiency. According to the standard view on the operation of private educational institutions, these are generally more pressured to operate within an efficiency-driven framework, as they develop their activity in a more competitive environment than their public counterparts. However, the presented specifications suggest that being a private school (15.6 percent of the sample) bears no significant impact on efficiency. Also, no significant impacts were generally found when this variable was in-

[15] As the standard deviation of total student enrolment among inefficient schools is of 531.3.

teracted with location dummies or continuous variables, such as student enrolment or outcome inequality indicators. However, as the coefficients in Column (B) of Table 6.3 show, when controlling for independent variations across countries and all other factors, private inefficient schools located in rural areas, villages or small towns are closer to the international efficiency frontier, even if for only a 10% significance level.

Dependent variable: Global efficiency scores $\left(\hat{\lambda}^{GE}_{FDH}\right)$		
Variables	(A)	(B)
School characteristics		
Total enrolment in school	-9.03e-05***	-7.85e-05***
	(4.73e-06)	(4.59e-06)
If the school is privately run	0.0288	0.0253
	(0.0230)	(0.0204)
Private * Enrolment	7.46e-06	-6.34e-06
	(1.09e-05)	(9.84e-06)
Private * Outcome inequality (St. Dev. Math scores)	-0.000475	-9.70e-05
	(0.000315)	(0.000277)
Rural area, village or small town	-0.0108**	-0.0103***
	(0.00431)	(0.00367)
Private * Rural area, village or small town	0.00315	-0.0187*
	(0.0125)	(0.0110)
Outcome inequality (St. Dev. Math scores)	-0.00290***	-0.00360***
	(0.000194)	(0.000181)
Outcome inequality (S90/S10 in Math scores)	0.400***	0.419***
	(0.0176)	(0.0159)
Other school-level controls	Yes	Yes
Country dummies	No	Yes
Constant	0.892***	0.951***
	(0.0182)	(0.0176)
σ	0.125***	0.110***
	(0.00153)	(0.00123)
Observations	6,515	6,515
Maximum Pseudo Log Likelihood	5126	5816
Wald Chi-square	1097	2520
p(Wald Chi-square)	0.000	0.000
Bootstrap Repetitions	500	500

Standard errors in parentheses.
P-values: *** p<0.01, ** p<0.05, * p<0.1
Country dummies for Australia, Austria, New Zealand and Sweden are omitted from the regression due to multicollinearity.
Other school level controls include indices of educational leadership, shortage of educational staff, as well as interactions between these and dummies for private and rural schools.
σ is the equivalent (in a truncated regression) to the root mean squared error in OLS, being the standard error of the regression.

Table 6.3. Final results: impact of school characteristics on efficiency

Independently of being private, the model specifications in both Columns of Table 6.3 suggest that being a school in a rural area, village or small town is significantly associated with higher scores for schools not operating at the international efficiency frontier. While the independent effect of being located in a rural environment or a small town has been extensively studied in the literature, the results generally show greater efficiency in urban schools (Witte and López-Torres, 2017). However, according to our definition of the education production function, efficiency takes into account the average socio-economic background of students within the schools. As schools in rural areas, villages or towns with less than 15 000 inhabitants (31.9 percent of the sample) have lower average ESCS than schools in cities, these are generally benefited in the evaluation of efficiency. In fact, running the same regression on for the case of global efficiency scores not considering average ESCS (raw efficiency), rural schools are generally less efficient. An hypothesis for this may be that schools located in urban areas traditionally deal with more ethnically and socially diverse student populations, showcasing greater diversity of needs than those in rural area and generating greater challenges for providing adequate education for the fulfilment of those needs. We explore such hypothesis in a section 6.2.2 where measures of inequality of students' household wealth are included.

Finally, Table 6.3 also reports the marginal linear association between outcome inequality measures and global efficiency scores. In particular, we look into two different measures. While the standard deviation of Math scores is associated with schools moving closer towards the international frontier, the inter-decile measure (S90/S10) of Math scores within the schools has an opposite qualitative interpretation. However, the effect of the latter looms significantly larger than the former. Within the typical inefficient school student Maths achievement in PISA for those in the 90th percentile of the achievement distribution is 1.48 times higher than those located in the 10th percentile. Holding all other factors constant, a standard deviation (0.17) increase in this inequality indicator for the average inefficient school leads to an increase of 6% in its global efficiency score – moving further from the international frontier. However, the results also show that schools with higher standard deviations in student achievement are independently associated with higher efficiency. But despite higher inter-decile inequality being a sufficient condition for higher standard deviations in student achievement, the marginal effect of the latter is still relatively smaller. While a standard deviation point increase in the S90/S10 measure of student achievement inequality would have an independent impact of 0.072 on the average inefficient school, the independent effect of a standard deviation point increase in the alternative measure would only be of -0.054.

6.2.2. *Students' characteristics*

Table 6.4 reports the results for when student characteristics and school-level policies are added to the model. In particular, Columns (A) and (B) in Table 6.4 emulate the initial specifications in Table 6.3 but only including results statistically significant at a

5% level. The qualitative interpretations of school characteristics remain unchanged when adding new environmental variables (Columns (A)-(D)). Column (B) employs the same model as Column(A) but with a greater amount of bootstrap repetitions. Column (C) includes additional variables, related to student characteristics. Importantly, when the new block of variables is added, the effect associated with being a private school becomes significant at a 1% significance level. The outcome inequality measure based on the standard deviation of students' achievement in Maths is also replaced by the standard deviation in Science performance, in order to avoid multicollinearity concerns[16].

Results in Column (C) suggest that the proportion of female students enrolled in the school is not significant for explaining differences in efficiency scores across schools [17]. Nonetheless, other characteristics have important impacts. Schools with high proportions of immigrants often struggle to provide adequate curricular offer to students coming from other country of origin or whose parents come from a different country (section 3.1). Our results suggest that while a higher proportion of second generation immigrants is significant in explaining differences in efficiency across schools, such may not be the case for first generation immigrants. The effect remains relatively stable when including other factors in the model specification (Column (D)). It suggests that schools with higher share of students whose parents are immigrants are also less able to provide more efficient allocation of resources, *ceteris paribus*. Such results have two important implications. First, that there is potential to explore differentiated policy options for students from immigrant backgrounds. The effects may suggest that while mechanisms to address the challenges arising from a high concentration of first generation students in schools may be well developed across countries, such is not the case for second generation immigrant students. Nevertheless, investigating that hypothesis is out of the scope of this dissertation. Second, the sign of the marginal effect lends support to the idea that schools with higher concentrations of students from immigrant backgrounds are also hindered in the efficient allocation of school resources, even when controlling for average ESCS.

[16]The correlation between standard deviation in Science and the S90/S10 measure in Maths (0.62) is lower than the first considered standard deviation in Maths and the alternative inequality measure (0.76). Estimating models without one of the outcome inequality indicators in the specification leads to the same qualitative interpretations in the other variables, while only having limited impacts in the coefficients. Nevertheless, we opt to maintain both regressors for sake of interpretation and given their opposite qualitative interpretation. Table 13.6, in Annex E: Correlations Tables, presents the correlations between the independent variables included in the regression.

[17]In particular, these results stand in contrast with the ones in Agasisti and Zoido (2015), that find significant positive effects.

Similarly, inefficient schools with higher proportion of students that have repeated a school year at least once are significantly further from the international school efficiency frontier. Such results are in line with the literature. The use of year repetition as a response to low performance by students has been shown to be a policy to bear little or short-lived academic benefits while substantially increasing individual and social costs – both through delaying entrance into the labour market and increasing the amount of funding required in the system (OECD, 2012). Our results support this view, and strongly so. An increase of 1 percentage point in the proportion of repeaters in the school is expected to move the average inefficient school between 11% to 14% away from the frontier (Table 6.5).

Differences in students' household possessions also influence the ability to deliver education closer to the best practice frontier. Schools where the composite index of household cultural possessions are higher, are also closer to the international PPF, controlling for the other factors. The effect is robust across the specifications, with no changes in its qualitative interpretation and only minor changes in its quantitative values. The result suggests that greater cultural endowment is associated with greater ease in providing quality education, given available resources. This is not surprising, as an extensive dedicated literature has been exploring the relationship between cultural capital and student achievement. In particular, Nordlander (2016), using data from Swedish secondary schools, finds that the cultural practices of students and their households, such as reading or visiting museums mediate part of the relationship between socio-economic background and academic success. Tramonte and Willms (2010), using PISA 2000 data, also point out for the positive effect of cultural capital on reading literacy, sense of belonging to school and occupational aspirations. But the effect does not only affect literacy in non-technical subjects, as some authors have earlier suggested (DiMaggio, 1982; Sullivan, 2002). For instance, Huang and Liang (2016), using TIMSS 2011 data for 32 different countries, find that cultural capital is significant in explaining students' results in Maths and Science. Andersen and Jæger (2015) complement this type of literature by also looking into within-school variability. According to the authors, the positive effect of cultural capital is stronger in schools with a greater number of lower achievers and more variability in grades. We further complement such hypothesis by assessing the independent effect of within-school variability of students' families wealth.

According to the results in Table 6.4, wealth inequality across students within the school, as measured by the standard deviation of the distribution, is associated with greater inefficiency in providing quality education. The indicator is based on a comprehensive measure of household wealth, taking into account household income, but also the possession of durable goods associated with wealth in each country (OECD, 2016a). The effect is robust to the introduction of further variables into the regression, as the qualitative interpretation remains unaltered from Column (C) to Column (D). Wealth inequality are then detrimental to school efficiency, showing no necessary trade-off between efficiency and equity. Also, introducing the effect of the variable does not imply a strong reduction, but rather an increase in the absolute effect of rural schools in efficient education. This indicates that the association between school location and proximity to the international efficiency frontier is not driven by a greater inequality of students' wealth in urban schools. In fact, restricting to the set of inefficient schools, a t-test to the difference between rural and urban schools does not give statistical support for differences in average within-school wealth inequality across school locations. Therefore, schools located in a rural area, small village or town are more efficient in providing quality education to their students, independently of the level of the economic inequality within the school.

Finally, students' characteristics effects were not significant for specific interactions with the dummy for private schools. However, when sequentially introducing further blocks of variables, the independent effect of being a private school becomes strongly statistically significant and robust to changes in the specification[18] (Columns (C)-(D), Table 6.4). These effects add to a literature with mixed results for the relevance of private ownership for the provision of efficient education (Witte and López-Torres, 2017). Nevertheless, in the context of our empirical study the results are not unexpected. Although schools that are private have higher average student achievement (about 18 PISA score points in Maths), they also enrol students from more advantaged socio-economic backgrounds (with a significant difference of 0.48 in the ESCS index). Given that the concept of school efficiency used in our analysis accounts for average ESCS as an input with which schools have to operate, the results have a somewhat expected sign. These are also in line with the results found in Agasisti (2013), in the context of the Italian education system,

[18]Such result may indicate that *private* was intitially reflecting the effect of ommited variables with it correlated. As these were sequentially introduced into the regression, and its partial effects independently pinned down, the variable became significant.

and for which the average socio-economic background of students in the school was also used as an input of a non-parametric efficiency analysis.

6.2.3. *School-level practices*

Practices at the school level (Column (D), Table 6.4) are also significant in explaining global efficiency differences across OECD countries.

According to our results, policies regarding the transparency of student results seem to play a role in the efficient provision of quality education. Schools are less inefficient than the reference group when achievement data of students is posted publicly within the school (45.94% of inefficient schools). These results, however, stand in contrast with those of Agasisti and Zoido (2015), who use PISA 2012 data, while employing a DEA analysis of efficiency (see Table 9.1, Annex A: Literature on School Efficiency). Nonetheless, the economic significance of such effect is one of the smallest among the factors considered, as applying such policy in the typical inefficient school would only lead to it being 1% to 7% closer to the international PPF, for a 95% confidence interval (Table 6.5).

Our results also suggest that having the school leader (or principal) taking direct responsibility over school budget allocations (78.6% of inefficient schools) is associated with less inefficiency (-0.009) holding all other factors constant. The effect here observed is in line with evidence that greater school autonomy in developed countries generally yields positive student outcomes (Hanushek et al., 2013). Nevertheless, school autonomy can assume different forms – from increased teachers' autonomy to set the curriculum to having the school management determining the offered courses. However, none of the other variables included for the purpose of assessing school autonomy proved to be significant at any reasonable level[19]. The investigation of the impact of different autonomy aspects on school efficiency merit careful attention. Reforms aimed at providing greater school autonomy have been argued on the basis of the large potential for municipalities and schools' staff to take into account the specific needs of their students' population. Our results support the validity of this rationale, at least for the case of school leaders' autonomy. However, extrapolation for the settlement of clear policy priorities deserves a deeper understanding of these mechanisms, submitted to a careful analysis of the institutional environment of each education system. The delegation of autonomy to public schools

[19] These included such schoool-level policies as having principals deciding on the courses offered and teachers determining the courses' content or the courses offered. See Annex C: Descriptive Statistics for descriptive statistics of these variables.

Dependent variable: Global efficiency scores $\left(\hat{\lambda}^{GE}_{FDH}\right)$				
Variables	(A)	(B)	(C)	(D)
School characteristics				
Total enrolment in school	-8.05e-05***	-8.05e-05***	-7.14e-05***	-6.58e-05***
	(4.19e-06)	(4.10e-06)	(4.07e-06)	(3.98e-06)
If the school is privately run			0.0156***	0.0126***
			(0.00460)	(0.00475)
Rural areas, village or small town	-0.0131***	-0.0131***	-0.0151***	-0.0162***
	(0.00365)	(0.00351)	(0.00351)	(0.00350)
Outcome inequality (St. Dev. Math scores)	-0.00360***	-0.00360***		
	(0.000180)	(0.000172)		
Outcome inequality (St. Dev. in Science scores)			-0.000903***	-0.000878***
			(0.000120)	(0.000121)
Outcome inequality (S90/S10 of Math scores)	0.416***	0.416***	0.182***	0.178***
	(0.0150)	(0.0148)	(0.0114)	(0.0113)
Student characteristics				
Proportion of enrolled girls at the school			-0.00858	
			(0.00960)	
Proportion of first generation immigrants			0.0106	
			(0.0164)	
Proportion of second generation immigrants			0.0364**	0.0389***
			(0.0142)	(0.0142)
Proportion of repeaters at school			0.165***	0.164***
			(0.0104)	(0.0101)
Composite index of household cultural possessions			-0.0159***	-0.0161***
			(0.00192)	(0.00189)
Wealth inequality (St. Dev. Household wealth index)			0.0697***	0.0707***
			(0.00750)	(0.00729)
School-level policies				
Student achievement data are posted publicly				-0.00976***
				(0.00337)
Principal decides on budget allocations				-0.00891**
				(0.00414)
If average class size is 15 students or less				0.0351***
				(0.00631)
Other school-level controls	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes
Constant	0.960***	0.960***	1.034***	1.046***
	(0.0167)	(0.0170)	(0.0179)	(0.0175)
σ	0.110***	0.110***	0.108***	0.107***
	(0.00126)	(0.00122)	(0.00121)	(0.00122)
Observations	6,517	6,517	6,508	6,528
Maximum Pseudo Log Likelihood	5814	5814	5917	5953
Wald Chi-square	2801	2675	2895	3068
p (Wald Chi-square)	0.000	0.000	0.000	0.000
Bootstrap Repetitions	500	2000	2000	2000

Standard errors in parentheses.
P-values: *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Country dummies for Australia, Austria, New Zealand and Sweden are omitted from the regression due to multicollinearity.
Other school level controls include indices of educational leadership and shortage of educational staff. Not significant in regression specification (D).
σ is the equivalent (in a truncated regression) to the root mean squared error in OLS, being the standard error of the regression.

Table 6.4. Final results: the impact of school characteristics, students characteristics and school-level policies on efficiency scores

should be balanced by the existence of local capacity to fulfil students' learning needs. Qualitative policy analyses among countries from OECD have been stressing this concern (OECD, 2017, pp.19-20).

Regarding the organisation of learning environments, small classes have the greatest negative impact on efficiency among the variables considered. Despite the argued potential gains from having smaller classes associated with greater individual support for students, international evidence has been stating the negligible or insignificant effects of class size reductions on average students' academic achievement (Wößmann, 2016). Nevertheless, the results need to be interpreted with care, as evidence has been pointing for heterogeneity of effects according to both the context of the school and the characteristics of students, such as for students from disadvantaged socio-economic backgrounds and in earlier years of education (OECD, 2017, p. 39). Moreover, no significant effects were associated with the grouping of students in classes according to ability.

Table 6.5 presents an alternative display of the final results in Column (D) of Table 6.4, ordered by sign and magnitude of the effects. In particular, this type of display highlights the impact of each independent on the distance of the average inefficient school to the international frontier. For instance, a 1 standard deviation positive variation in the total number of students enrolled in the average inefficient school leads to it being 13% to 17% percent closer to best practice, for a 95% confidence interval. On the other hand, the average inefficient school that has average class size with 15 students or less is 10% to 20% further from the international efficiency frontier, controlling for all other factors.

	Type of impact	Lower bound	Upper bound
Positive impact		%	%
Total enrolment in school	1 s.d.	13.43	17.04
Composite index of household cultural possessions	1 s.d.	5.36	8.57
Outcome inequality (St. Dev. in Science scores)	1 s.d.	4.77	8.29
Rural areas, village or small town	If = 1	4.07	10.05
Student achievement data are posted publicly	If = 1	1.38	7.13
Principal decides on budget allocations	If = 1	0.35	7.42
Negative impact		%	%
If average class size is 15 students or less	If = 1	-20.69	-9.91
Outcome inequality (S90/S10 of Math scores)	1 s.d.	-14.90	-11.60
Proportion of repeaters at school	1 s.d.	-14.26	-11.19
If the school is privately run	If = 1	-9.55	-1.43
Wealth inequality (St. Dev. Household wealth index)	1 s.d.	-8.12	-5.39
Proportion of second generation immigrants	1 s.d.	-3.43	-0.57

Own calculations based on Column (D) of Table 6.4.
Lower and upper bounds are computed for a 95% confidence interval. It presents how much closer the average inefficient school (with global efficiency score 1.2294) gets to the international efficiency frontier, given a 1 standard deviation impact of the independent variable (in the case of continuous variables) or a change of status in the case of dummies.
Variables are ordered by sign and magnitude of the impact.

Table 6.5. Summary of final results: how close to the international frontier

6.2.4. *Reliability of the results*

In order to test the global significance of the regressions, the Wald Chi-square test was performed. All regressions proved to be globally statistically significant as the null hypothesis of the test (all coefficients being not different from zero) was rejected in all cases.

The statistical power of each model was also analysed in light of the significance and size of each estimated σ. σ is the estimated standard error of the regression, and corresponds to the root mean squared error (RMSE) in a standard OLS. Therefore, the smaller its value, the larger the fitness of the model. The value of the estimated σ in the last model (0.107) compares to the unconstrained standard deviation of global efficiency scores among inefficient schools of 0.125.

We have also computed a pseudo-R^2 for the specification in Column (D) and have a rough estimate of the percentage of variation in the dependent variable that is explained by variations in the included covariates. In order to do that, we have made linear predictions of the results for each school given the model and correlated these with the estimated global efficiency score of each school. We then squared the result to obtain a pseudo-R^2 of 0.359, meaning that further research has large potential to add additional explanatory factors to the final model here presented.

Finally, the economic importance of each environmental variable was also analysed in light of the size of the effects. Despite the statistical significance of the environmental variables included in the regression, some of the variables had more limited economic interpretations. Such results further support the finding in section 6.1.2, that differences in global efficiency scores are mostly driven by country effects, rather than variation in the management of schools. Despite the inclusion of country dummies to control for the independent effect of operating in a given education system, a further hypothesis – out of the scope of this work – could be to extend the empirical research to include institutional factors across countries. We have additionally explored country-specific effects by running separate regressions for each country, with the estimated country-level efficiency scores (λ_{FDH}^{ScE}) as the dependent and the factors included in Column (D) as the covariates, with 250 bootstrap repetitions. The results point towards large heterogeneity across countries and the statistical insignificance of most of these, suggesting that the factors leading inefficient schools closer to the international frontier differ from those leading them closer to country-level PPFs[20]. The most consistent result, though is total student enrolment in school which was significantly and positively associated with efficiency in most of the countries analysed. Nevertheless, we abstain from providing a systematic analysis of the results given its limited reliability associated with the small sample size of some countries.

The further reliability of the results was confirmed by robustness checks, reported in the next section.

6.3. Robustness checks

6.3.1. *First stage efficiency computation*

All final results were submitted to extensive robustness checks. The efficiency scores computed through the FDH estimator were robust to further changes in the specification of the model. In particular, we have tested education production functions with different combinations of outputs. Correlations across the alternative specifications are high. Computing global efficiency scores with Science and Reading scores (rather than Maths) as the outputs yields a Pearson correlation coefficient of 0.92. On the other hand, considering Maths and Reading as the outputs yields a correlation coefficient of 0.93. The results are also robust to changes in the plausible values considered for the outputs (holding Pearson correlation coefficients of about 0.94)[21].

Correlations between the different methods used to compute the efficiency scores were also high. Table 13.3 in Annex E: Correlations Tables presents the values. As also depicted in Figure 6.8 efficiency scores computed through full frontier FDH are strongly correlated with their partial frontiers' counterparts. Furthermore, the correlation between order-m and order-α is also close to unity (0.97).

6.3.2. *Second stage regressions*

For the analysis of the impact of environmental variables on efficiency scores, two types of robustness checks were performed. For each block of introduced variables, the sensitivity of the coefficients was both tested for changes in the independent variables and the dependent.

Changes in the covariates were sequentially introduced. In particular, we have tested for alternative measures of within-school outcome inequality, namely S75/S25, as well as

[20] The full set of results for the individual regressions by country are available at request.

[21] The full correlation table across model specifications can be found in Tables 13.1 and 13.2 in Annex E: Correlations Tables.

for different types of outcomes (namely, achievement in Reading). The models were also tested for different measures of socio-economic inequality (namely the standard deviation of ESCS within school), with no differences to be registered. Changing these measures produced no variation in the qualitative interpretation of the results, and only limited differences in the quantitative marginal effects. There were also no significant effects associated with school leader's autonomy in the allocation of the budget in private schools. A composite index variable for school autonomy was also introduced as a potential explanatory factor but, due to its specific discrete nature[22], the more fine-grained measure of autonomy in budget allocations was preferred.

Additionally, alternative models for the computation of efficiency scores were considered for the measurement of the underlying global school efficiency. Different regressions were ran for models with alternative plausible values of performance and for different combinations of PISA subjects. Despite minimal changes in the marginal effects, the qualitative interpretation remained generally stable across the different models. Only when Reading is introduced as the output of the education production function (in exchange of Mathematics or Science), the type of school ownership (private or public) becomes insignificant to explain differences in efficiency across schools[23].

Finally, the covariates in the final model presented in Table 6.4 (Column (D)) were alternatively regressed on global efficiency scores computed through order-m and order-α. Again, qualitative interpretations remained unchanged for most variables. Nevertheless, *private* changed sign using order-α and became insignificant using order-m, having the less stable effect across models. Similarly, when using efficiency scores computed through order-m, the effect of having a average class size of 15 students or less becomes insignificant. Table 14.1, in Annex F: Alternative Dependent Variables, presents the final results for each of the alternative models.

[22]Given the nature of this particular composite index, its distribution shows a non-continuous variable, varying from 0 (no autonomy) to 1 (full autonomy) in discrete steps. Distributions of all variables can be made available at request.

[23]The robustness checks outputs are available at request.

7. Conclusion

Not every school is efficient in providing quality education to children. The complexity of the educational process, as well as the intangibility of the inputs and outputs usually involved in it often hamper the ability of school management and educational authorities to take efficiency-driven decisions. Furthermore, as most schools are public, not-for-profit, and aim to fulfil social goals, these are often far from operating in a competitive environment.

This dissertation empirically supports this claim. According to our model of the education production function, technical inefficiencies in the provision of quality education to 15 year-old students are high: some schools do substantially better with the same level of human and material resource intensiveness than others. In order to confirm this hypothesis, we have used a flexible model for inferring the educational process directly from the data. Contrary to most studies in the field, we have opted to use a non-convex estimator (FDH) in our analysis, thus ensuring consistency in case of non-convexities in the true production possibilities set. Furthermore, we have controlled for differences in socio-economic and cultural background (ESCS). Through this design, we were able to not only account for the efficient allocation of human and material resources but also the ability of each school to face the adverse contextual circumstances in which it operates, implicitly assuming that schools directly incur in the cost of bringing students from disadvantaged backgrounds to higher achievement. We conclude that imputing average ESCS of the enrolled students directly into the unconditional education production function produces significantly different results. Schools that were evaluated as inefficient in the unconditional model are – in the conditional one – at the international efficiency frontier. Even so, inefficient schools could have increased student outcomes, on average, by 18.2 percent were they operating according to best practice.

Nonetheless, the distribution of the computed efficiency scores shows considerable heterogeneity. In order to control for specific production technologies of the countries analysed – in this case, broad institutional and cultural differences across education systems – we have used a metafrontier approach to separate national-level efficiency frontiers from the enveloping international one. We conclude, in line with others in the literature (e.g. Cordero et al., 2017) that country effects loom larger than school-level effects in covering the distance from inefficient schools to the international efficiency frontier. Moreover, there are substantial differences in the distribution of global efficiency scores, i.e., those measured through the international frontier, and country-level efficiency scores, i.e., those measured through each national efficiency frontier. Analyses of such differences allow to draw further nuances in the investigation of the structure of the efficiency distributions. In Iceland, while no schools are part of the international efficiency frontier, 54% of the sample of schools defines the national efficiency frontier – the highest percentage among the full sample of countries. Greater insights can be taken by exploring the factors that drive education systems to have more schools highly concentrated close to the national PPFs. Unfortunately, the size of the school samples was insufficient in most countries in order to draw reliable conclusions about the effects driving schools closer to national frontiers.

Further controlling for super-efficiency allowed us to exclude the effect of potential outlier behaviour and circumvent the deterministic nature of the estimator. Correlations across methods were quite high and increased the robustness of the assessment. Nevertheless, qualitative and quantitative changes were substantial. The average inefficient school could have still increased student outcomes by 14.1%, holding inputs constant – using order-m. The use of order-α provides a somewhat more conservative measure of 9.1%, also given to the higher share of schools trimmed off due to its classification as super-efficient. These results stress the importance of considering alternative model specifications in the assessment of school efficiency. In light of the heterogeneity in the league table of countries and schools, policy-making based on efficiency benchmarking should be aware of the limitations of the models producing the results. Empirical frontier analyses are usually sensible to the technique used and the number of inputs and outputs considered (Grosskopf et al., 2014; Johnes, 2004). Furthermore, school efficiency should be only interpreted in light of the inputs chosen to be part of the education production function. Nonetheless, the extensive robustness tests to which the models were submitted and their consistent high correlations lend greater reliability to the interpretations. A further extension could be to also develop models with convex non-parametric estimation (e.g., DEA) or through the estimation of stochastic parametric frontiers to further support our findings. Another add-on to our empirical strategy in further research could be to introduce bootstrapping techniques in the first stage of the regression, as proposed in

Simar and Wilson (2007).

The second stage of the analysis also allowed to present some insights for policy-making and school management. The results confirm that the institutional design of school-level policies and practices is important to determine their ability to provide efficient education. Our findings suggest that school consolidation policies have the potential to develop economies of scale and contribute to efficient quality education, as larger schools are closer to the international efficiency frontier. Also, providing greater autonomy in budget allocation to school principals was estimated to move the average inefficient school up to 7% of the distance to the international efficiency frontier. On the other hand, schools with small classes are significantly less efficient.

Importantly our results do not support a clear trade-off between efficiency and equity. While alternative measures of outcome inequality provide substitute effects, high concentrations of second generation immigrants and repeaters are associated with greater *in*efficiency. On the other hand, schools where there is less inequality of wealth across students' households and a higher level of cultural capital are closer to the frontier, controlling for all other environmental factors.

These findings reinforce the potential for policy responses that go beyond pure educational interventions. For instance, concentration of families from immigrant backgrounds and limited cultural capital in the same neighbourhoods – either through pure preferential attachment or otherwise somewhat explicit economic incentives – limits the ability of providing quality education, since schools most usually draw their students from their surroundings. Explicit policies for providing greater school choice or the supported transportation of students from socio-economically disadvantaged neighbourhoods to schools in other communities are possible options to pursue, given the results from our analysis. Nevertheless, it was out of the scope of this dissertation to focus on the details of the potential of such interventions – for which a more fine-grained research design would be required. The nature of our research design was mostly exploratory. The framework of analysis here utilised could be further developed to explore some causal interpretations of specific effects, for instance by developing an instrumental variables design in the second stage of the analysis.

Our results show that the efficient provision of quality education depends on a complex interplay between class- and school-level economies of scale, with greater autonomy in budgetary decisions, less residential segregation, and greater support for students and their households to develop cultural habits.

8. References

Afonso, A. and M. St. Aubyn (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic Modelling 23*(3), 476–491.

Afriat, S. N. (1972). Efficiency Estimation of Production Functions. *International Economic Review 13*(3), 568–598.

Agasisti, T. (2013). The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA2006 data. *Education Economics 21*(5), 520–544.

Agasisti, T. (2014). The efficiency of public spending in education: An empirical comparison of EU countries. *European Journal of Education 49*(4), 543–557.

Agasisti, T. and P. Zoido (2015). The Efficiency of Secondary Schools in an International Perspective: Preliminary Results from PISA 2012. *OECD Education Working Papers, No. 117*.

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977). Formulation and Estimation of Stochastic Production Function Models. *Journal of Econometrics 6*, 21–37.

Andersen, I. G. and M. M. Jæger (2015). Cultural capital in context: Heterogeneous returns to cultural capital across schooling environments. *Social Science Research 50*, 177–188.

Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005). Nonparametric Frontier Estimation: a Conditional Quantile-Based Approach. *Econometric Theory 21*(2), 358–389.

Banker, A. R. D., A. Charnes, and W. W. Cooper (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science 30*(9), 1078–1092.

Banker, R. D. and R. C. Morey (1986a). Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research 34*(4), 513–521.

Banker, R. D. and R. C. Morey (1986b, dec). The Use of Categorical Variables in Data Envelopment Analysis. *Management Science 32*(12), 1613–1627.

Bifulco, R. and S. Bretschneider (2001). Estimating School Efficiency: A Comparison of Methods Using Simulated Data. *Economics of Education Review 20*(5), 417–429.

Bowles, S. and H. Gintis (2002). Schooling in Capitalist America Revisited. *Sociology of Education 75*(1), 1–18.

Cazals, C., J. P. Florens, and L. Simar (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics 106*(1), 1–25.

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

Cherchye, L., K. De Witte, E. Ooghe, and I. Nicaise (2010, apr). Efficiency and equity in private and public education: A nonparametric comparison. *European Journal of Operational Research 202*(2), 563–573.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review 104*(9), 2633–2679.

Cordero, J. M., D. Santín, and R. Simancas (2017). Assessing European primary school performance through a conditional nonparametric model. *Journal of the Operational Research Society 68*(4), 364–376.

Cordero-Ferrera, J. M., F. Pedraja-Chaparro, and J. Salinas-Jiménez (2008). Measuring efficiency in education: an analysis of different approaches for incorporating nondiscretionary inputs. *Applied Economics 40*(10), 1323–1339.

Daouia, A. and L. Simar (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics 140*(2), 375–400.

Daraio, C. and L. Simar (2005). Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach. *Journal of Producitivity Analysis 24*(1), 93–121.

Daraio, C. and L. Simar (2007a). *Advanced Robust and Nonparametric Methods in Efficiency Analysis*. Berlin Heidelberg: Springer.

Daraio, C. and L. Simar (2007b). The Measurement of Efficiency. In *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, Chapter 2, pp. 13–42. Berlin Heidelberg: Springer.

Debreu, G. (1951). The Coefficient of Resource Utilization. *Econometrica 19*(3), 273–292.

Deprins, D., L. Simar, and H. Tulkens (1984). Measuring Labor Efficiency in Post Offices. In M. Marchand, P. Pestieau, and H. Tulkens (Eds.), *The Performance of Public Enterprises: Concepts and Measurement*, pp. 243–267. Amsterdam: Elsevier.

Diewert, W. E. and C. Parkan (1983). Linear Programming Tests of Regularity Conditions for Production Functions. In *Quantitative Studies on Production and Prices*, pp. 131–158. Heidelberg: Physica-Verlag HD.

DiMaggio, P. (1982). Cultural Capital and School Success : The Impact of Status Culture Participation on the Grades of U.S. High School Students. *American Sociological Review 47*(2), 189–201.

Dorfman, R., P. A. Samuelson, and R. M. Solow (1958). *Linear Programming and Economic Analysis*. New York: McGraw-Hill.

Farkas, G. (2003). Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes. *Annual Review of Sociology 29*, 541–562.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General) 120*(3), pp. 253–290.

Farrell, M. J. (1959). The Convexity Assumption in the Theory of Competitive Markets. *The Journal of Political Economy 67*(4), 377–391.

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

Greene, W. H. (1980a). Maximum Likelihood Estimation of Econometric Frontier Functions. *Journal of Econometrics 13*, 27–56.

Greene, W. H. (1980b). On the Estimation of a Flexible Frontier Production Model. *Journal of Econometrics 13*, 101–115.

Grosskopf, S., K. J. Hayes, and L. L. Taylor (2014, jun). Efficiency in education: Research and implications. *Applied Economic Perspectives and Policy 36*(2), 175–210.

Häkkinen, I., T. Kirjavainen, and R. Uusitalo (2003). School resources and student achievement revisited: New evidence from panel data. *Economics of Education Review 22*(3), 329–335.

Hanoch, G. and M. Rothschild (1972). Testing the Assumptions of Production Theory: A Nonparametric Approach. *Journal of Political Economy 80*(2), 256–275.

Hanushek, E. A. (2006). School Resources. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 14, pp. 865–908. Elsevier B. V.

Hanushek, E. A., S. Link, and L. Woessmann (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics 104*, 212–232.

Hanushek, E. A. and J. A. Luque (2003). Efficiency and equity in schools around the world. *Economics of Education Review 22*(5), 481–502.

Huang, H. and G. Liang (2016). Parental cultural capital and student school performance in mathematics and science across nations. *The Journal of Educational Research 3*(3), 286–295.

Jehle, G. a. and P. J. Reny (2001). *Advanced Microeconomic Theory.* Prentice Hall.

Johnes, J. (2004). Efficiency Measurement. In G. Johnes and J. Johnes (Eds.), *International Handbook on the Economics of Education*, pp. 613–741. Edward Elgar Publishing.

Jones, J. T. and R. W. Zimmer (2001, dec). Examining the impact of capital on academic achievement. *Economics of Education Review 20*(6), 577–588.

Kang, B.-G. and K. V. Greene (2002, jan). The Effects of Monitoring and Competition on Public Education Outputs: A Stochastic Frontier Approach. *Public Finance Review 30*(1), 3–26.

Koopmans, T. C. (1951a). *Activity Analysis of Production and Allocation.* New York: John Wiley & Sons, Inc.

Koopmans, T. C. (1951b). Analysis of Production as an Efficient Combination of Activities. In *Activity Analysis of Production and Allocation.* New York: John Wiley & Sons, Inc.

Levin, H. M. (2015). Issues in Educational Privatization. In H. F. Ladd and M. E. Goertz (Eds.), *Handbook of Research in Education Finance and Policy*, Chapter 24. New York: Routledge.

Levin, H. M., D. T. Jamison, and R. Radner (1976). Concepts of Economic Efficiency and Educational Production. In J. T. Froomkin, D. T. Jamison, and R. Radner (Eds.), *Education as an Industry*, pp. 149 – 198. NBER.

Lovell, C. K. (1993). Production frontiers and productive efficiency. In H. O. Fried, C. A. K. Lovell, and S. S. Schmidt (Eds.), *The measurement of productive efficiency: techniques and applications*, pp. 3–67. New York: Oxford University Press.

Mayston, D. J. (2017). Convexity, quality and efficiency in education. *Journal of the Operational Research Society 68*(4), 446–455.

Nordlander, E. (2016). Are Young People's Activities an Important Link between Class Origin and Educational Attainment? *European Sociological Review 32*(2), 238–250.

O'Donnell, C. J., D. S. P. Rao, and G. E. Battese (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics 34*(2), 231–255.

OECD (2012). *Equity and Quality in Education: Supporting Disadvantaged Students and Schools.* Paris: OECD Publishing.

OECD (2013). What is the Impact of the Economic Crisis on Public Education Spending. Technical Report 18, OECD, Paris.

OECD (2016a). *PISA 2015 Results (Volume I): Excellence and Equity in Education.* PISA. Paris: OECD Publishing.

OECD (2016b). PISA 2015 Technical Report.

OECD (2017). *The Funding of School Education: Connecting Resources and Learning.* Paris: OECD Publishing.

Olesen, O. B. and N. C. Petersen (2015). Stochastic Data Envelopment Analysis—A review. *European Journal of Operational Research 251*(1), 2–21.

Portela, M. C. A. S. and E. Thanassoulis (2001). Decomposing school and school-type efficiency. *European Journal of Operational Research 132*(2), 357–373.

Ray, S. C. (2004). *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. Cambridge: Cambridge University Press.

Richmond, J. (1974). Estimating the Efficiency of Production. *International Economic Review 15*(2), 515–521.

Rivkin, S. G. (1995). Black / White Differences in Schooling and Employment. *The Journal of Human Resources 30*(4), 826–852.

Scheerens, J., H. Luyten, and J. Ravens (2011). *Measuring Educational Quality by Means of Indicators*. Springer Netherlands.

Shephard, R. W. (1953). *Cost and Production Functions*. Princeton: Princeton University Press.

Simar, L. and P. W. Wilson (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research 115*(3), 459–471.

Simar, L. and P. W. Wilson (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics 136*(1), 31–64.

Sullivan, A. (2002). Bourdieu and Education: How useful is Bourdieu's Theory for Researchers. *The Netherlands' Joumal of Social Sciences 38*(2), 144–166.

Sutherland, D., F. Gonand, and R. Price (2009). Improving public spending efficiency in primary and secondary education. *OECD Journal: Economic Studies 2009*(1), 1–30.

Tauchmann, H. (2011). Partial Frontier Efficiency Analysis for Stata. *Stata Journal* (25).

Thanassoulis, E. and M. C. A. S. Portela (2002). School outcomes: sharing the responsibility between pupil and school. *Education Economics 10*(2), 183–207.

Tramonte, L. and J. D. Willms (2010). Cultural capital and its effects on education outcomes. *Economics of Education Review 29*(2), 200–213.

Varian, H. R. (1984). The Nonparametric Approach to Production Analysis. *Econometrica 52*(3), 579–597.

Witte, K. D. and M. Kortelainen (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics 45*(17), 2401–2412.

Witte, K. D. and L. López-Torres (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society 68*(4), 339–363.

Wolff, E. N. (2015). Educational expenditures and student performance among OECD countries. *Structural Change and Economic Dynamics 33*, 37–57.

Wolff, E. N., W. J. Baumol, and A. N. Saini (2014). A comparative analysis of education costs and outcomes: The United States vs. other OECD countries. *Economics of Education Review 39*, 1–21.

Wößmann, L. (2008). Efficiency and equity of European education and training systems. *Journal of Economic Perspectives 30*(3), 3–32.

Wößmann, L. (2016). The Importance of School Systems: Evidence from International Differences in Student Achievement. *Journal of Economic Perspectives 30*(3), 3–32.

9. Annex A: Literature on School Efficiency

Table 9.1. Summary of non-parametric international frontier studies of efficiency measurement in school education

Paper	Data	Main Methods	Inputs	Outputs	External factors	Main Results
Agasisti and Zoido (2015)	**PISA** 2012; 30 OECD countries	Two-stage DEA: 1) Efficiency scores calculation through **DEA** (with **bootstrapping**); 2) **Tobit** regression in a set of external factors	**Teacher-student ratio** (1/St_Ratio); Proxy for the quantity of material resources(Computer_n - **number of computers per student**); Measure of socioeconomic background (ESCS)	Average score in **mathematics** (pv1math); Average score in **reading** (pv1reading)	**School's general characteristics** (e.g., general / vocational orientation; ownership; class size, school size); **Student´s characteristics** (e.g. proportion of immigrants and repeaters); **School´s practice** (e.g., inter-relations, responsibility of principal in budget allocation)	1) On average, schools can raise scores by 27%; 2) Characteristics of the students explain most of the variation (females, migrants, socioeconomic background), but also some school related factors (extracurricular activities, principal's leadership style, budget autonomy)
Cordero et al. (2017)	**PIRLS** 2011 (Main); Data on social indicators and beliefs from the World Bank; Sample of primary schools in 16 European countries	**Order-*m*** (**FDH** assumptions, **bootstrapping** and conditional analysis)	Number of **teachers per 100 students**; **Instruction hours per week**; Number of **computers per 100 students**; Socioeconomic status of students	Student achievement at **reading**	**School-specific** (e.g., early literacy skills before entering school, parental involvement at home and school, location [urban/rural]); **Country-specific** (e.g., GDP per capita, public expenditure per student in primary education as a percentage of GDP (in 2011), perceived hard work, responsibility and perseverance at the country-level)	1) On average, test scores could have increased by 10% if inefficient schools would perform efficiently; 2) Inefficiency is mostly explained by the operating environment in the country (60%) rather than school context (40%); 3) Heterogeneity among different countries is more relevant than among schools in the same country; 4) Hard work, responsibility and perseverance have a favourable influence in efficiency (more attention should be devoted to non-cognitive skills to increase efficiency in education)
Agasisti (2014)	**PISA** 2006 and 2009 (Main); **Education at a Glance** dataset	Two-stage DEA: 1) Efficiency scores calculation through **DEA** (with **bootstrapping**); 2) **OLS** regression in a set of external factors	**Expenditure per student** (measured through PPP$); **Student-teachers ratio**	Student achievement at **math**; Student achievement at **science**	**GDP per capita**; Average **teachers' salaries** (measured in PPP$); Proportion of students with regular access to Internet at school and at home (proxy for **digital literacy**); Proportion of **public spending in education**; **Instructional time** (measured in hours per year)	1) Switzerland, the Netherlands and Finland are allocatively efficient; 2)A 10% saving of resources is possible given efficient operation; 3) GDP per capita is negatively correlated with efficiency (inconsistent with Afonso and St. Aubyn (2006) and Cordero et al. (2017)); 4) There was a convergence process of relative efficiency among European countries in 2006-2009 (with mean efficiency increasing)

Table 9.1. Summary of Non-parametric International Frontier Studies of Efficiency Measurement in School Education(cont.)

Paper	Data	Main Methods	Inputs	Outputs	External factors	Main Results
Giménez et al. (2007)	**TIMSS** 1999; 31 countries	One-stage **DEA** (with further correction for external factors [separation between global and management technical efficiency; separation between short-run and long-run inputs]	Intensity of **teaching resources** (number of teacher hours per student and per year); **Facilities** (index of the adequacy of teaching facilities); **Materials** (index reflecting the availability and appropriateness of teaching materials); **Quality of teaching staff** (average index of self-expressed confidence to teach mathematics and science)	Academic performance in **math**; Academic performance in **science**	Positive **attitudes towards studying** (attitudes towards maths and science and time spent studying at home); **Availability of resources at home** (incl. % students > 25 books, % students with a desk at home and parent´s educational level); **Family income level** (incl. GNP per capita in PPP and % students with computers at home); **Expectations and conception of difficulty** (including time spent studying at home, expectations of pursuing higher education)	1) When environmental variables (efficiency of management) are taken into account the number of efficient countries in education increases; 2) Average increase in academic outcomes could be 10% (6% attributable to environmental factors and 4% to inefficiency of the system); 3) Quality of the teaching staff is not the variable over which the governments should primarily act; 4) Countries like Australia, Canada, the US and New Zealand have the potential to increase academic outcomes, both through better efficiency and a decrease in the inputs used
Afonso and St. Aubyn (2006)	**PISA** 2003; 25 countries	Two-stage DEA: 1) Efficiency scores computation through **DEA** (with bootstrapping); 2) **Tobit** regression in a set of external factors	Teacher per 100 students (2000-02 average); Hours per year in school (2000-02 average)	Country average students´ performance at **maths**, **reading**, **problem solving** and **science**	GDP per capita; Parent´s educational attainment (% population that attained at least upper secondary education, 2000-02 average); Public-to-total expenditure ratio (2000-01 average)[not significant]	1) Finland, Korea and Sweden are the most efficient education systems; 2) On average, countries could have increased results by 11.6% using the same resources; 3) The higher parent´s education and GDP per capita, the lower the inefficiency; 4) When efficiency scores are fully corrected for differences in GDP per capita and parents´ educational attainment across countries there are significant differences (most efficient: Portugal, Korea and Australia)

10. Annex B: Summary Statistics by Country

		Math scores		Science scores	
Country	Nr. Schools	Mean	St. Deviation	Mean	St. Deviation
Japan	197	532.84	61.44	538.13	63.76
Estonia	156	521.64	35.12	536.75	40.42
Finland	152	513.44	27.15	534.75	28.95
Korea	167	522.20	54.96	514.18	49.71
Netherlands	86	518.25	64.88	514.75	73.48
Germany	167	510.21	59.78	516.32	67.49
Canada	584	504.10	40.01	517.76	41.43
Belgium	219	512.52	63.23	505.36	68.67
Switzerland	190	515.79	58.24	501.20	62.84
Poland	154	509.43	36.60	505.91	38.68
New Zealand	133	496.29	44.56	516.03	50.03
United Kingdom	367	496.58	44.49	511.34	48.33
Ireland	148	503.97	33.23	503.93	36.37
Norway	170	504.84	30.96	502.88	32.47
Portugal	161	500.42	42.07	506.17	40.22
Sweden	179	502.14	42.65	502.79	47.36
Australia	607	490.48	48.33	506.87	52.15
Denmark	235	502.72	36.73	490.84	45.25
Czech Republic	304	496.05	61.16	496.42	65.72
France	208	493.56	68.16	496.42	72.16
Spain	176	490.78	32.32	498.09	32.51
Austria	226	491.50	65.96	490.39	67.27
Italy	313	491.30	62.43	486.52	62.88
United States	144	475.10	41.10	501.56	43.60
Slovenia	268	490.57	61.04	486.09	70.42
Latvia	173	484.21	34.81	490.91	38.18
Iceland	85	484.99	30.90	473.08	28.85
Hungary	211	468.42	71.34	466.27	77.40
Slovak Republic	263	473.89	57.71	458.77	58.62
Israel	141	464.36	68.87	465.25	69.06
Greece	191	450.26	61.25	450.12	63.62
Chile	196	431.46	67.97	454.29	66.64
Turkey	138	414.59	59.02	421.72	56.96
Mexico	209	413.20	41.00	421.14	40.39
Total	7318	490.86	58.02	495.07	60.94

Note: Countries are ordered by the average mean scores at Mathematics and Science

Table 10.1. Sample summary statistics of the outputs by country

		Teacher-student ratio		Computer-student ratio		ESCS	
Country	Nr. Schools	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
Australia	607	0.08	0.03	1.59	1.80	6.52	0.74
Austria	226	0.11	0.06	1.12	1.00	6.29	0.77
Belgium	219	0.13	0.07	1.03	0.86	6.49	0.75
Canada	584	0.07	0.05	1.22	0.92	6.90	0.62
Chile	196	0.06	0.03	0.75	0.59	5.80	1.53
Czech Republic	304	0.09	0.06	1.05	0.63	5.88	0.72
Denmark	235	0.09	0.07	1.01	0.77	6.88	0.74
Estonia	156	0.09	0.03	0.82	0.54	6.28	0.58
Finland	152	0.10	0.04	0.83	0.81	6.59	0.45
France	208	0.09	0.04	0.81	0.65	5.94	0.69
Germany	167	0.07	0.03	0.59	0.37	6.40	0.78
Greece	191	0.11	0.05	0.26	0.17	5.98	0.89
Hungary	211	0.14	0.16	0.81	0.86	5.67	1.13
Iceland	85	0.12	0.10	1.87	1.41	7.21	0.46
Ireland	148	0.08	0.04	0.70	0.55	6.40	0.63
Israel	141	0.10	0.05	0.44	0.35	6.37	0.68
Italy	313	0.12	0.06	0.66	0.64	6.08	0.81
Japan	197	0.12	0.12	0.54	0.67	5.88	0.56
Korea	167	0.07	0.03	0.38	0.43	5.87	0.52
Latvia	173	0.11	0.04	0.94	0.45	5.49	0.69
Mexico	209	0.05	0.04	0.34	0.39	4.37	1.25
Netherlands	86	0.06	0.02	0.63	0.35	6.49	0.58
New Zealand	133	0.07	0.02	1.20	0.78	6.43	0.59
Norway	170	0.07	0.02	0.86	0.43	6.96	0.42
Poland	154	0.13	0.05	0.48	0.33	5.56	0.64
Portugal	161	0.11	0.06	0.55	0.62	5.50	1.02
Slovak Republic	263	0.08	0.04	0.98	0.62	5.96	0.84
Slovenia	268	0.18	0.21	0.82	0.85	6.00	0.74
Spain	176	0.10	0.09	0.78	0.57	5.39	1.05
Sweden	179	0.09	0.03	0.94	0.46	6.74	0.55
Switzerland	190	0.10	0.03	0.78	0.89	6.41	0.70
Turkey	138	0.08	0.05	0.20	0.25	3.93	1.08
United Kingdom	367	0.07	0.03	1.08	0.69	6.58	0.64
United States	144	0.07	0.02	1.16	1.02	6.36	0.82
Total	7318	0.09	0.07	0.90	0.91	6.17	1

Table 10.2. Summary statistics of the inputs by country

11. Annex C: Descriptive Statistics

Continuous variables	Description	Obs.	Mean	Std. Dev.	Min	Max
enrol	Total number of students at school	6699	709.55	554.72	13.00	7159.00
edushort	Composite index on the adequacy of educational materials at school	7247	-0.06	0.98	-1.32	3.63
lead	Composite index on the quality of educational leadership at school	7217	0.07	1.02	-6.74	4.43
staffshort	Composite index on the adequacy of educational staff at school	7232	-0.07	1.00	-1.68	3.72
sd__pv1math	Standard deviation of the mathematics scores in the school	7256	71.73	15.16	0.41	187.24
sd__pv1scie	Standard deviation of the science scores in the school	7256	76.26	17.15	2.44	178.17
s90__10__pv1math	S90/10 ratio of the mathematics scores in the school	7318	1.47	0.17	1.00	4.98
s90__10__pv1scie	S90/10 ratio of the science scores in the school	7318	1.51	0.18	1.00	2.56
s75__25__pv1math	S75/25 ratio of the mathematics scores in the school	7318	1.23	0.09	1.00	2.27
s75__25__pv1scie	S75/25 ratio of the science scores in the school	7318	1.24	0.09	1.00	2.14
Dummies			%			
private	If the school is privately owned	7318	15.59	0.36	0	1
rural	If the school is located in a rural location	7116	31.93	0.47	0	1

Table 11.1. Summary statistics of schools' characteristics

Variables	Description	Obs.	Mean	Std. Dev.	Min	Max
prop_fem_schl	Proportion of female students in school	7318	0.49	0.18	0.00	1.00
prop_immig_1	Proportion of first generation immigrants at school	7318	0.06	0.10	0.00	1.00
prop_immig_2	Proportion of second generation immigrants at school	7318	0.07	0.12	0.00	1.00
prop_rept	Proportion of repeaters at school	7318	0.10	0.18	0.00	1.00
cultposs	Compsite index on students' cultural posessions	7318	0.00	1.00	-3.76	5.02
sd_cultposs	Standard deviation of students' cultural posessions	7256	0.91	0.20	0.00	2.68
sd_wealth	Standard deviation of students' wealth	7242	0.77	0.22	0.00	3.04
sd_escs	Standard deviation of ESCS	7256	0.75	0.18	0.01	2.57
sd_pared	School standard deviation of students' parents education in years	7256	2.27	0.83	0.00	7.01

Table 11.2. Summary statistics of students' characteristics

Continuous Variables	Description	Obs.	Mean	Std. Dev.	Min	Max
schaut	Composite index of school autonomy	7315	0.72	0.23	0	1
Dummies			%			
achv_public	Achievement data are posted publicly	7208	45.78	0.50	0	1
ability_out	Students are grouped in different classes according to ability	7173	5.91	0.24	0	1
ability_in	Students are grouped according to ability within classes	7207	4.72	0.21	0	1
budget_alloc	School leader decides budget allocations within the school	7315	78.02	0.41	0	1
curric_teach	Teachers determine courses' content	7315	67.66	0.47	0	1
curric_offer_p	School leader determines which courses are offered	7315	66.47	0.47	0	1
curric_offer_t	Teachers detemine which courses are offered	7315	49.35	0.50	0	1
teachevl_insp	External inspection is used to monitor teachers' practices	7214	39.77	0.49	0	1
small_class	If classes have, on average, less than 15 students	7318	5.45	0.23	0	1

Table 11.3. Summary statistics of school-level policies

12. Annex D: Summary of Efficiency Scores

	Global efficiency		Country effect		School effect	
Country	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
Japan	1.245	0.151	1.003	0.013	1.242	0.152
Netherlands	1.246	0.200	1.029	0.019	1.209	0.181
Estonia	1.248	0.095	1.104	0.055	1.133	0.094
Finland	1.257	0.067	1.164	0.034	1.080	0.056
Korea	1.259	0.162	1.059	0.033	1.188	0.139
Canada	1.279	0.119	1.064	0.039	1.201	0.104
Germany	1.283	0.176	1.078	0.037	1.190	0.156
Portugal	1.302	0.132	1.133	0.043	1.149	0.103
New Zealand	1.305	0.134	1.127	0.066	1.160	0.127
United States	1.307	0.125	1.140	0.054	1.148	0.103
Switzerland	1.307	0.157	1.073	0.061	1.221	0.159
Poland	1.310	0.100	1.108	0.029	1.183	0.094
United Kingdom	1.311	0.123	1.063	0.033	1.234	0.117
Norway	1.319	0.080	1.188	0.031	1.111	0.074
Belgium	1.321	0.170	1.059	0.043	1.250	0.174
Ireland	1.321	0.099	1.180	0.042	1.120	0.088
Sweden	1.322	0.114	1.061	0.068	1.251	0.132
Australia	1.330	0.140	1.095	0.083	1.219	0.146
Spain	1.334	0.097	1.196	0.037	1.117	0.084
Denmark	1.334	0.105	1.179	0.060	1.135	0.107
Austria	1.344	0.182	1.131	0.073	1.193	0.175
Czech Republic	1.350	0.169	1.077	0.065	1.256	0.160
Mexico	1.356	0.227	1.157	0.109	1.167	0.125
France	1.359	0.211	1.105	0.028	1.229	0.186
Italy	1.363	0.186	1.109	0.043	1.231	0.172
Latvia	1.367	0.110	1.219	0.039	1.122	0.095
Slovenia	1.369	0.183	1.067	0.052	1.285	0.183
Iceland	1.386	0.088	1.304	0.053	1.064	0.071
Israel	1.426	0.220	1.186	0.093	1.207	0.198
Slovak Republic	1.431	0.189	1.142	0.102	1.260	0.188
Chile	1.433	0.250	1.122	0.074	1.278	0.209
Hungary	1.441	0.226	1.101	0.056	1.313	0.222
Turkey	1.450	0.268	1.135	0.099	1.274	0.187
Greece	1.479	0.225	1.214	0.078	1.224	0.213
Total	1.337	0.171	1.113	0.081	1.205	0.158

Table 12.1. Decomposition of the efficiency scores by country, using FDH (raw efficiency)

	Overall efficiency		Country effect		School effect	
Country	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
Korea	1.106	0.097	1.051	0.079	1.054	0.071
Turkey	1.114	0.127	1.015	0.039	1.098	0.118
Japan	1.114	0.116	1.056	0.085	1.056	0.071
Mexico	1.115	0.131	1.042	0.076	1.069	0.077
Portugal	1.116	0.084	1.068	0.058	1.046	0.059
Spain	1.127	0.080	1.094	0.060	1.029	0.038
Poland	1.151	0.088	1.046	0.054	1.101	0.077
Estonia	1.156	0.079	1.127	0.061	1.026	0.044
Netherlands	1.164	0.139	1.109	0.105	1.051	0.090
Germany	1.181	0.110	1.120	0.072	1.054	0.067
Latvia	1.191	0.098	1.142	0.062	1.043	0.060
Czech Republic	1.193	0.117	1.129	0.077	1.057	0.073
United States	1.200	0.091	1.151	0.062	1.042	0.053
Finland	1.201	0.056	1.168	0.035	1.029	0.036
New Zealand	1.205	0.083	1.167	0.074	1.033	0.055
France	1.213	0.138	1.160	0.124	1.047	0.068
Switzerland	1.219	0.131	1.104	0.076	1.106	0.109
United Kingdom	1.223	0.085	1.153	0.068	1.061	0.065
Ireland	1.223	0.077	1.183	0.062	1.034	0.043
Chile	1.232	0.156	1.143	0.073	1.077	0.100
Canada	1.233	0.105	1.110	0.058	1.112	0.076
Australia	1.243	0.098	1.135	0.070	1.098	0.084
Slovenia	1.250	0.138	1.167	0.084	1.071	0.096
Italy	1.250	0.141	1.167	0.088	1.071	0.093
Austria	1.259	0.142	1.191	0.098	1.057	0.085
Hungary	1.262	0.148	1.157	0.090	1.093	0.116
Belgium	1.271	0.150	1.195	0.109	1.063	0.086
Sweden	1.275	0.091	1.200	0.091	1.066	0.090
Slovak Republic	1.276	0.123	1.181	0.086	1.082	0.097
Denmark	1.294	0.089	1.244	0.060	1.040	0.055
Norway	1.298	0.074	1.241	0.058	1.047	0.061
Greece	1.325	0.169	1.203	0.096	1.103	0.124
Israel	1.333	0.168	1.256	0.101	1.061	0.094
Iceland	1.377	0.088	1.337	0.064	1.030	0.053
Total	1.222	0.130	1.145	0.096	1.068	0.084

Table 12.2. Decomposition of the efficiency scores by country using FDH (after accounting for ESCS)

	Global efficiency		Country effect		School effect	
Country	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
Japan	1.057	0.090	1.004	0.073	1.054	0.071
Portugal	1.075	0.061	1.029	0.045	1.045	0.058
Korea	1.075	0.086	1.022	0.069	1.053	0.070
Spain	1.081	0.060	1.051	0.042	1.029	0.037
Poland	1.084	0.061	0.987	0.040	1.100	0.076
Estonia	1.086	0.066	1.060	0.054	1.025	0.044
Mexico	1.101	0.105	1.033	0.057	1.065	0.075
Turkey	1.104	0.117	1.008	0.033	1.095	0.114
Netherlands	1.117	0.125	1.064	0.097	1.050	0.090
Latvia	1.119	0.073	1.075	0.046	1.042	0.059
Germany	1.125	0.101	1.069	0.064	1.052	0.067
Finland	1.126	0.051	1.096	0.036	1.028	0.036
Czech Republic	1.128	0.105	1.071	0.068	1.053	0.073
New Zealand	1.139	0.079	1.104	0.075	1.032	0.054
United States	1.145	0.076	1.101	0.051	1.040	0.052
Switzerland	1.145	0.111	1.040	0.074	1.103	0.107
France	1.147	0.124	1.099	0.117	1.045	0.066
United Kingdom	1.156	0.076	1.094	0.065	1.058	0.062
Ireland	1.157	0.057	1.120	0.046	1.033	0.042
Canada	1.159	0.084	1.051	0.046	1.103	0.072
Belgium	1.160	0.113	1.093	0.082	1.062	0.085
Australia	1.164	0.089	1.078	0.056	1.081	0.077
Slovenia	1.167	0.116	1.094	0.077	1.068	0.095
Italy	1.169	0.118	1.096	0.082	1.068	0.090
Austria	1.170	0.123	1.111	0.090	1.054	0.083
Sweden	1.184	0.082	1.119	0.086	1.061	0.085
Hungary	1.190	0.124	1.094	0.086	1.090	0.114
Chile	1.191	0.129	1.112	0.063	1.071	0.096
Denmark	1.199	0.074	1.156	0.054	1.038	0.054
Norway	1.201	0.066	1.150	0.058	1.046	0.059
Slovak Republic	1.205	0.113	1.120	0.081	1.078	0.095
Israel	1.263	0.152	1.192	0.101	1.059	0.093
Greece	1.265	0.147	1.152	0.090	1.100	0.122
Iceland	1.265	0.075	1.229	0.060	1.030	0.053
Total	1.153	0.109	1.085	0.081	1.064	0.081

Table 12.3. Decomposition of the efficiency scores by country using Order-m (m=85; B=200)

	Overall efficiency		Country effect		School effect	
Country	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
Japan	0.983	0.082	0.946	0.056	1.039	0.067
Korea	1.001	0.083	0.958	0.061	1.046	0.067
Portugal	1.003	0.058	0.966	0.048	1.039	0.056
Poland	1.008	0.059	0.930	0.054	1.086	0.075
Spain	1.009	0.055	0.985	0.042	1.024	0.038
Estonia	1.016	0.062	0.995	0.049	1.021	0.044
Netherlands	1.039	0.120	0.993	0.096	1.047	0.090
Latvia	1.042	0.067	1.007	0.038	1.035	0.056
Czech Republic	1.046	0.093	1.010	0.049	1.035	0.078
Finland	1.049	0.046	1.027	0.031	1.021	0.032
Germany	1.050	0.094	1.009	0.055	1.041	0.066
France	1.064	0.110	1.034	0.095	1.030	0.063
New Zealand	1.067	0.075	1.042	0.064	1.024	0.050
Switzerland	1.068	0.106	0.984	0.064	1.086	0.100
United States	1.068	0.070	1.040	0.048	1.027	0.051
Turkey	1.070	0.110	1.001	0.041	1.069	0.102
Mexico	1.072	0.083	1.023	0.052	1.048	0.072
Belgium	1.079	0.101	1.026	0.068	1.053	0.080
United Kingdom	1.081	0.072	1.038	0.053	1.042	0.055
Canada	1.083	0.077	1.008	0.041	1.074	0.067
Ireland	1.083	0.050	1.053	0.037	1.029	0.041
Slovenia	1.085	0.106	1.035	0.058	1.048	0.093
Austria	1.088	0.113	1.063	0.069	1.024	0.083
Australia	1.090	0.082	1.045	0.037	1.043	0.072
Italy	1.090	0.104	1.041	0.062	1.047	0.081
Sweden	1.103	0.077	1.065	0.069	1.038	0.074
Hungary	1.108	0.112	1.037	0.073	1.071	0.109
Norway	1.117	0.059	1.083	0.042	1.032	0.048
Chile	1.118	0.120	1.076	0.068	1.039	0.089
Denmark	1.118	0.067	1.087	0.044	1.029	0.052
Slovak Republic	1.123	0.103	1.070	0.068	1.051	0.089
Israel	1.176	0.144	1.129	0.087	1.041	0.089
Iceland	1.179	0.070	1.147	0.055	1.028	0.053
Greece	1.179	0.132	1.091	0.075	1.082	0.117
Total	1.077	0.099	1.031	0.070	1.046	0.076

Table 12.4. Decomposition of the efficiency scores by country using Order-α (α=0.95)

		International frontier		National frontier	
Country	Nr. Schools	Total	Percent	Total	Percent
Australia	607	2	0.33	111	18.29
Austria	226	0	0.00	109	48.23
Belgium	219	1	0.46	84	38.36
Canada	584	10	1.71	63	10.79
Chile	196	7	3.57	72	36.73
Czech Republic	304	2	0.66	107	35.20
Denmark	235	0	0.00	99	42.13
Estonia	156	1	0.64	84	53.85
Finland	152	0	0.00	66	43.42
France	208	6	2.88	91	43.75
Germany	167	8	4.79	56	33.53
Greece	191	0	0.00	56	29.32
Hungary	211	3	1.42	76	36.02
Iceland	85	0	0.00	46	54.12
Ireland	148	1	0.68	61	41.22
Israel	141	0	0.00	69	48.94
Italy	313	2	0.64	112	35.78
Japan	197	37	18.78	86	43.65
Korea	167	22	13.17	71	42.51
Latvia	173	0	0.00	74	42.77
Mexico	209	54	25.84	66	31.58
Netherlands	86	15	17.44	39	45.35
New Zealand	133	0	0.00	69	51.88
Norway	170	0	0.00	75	44.12
Poland	154	4	2.60	28	18.18
Portugal	161	16	9.94	66	40.99
Slovak Republic	263	0	0.00	87	33.08
Slovenia	268	0	0.00	104	38.81
Spain	176	9	5.11	71	40.34
Sweden	179	2	1.12	65	36.31
Switzerland	190	5	2.63	55	28.95
Turkey	138	35	25.36	42	30.43
United Kingdom	367	2	0.54	115	31.34
United States	144	2	1.39	53	36.81
Total	7318	246	3.36	2258	30.86

Table 12.5. Number of efficient schools at the international and national frontiers (FDH, after accounting for ESCS)

13. Annex E: Correlations Tables

	FDH (pv1math, pv1scie)	*FDH (pv1math, pv1scie)*	*FDH (pv1read, pv1scie)*
FDH (pv1math, pv1scie)	1		
FDH (pv1math, pv1scie)	0.9303	1	
FDH (pv1read, pv1scie)	0.9177	0.9578	1

Notes: Outputs in parentheses; Inputs: *escs*, *tsratio*, *comp*

Table 13.1. Pearson correlations of global efficiency scores across specifications with different subject outputs

	FDH (pv1math, pv1scie)	*FDH (pv2math, pv2scie)*	*FDH (pv3math, pv3scie)*
FDH (pv1math, pv1scie)	1		
FDH (pv2math, pv2scie)	0.9441	1	
FDH (pv3math, pv3scie)	0.9416	0.9416	1

Notes: Outputs in parentheses; Inputs: *escs*, *tsratio*, *comp*

Table 13.2. Pearson correlations of global efficiency scores across specifications with alternative plausible values

Pearson correlations			
	FDH	Order-α	Order-m
FDH	1		
Order-α	0.9008	1	
Order-m	0.9463	0.9759	1
Spearman correlations			
	FDH	Order-α	Order-m
FDH	1		
Order-α	0.8928	1	
Order-m	0.9427	0.9719	1

Notes: Output-oriented scores; m=85; α=0.95; Inputs: *escs*, *tsratio*, *comp*; Outputs: *pv1math*, *pv1scie*.

Table 13.3. Pearson and Spearman correlations of global efficiency scores across models (FDH, Order-m, Order-α)

	pv1math	*pv1scie*	*tsratio*	*comp*	*escs*
pv1math	1				
pv1scie	0.938	1			
tsratio	-0.037	-0.067	1		
comp	-0.007	0.017	0.109	1	
escs	0.658	0.662	-0.055	0.095	1

Table 13.4. Pearson's correlations between inputs and outputs

	pv1math	*pv1scie*	*tsratio*	*comp*	*escs*
pv1math	1				
pv1scie	0.9321	1			
tsratio	-0.041	-0.093	1		
comp	0.003*	0.0287	0.105	1	
escs	0.646	0.662	-0.103	0.147	1

* Correlation not significant at a 5% level.

Table 13.5. Spearman's correlations between inputs and outputs

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1.enrol	1											
2.private	0.03	1										
3.rural	-0.30	-0.13	1									
4.s90_10_pv1scie	-0.03	-0.08	0.09	1								
5.sd_pv1math	0.07	-0.05	0.05	0.62	1							
6.prop_immig_2	0.11	-0.01	-0.17	0.09	0.04	1						
7.prop_rept	-0.05	0.01	0.09	0.08	-0.04	0.06	1					
8.cultposs	0.06	0.12	-0.12	-0.18	0.05	-0.09	-0.30	1				
9.sd_wealth	0.08	-0.02	0.01	0.09	0.04	0.04	0.08	-0.04	1			
10.achv_public	0.17	0.01	-0.04	0.02	0.04	0.03	-0.12	0.05	0.04	1		
11.budget_alloc	0.03	0.02	0.04	0.09	0.09	0.07	-0.13	0.05	0.01	0.05	1	
12.small_class	-0.20	0.00	0.17	0.04	0.01	0.00	0.08	-0.04	0.00	-0.07	0.01	1

Table 13.6. Pearon's correlations between the environmental variables in the final model (globally inefficient schools)

14. Annex F: Alternative Dependent Variables

Dependent variable: Global Efficiency scores $\left(\hat{\lambda}^{GE}_{FDH,\alpha,m}\right)$			
Variables / Dependent variable method	FDH	Order-α	Order-m
School characteristics			
Total enrolment in school	-6.58e-05***	-5.88e-05***	-4.38e-05***
	(3.98e-06)	(5.33e-06)	(3.77e-06)
If the school is privately run	0.0126***	-0.0148***	0.00150
	(0.00475)	(0.00564)	(0.00429)
Rural areas, village or small town	-0.0162***	-0.0116***	-0.0166***
	(0.00350)	(0.00447)	(0.00320)
Outcome inequality (St. Dev. in Science scores)	-0.000878***	-0.00149***	-0.00111***
	(0.000121)	(0.000149)	(0.000110)
Outcome inequality (S90/S10 of Math scores)	0.178***	0.182***	0.183***
	(0.0113)	(0.0122)	(0.0102)
Student characteristics			
Proportion of second generation immigrants	0.0389***	0.0489***	0.0313**
	(0.0142)	(0.0164)	(0.0132)
Proportion of repeaters at school	0.164***	0.180***	0.178***
	(0.0101)	(0.0102)	(0.00931)
Composite index of household cultural possessions	-0.0161***	-0.0164***	-0.0191***
	(0.00189)	(0.00222)	(0.00162)
Wealth inequality (St. Dev. Household wealth index)	0.0707***	0.0787***	0.0711***
	(0.00729)	(0.00824)	(0.00655)
School-level policies			
Student achievement data are posted publicly	-0.00976***	-0.00878**	-0.00924***
	(0.00337)	(0.00427)	(0.00297)
Principal decides on budget allocations	-0.00891**	-0.00891*	-0.00893**
	(0.00414)	(0.00522)	(0.00387)
If average class size is 15 students or less	0.0351***	0.0131*	0.00686
	(0.00631)	(0.00741)	(0.00603)
Other school-level controls	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes
Constant	1.046***	0.905***	0.957***
	(0.0175)	(0.0198)	(0.0161)
σ	0.107***	0.0925***	0.0913***
	(0.00122)	(0.00173)	(0.00114)
Observations	6,528	5,260	6,212
Maximum Pseudo Log Likelihood	5953	7463	7318
Wald Chi-square	3068	1619	3042
p (Wald Chi-square)	0.000	0.000	0.000
Bootstrap Repetitions	2000	500	500

Standard errors in parentheses.
P-values: *** p<0.01, ** p<0.05, * p<0.1
Country dummies for Australia, Austria, New Zealand and Sweden are omitted from the regression due to multicollinearity.

Table 14.1. Final results: comparison of marginal impacts across different dependent variable models