A mutilevel decomposition of school performance using robust nonparametric frontier techniques*

Claudio Thieme† Diego Prior‡ Emili Tortosa-Ausina§

August 2, 2011

Abstract

This article proposes a methodology for evaluating educational performance, from a multilevel perspective. We consider the use of frontier techniques rather than regression equations—the latter of which do not explore variations in students' outcomes within the same school, as this variation is hidden behind an average. Similar to some recent literature contributions, we use partial frontier approaches to mitigate the influence of outliers and the curse of dimensionality, yielding statistically robust results. In contrast to previous studies that use partial frontiers, we consider in our estimation idiosyncratic variables at the school, class, and student levels. Our model is applied to a sample of students in the fourth year of primary school in urban schools in Chile. The results are in line with previous ones that found that less than 30% of the variance in students' educational attainment could be attributed to their schools. Our application also corroborates the assertion that a model that considers only student-level variables would yield high inefficiencies that cannot be attributed to the school management, but rather to inadequate resource-endowment policy. In other words, when one does not consider specific variables concerning the resources allocated to the schools, the performance of those school is undervalued, largely because inefficiencies caused by suboptimal resource endowments or difficulties that arise from the socioeconomic environment are instead attributed to poor school management.

Key words and phrases: efficiency, multilevel analysis, order-m, school effectiveness.

JEL Classifications: C61, H52, I21.

Communications to: Emili Tortosa-Ausina, Departament d'Economia, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló de la Plana, Spain. Tel.: +34 964387168, fax: +34 964728591, e-mail: `tortosa@uji.es`.

*Claudio Thieme acknowledges the financial support of Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT), grant #11085061. Diego Prior acknowledges the financial support of Ministerio de Ciencia e Innovación (ECO2010-18967/ECON). Emili Tortosa-Ausina acknowledges the financial support of Fundació Caixa Castelló-Bancaixa (P1.1B2008-46), Ministerio de Ciencia e Innovación (ECO2008-03813/ECON and ECO2008-05908-C02-01/ECON) and Generalitat Valenciana (PROMETEO/2009/066). The usual disclaimer applies.

†Universidad Diego Portales.

‡Universitat Autònoma de Barcelona and IESEG School of Management.

§Universitat Jaume I and Instituto Valenciano de Investigaciones Económicas (Ivie).

1. Introduction

Over the last 20 years, there has been growing interest among academics and policy-makers alike vis-à-vis school effectiveness research. The central hypothesis of these research initiatives postulates that certain characteristics of a school under analysis can impact its students' educational attainment, and that that impact holds even after controlling for the students' socioeconomic, academic, and demographic characteristics (Goldstein and Woodhouse, 2000; Opdenakker and Van Damme, 2001; Phillips, 1997; Sammons *et al.*, 1997; Bosker and Witziers, 1996).

There has been remarkable methodological progress in this line of research, mostly due to the development of multilevel models (Bryk and Raudenbush, 2002; Goldstein, 1995) that have improved both the definition and measurement of the underlying causes of students' learning processes (Aitkin and Longford, 1986). The general consensus is that students' educational attainment depends on both their personal circumstances and on the idiosyncratic characteristics of their schools and those schools' catchment areas. In order to model these scenarios, the different *levels* are considered hierarchical systems of the students and schools; in these, individuals and groups are stratified into different clusters, using variables defined for each of these levels (Hox, 2002).

This progress in the field has facilitated the resolution of one of the main methodological problems faced by pioneering studies-namely, the inability to decompose the variety of nested effects that explain students' educational achievements. The new methodological proposals enable one to ascertain the share of each student's educational attainment that can be attributed to the various variables measured at different levels (i.e., student, class, school, and district); this constitutes information relevant to the design of specific policy measures at each level—i.e., student, school, and environment-thus improving service delivery in this sector.

The educational attainment of students is usually measured by using common test scores from all schools, whereas the average result of the tests for a given school is assumed to be an indicator of its educational attainment. The variance *among* schools with respect to total variance (i.e., among all students at different schools) is defined as the "gross effect" of the school. In contrast, the variance *within* a school, which cannot be explained by control variables specific to each school (such as, for instance, the level of resources endowed by the government or the average socioeconomic level among the students) is considered the "net effect" of the school. It is expected that a significant proportion of within-school variance can be explained

by factors specific to each school.

The results of the multilevel studies available thus far indicate that the variance among *students* that can be attributed to their individual characteristics is the most important effect. Additionally, the school's effect, once the socioeconomic level of the students has been controlled for, ranges between 10% and 30%; it is higher in mathematics than in either languages or science, and it is also higher in primary education than in secondary education (Cervini, 2009; Murillo, 2010; Blanco, 2010). Part of this literature also indicates that the educational and socioeconomic characteristics of the students explain not only the differences in the educational attainment of students *within* a school, but also *among* schools. In some countries, this would be related to the schools' selection, on educational and socioeconomic bases, of their students (see, for instance, Elacqua *et al.*, 2006).

Most of the research initiatives used to measure school efficiency have been developed in the fields of education and economics. One of the techniques that have been used more intensely therein is regression analysis. However, as indicated by Silva Portela and Thanassoulis (2001), regression equations "do not explore the variation in pupils' outcomes inside the same school as this variation is hidden behind an average". These concerns have also been raised by Goldstein (1997), who provide an example warning about the possibly misleading nature of an aggregate-level analysis. In addition, as indicated by De Witte *et al.* (2010), school data are usually nested (e.g., students within classes, classes within schools, schools within districts, districts within local education authorities, etc.) and, in such a case, the parametric ordinary least squares (OLS) estimates can be biased, in that the presence of intra (or within) unit correlation can lead to underestimations of the standard error of the regression coefficients (De Witte *et al.*, 2010, p.1224). In addition, the variables selection within this approach is usually restricted to only one output.

In contrast, over the last few years some methods have been used that, among other advantages, allow for an extension of the bundle of outputs and inputs. Most of these methods have been developed in the field of efficiency and productivity measurement using frontier techniques; among them, Data Envelopment Analysis (DEA) stands out on the basis of both the number and relevance of applications.[1] DEA is a linear programming technique initially developed by Charnes *et al.* (1978) to measure the productive efficiency of the so-called decision-

[1]Among nonparametric methods, DEA is the most popular technique used to measure efficiency while using frontier techniques, whereas SFA (Stochastic Frontier Analysis) is the most popular within the parametric field. See Fried *et al.* (1993) and Fried *et al.* (2008) for interesting panoramas of both branches of research within frontier efficiency analysis.

making units (DMUs). It has been extensively applied to the assessment of efficiency among many types of DMUs, including banks, retail outlets, municipalities, hospitals, and schools, among many other applications. Some of its most valued virtues are that it neither stipulates a functional form for the cost (or profit, or revenue) functions, nor for the distribution of efficiencies; it therefore closely *envelops* the data. The existing literature is sizeable, but the unfamiliarized reader can become acquainted by reviewing the recent surveys of Emrouznejad *et al.* (2008) or Cook and Seiford (2009), for instance. In addition to allowing for the simultaneous modeling of several inputs and outputs, DEA also has another appealing feature: it allows for comparisons of each unit with *optimal* or *efficient* values, since it is based on estimations of an efficient frontier where *best-practice* DMUs lie. DEA also has a variant-the Free Disposal Hull (FDH) (Deprins *et al.*, 1984)—which differs from DEA in its removal of the convexity assumption. In practical terms, this implies that each DMU is compared only to other existing DMUs, and that it cannot be evaluated against convex combinations of efficient units. Therefore, FDH is even more flexible than DEA, since there are even fewer required assumptions.

In the field of education, several studies have applied these techniques in order to assess different aspects related to school efficiency. Some interesting literature contributions that consider DEA while using school-level data include Bessent *et al.* (1982), Ruggiero *et al.* (1995), Mancebón and Mar Molinero (2000), Bifulco and Bretschneider (2001), Mizala *et al.* (2002), and Ouellette and Vierstraete (2005), among others. Those that use FDH are much fewer in number and include, among others, Oliveira and Santos (2005).

Studies using student-level data in DEA assessments are more recent, starting with Thanssoulis (1999). Silva Portela and Thanassoulis (2001) have made substantial methodological progress by proposing a DEA approach that identifies the sources of student under-attainment: lack of effort on the part of the student, school effectiveness, and the type of funding regime under which the school operates. Their variable-returns-to-scale (VRS) DEA model measures student efficiency while considering a *global* frontier (*student-within-all-schools-efficiency*) and *local* frontiers specific to each school (*student-within-school-efficiency*). The distance to the local frontier corresponds to the student's effect, whereas the distance between the local and global frontiers reflects the school's effect. Followers of this approach include Thanassoulis and Silva Portela (2002), Silva Portela and Camanho (2010), and Mancebón and Muñiz (2007), among others.

In this paper, we extend the methods of Silva Portela and Thanassoulis (2001) in several directions. Whereas Silva Portela and Thanassoulis (2001) and many of their followers use either DEA or FDH, we exploit some alternative new concepts of efficiency, as well as new

nonparametric estimators. Specifically, we base our analysis on the order-m partial frontiers described by Cazals *et al.* (2002), which offer several advantages over previously used efficiency estimation methods. Although compared to parametric methods DEA and FDH have the relevant advantage of not imposing a particular functional form on the relationship between production inputs and outputs, they do have some drawbacks. As indicated not only byCazals *et al.* (2002) but also by Simar (2003), Simar and Wilson (2008), and Wheelock and Wilson (2009), both DEA and FDH are highly sensitive to extreme values and noise in the data, and suffer from the well-known curse of dimensionality (Simar and Wilson, 2008). In contrast, order-m estimators are robust with respect to extreme values and noise, and are $\sqrt{n}$ consistent: they do not suffer from the curse of dimensionality. In the field of education, only De Witte *et al.* (2010) consider the use of order-m estimators; however, they do not consider explicitly multilevel models, as we do. Specifically, we include school-level variables that refer to controllable and noncontrollable inputs for each school.

In the current study, we also combine the application of partial order-m frontiers with the concept of the "metafrontier" (Battese *et al.*, 2004; O'Donnell *et al.*, 2008)—one that is especially helpful when working with observations that are stratified into different levels and can be evaluated using different frontiers. In addition, with respect to the work of Silva Portela and Thanassoulis (2001), we extend the analysis to include contextual variables, not only at the student level but also at the school level.

Moreover, we also provide a relevant application that focuses on the case of Chile, a country that has taken serious initiatives in the area of making improvements to public service delivery, especially in such important areas as the provision of educational choice, incentives, and information. As indicated by Mizala *et al.* (2007), among these initiatives, a critical input is an assessment of school performance, which includes in several cases a ranking of the institutions that is to be used, as required, to inform parents or to allocate rewards (or penalties) in accountability-type schemes; such instruments are also used in other countries, including the United States. In this regard, the partial frontiers that we use will yield a more precise ranking of schools, thus improving the informativeness of either DEA or FDH. In our particular application, the data from the Chilean educational system consist of a sample of 11,319 students studying in the fourth year of primary school, corresponding to 176 elementary-level urban schools.

To present and discuss our proposal, the rest of this paper is organized as follows. Section 2 presents the model and methods, section 3 describes the data, and section 4 discusses the

results. Finally, section 5 provides concluding remarks.

2. An order-m multilevel frontier proposal

2.1. The decomposition of the multilevel frontier model

The immediate antecedent of our proposal is the study by Thanassoulis and Silva Portela (2002). In that work, the authors consider two frontiers: the local frontier, specific to each school oriented to an estimation of student-within-school efficiency, and the global frontier, used to estimate student-within-all-schools efficiency. The distance to the local frontier depends on the student's efficiency (the so-called student's effect, henceforth STE), whereas the distance separating the local and the global frontiers expresses the school efficiency (the so-called school's effect, or SCE). Figure 1 documents the rationale that generates these two effects.

Student (c) under analysis obtains the output level represented by y_c , which has the input level x_c. When comparing the academic performance of this student to the local frontier (i.e., that corresponding to school d where student c is enrolled), it is obvious that student c is inefficient, as on the frontier we find more-efficient students who attend the same school and obtain better results (y') with the same level of inputs (x_c). Accordingly, the student's effect (the *student-within-school efficiency*, in terms of Thanassoulis and Silva Portela (2002)) can be determined as a ratio: the potential output divided by the actual output ($STE = \alpha' = y'/y_c$). The student's effect is higher than unity when the student is inefficient (as in the case presented in Figure 1), and equal to unity otherwise. When compared to the overall frontier (metafrontier or the student-within-all-schools efficiency, in terms of Thanassoulis and Silva Portela (2002)), the efficiency coefficient for the student under analysis is $OE = \alpha'' = y''/y_c$. Having these two reference frontiers, the school's effect (a sort of technology-gap ratio separating the school-specific frontier from the overall frontier) is determined by comparing the overall and local frontiers ($SCE_1 = y''/y' = OE/STE$).

In summary, the proposal of Thanassoulis and Silva Portela (2002) (henceforth, model 1) concludes by defining the following decomposition, in which global efficiency can be decomposed into two effects, namely:

$$\text{Overall efficiency} = \text{Student's effect} \times \text{School's effect} \tag{1}$$

or

$$OE = \alpha'' = \alpha' \times \frac{\alpha''}{\alpha'} = STE_1 \times \ SCE_1 \tag{2}$$

We will refer to expression (1) as the bipartite decomposition of a school's overall efficiency. As mentioned, we partially follow this proposal, as our interest is in taking the student as a unit of analysis. However, in contrast to Thanassoulis and Silva Portela (2002), here our attempt is to develop a decomposition underpinned by multilevel analysis. This means basically that the metafrontier needs to consider not only student data but also additional variables regarding the internal (i.e., resources allocated) and external (i.e., environmental factors) conditions of each school. In what follows, we develop the multilevel frontier as well as the proposed decomposition, using the scenario depicted in Figure 2. In this figure, we start with the aforementioned model 1 and define two additional proposals (models 2 and 3).

Regarding the similarities between figures 1 and 2, it is easy to verify that the student's effect (STE) is exactly the same for all three models. Therefore, to estimate STE (recall that $STE_1 = \alpha' = y'/y_c \geq 1$), we compare the observed output of student c (y_c) and the maximum output achieved by another student with capabilities similar to those of c, who is enrolled in the same school d. However, when quantifying the school's effect, we consider additional variables relating to the resources available to each school, as it may well be that not all schools are endowed with the same level of resources. In doing this, it is worth estimating y_1—say, the maximum output level a student can achieve-while taking into account his or her specific abilities, c, and the resources available at school d where he or she is enrolled. With this new output level, it is possible to define better the school's effect ($SCE_2 = y_1/y') \geq 1$), as it is estimated by comparing two output levels that correspond to different schools that have been endowed with equivalent resource levels but present significant differences in terms of the efficiency with which they manage their allocated resources. This means that the maximum level of output, y_1, is achievable for student c with the resources allocated to the school d.

It is now clear that SCE_2 is only one part of what appears in Figure 1, indicating the error potentially generated when information concerning the resources allocated to each school is not considered. In other words, the proposal by Thanassoulis and Silva Portela (2002), in not having taken into account the resources allocated to each school, is affected by a potential overestimation of the school's effect.

Having this new reference in the frontier, it is relatively straightforward to expand the decomposition of the overall efficiency and introduce the resources endowment effect (hence-

forth, REE)—something that is usually beyond the scope of a school's decision-making, as resource allocation is usually a decision-making matter handled by education authorities. More formally, the description of this effect is as follows:

Resource endowments' effect (REE): This technology gap appears to be significant when students with different performance levels are placed in schools with different resource endowments. When this is the case, a specific efficiency coefficient $[(REE_2 = y''/y_1) \geq 1]$ determines the importance of this effect. Obviously, when $REE_2 = 1$, there is no gap caused by a lack of resources.

In summary, the decomposition corresponding to model 2 is:

$$\text{Overall efficiency} = \text{Student's effect} \times \text{School's effect} \times \text{Resource endowments' effect} \tag{3}$$

or

$$OE = \alpha'' = \alpha' \times \frac{\alpha_1}{\alpha'} \times \frac{\alpha''}{\alpha_1} = STE_1 \times \ SCE_2 \times \ REE_2 \tag{4}$$

where $\alpha_1 = y_1/y_c$.

We will also refer to model 2, or to the decomposition in (3) as the tripartite decomposition of the school's overall efficiency.

Now we develop model 3 by introducing an additional factor, the spillover effect (SPE) that can modify the school's effect. According to Patrinos (1995), some of the differences in the results that students achieve are related to differences in socioeconomic and environmental factors inside the classroom. When this is the case, a spillover effect appears if, for instance, students enrolled in schools have colleagues with superior socioeconomic conditions that predispose them to obtain better results. Accordingly, the spillover effect caused by this capability gap ($SPE_3 = y''/y_2 \geq 1$) indicates the extent to which differences in students' socioeconomic conditions cause differences in their academic results. When these conditions do not have any impact on student performance, $SPE_3 = 1$. A more formal definition of the SPE follows:

Spillover effect (SPE): As Patrinos (1995) point out, in education, there is a potential spillover effect when a student experiences positive externalities on account of the enrollment of other students having, on average, better socioeconomic conditions than him or her that improve their academic capabilities. This means that, in order to reinforce his or her identification with the group, the student will spend extra effort emulating his or

her peers by behaving in accordance with the internal environment. This gap captures the potential improvement the student can realize by taking advantage of the positive externality caused by emulating advantaged peers, if placed in another school.

We will not repeat here the other components of model 3, as they are found in model 2 and their definitions are the same. Following a process similar to that previously described, model 3 can therefore be defined as follows:

$$\begin{aligned}&\text{Overall efficiency} = \\ &\quad = \text{Student's effect} \times \text{School's effect} \times \text{Spillover effect} \times \text{Resources' endowment effect} \qquad (5)\end{aligned}$$

or, more succinctly,

$$OE = \alpha'' = \alpha' \times \frac{\alpha_2}{\alpha'} \times \frac{\alpha_1}{\alpha_2} \times \frac{\alpha''}{\alpha_1} = STE_1 \times \ SCE_2 \times \ SPE_3 \times \ REE_3 \qquad (6)$$

where $\alpha_2 = y_2/y_c$.

Analogous to the decompositions in (1) and (3), we will refer to expression (5) as the quadripartite decomposition of the school's overall efficiency.

2.2. The order-m estimation of the frontier efficiency coefficients

An important decision to be made before one starts to estimate inefficiency levels and benchmarks in the frontier (y', y'', y_1, y_2) relates to the specification of the prevalent technology used in the teaching process. This specification is not trivial, as it has direct implications vis-à-vis the school's efficiency level. So, when we assume a convex technology, the DEA models operate with virtual points, thus establishing linear combinations among real observations. Alternatively, a nonconvex technology defines real observations as a frontier reference. As a consequence, each inefficient student will be related with another, more-efficient student (i.e., his or her peer), without needing to determine nonexistent points through the combination of real observations. This is precisely the thrust of the FDH evaluation process.

The existing literature highlights some important limitations concerning nonparametric frontier estimation methods: the curse of dimensionality, their lack of statistical properties—as they are deterministic in nature—and the potential impact of outliers. This last issue has been treated in Thanassoulis and Silva Portela (2002), following the method proposed by Thanssoulis (1999). The proposal consists of the identification and elimination of the extreme,

super-efficient cases. However, this is controversial, as the simple elimination of super-efficient units could hide important information; assuming that extreme efficiency is not caused by any error, because these observations provide valuable information, their elimination would increase the overall efficiency value by magnifying mediocrity and reducing the potential efficiency gains that could be achieved.

To cope with the aforementioned limitations, some proposals have established the statistical properties of the FDH estimator (Kneip *et al.*, 1998; Simar and Wilson, 2000), as well as those of other nonparametric efficiency indicators. From these studies, it can be deduced that the FDH models experience dimensionality problems due to their slow convergence rates; at the same time, however, they have quite appealing statistical properties, since they are consistent estimators for any monotone boundary (i.e., by imposing only strong disposability). Moreover, when the true technology is convex, the FDH estimator converges to the true estimator, albeit at a slow rate. In contrast, a convex model causes a specification error when the true technology is nonconvex. See Park *et al.* (2000) or the literature review in Simar and Wilson (2000).

Here we assume a nonconvex technology (meaning, real students will be compared to other real but more-efficient students); however, to sort out some of the problems related to the FDH models, the efficiency scores will be determined through the use of an order-m estimation process. We will initially define the FDH evaluation process and, afterwards, the order-m will be introduced.

Let us assume we have information on the input and output vectors ($x_c = (x_{c,1}, x_{c,2}, \ldots, x_{c,i}, \ldots, x_{c,I})$ and $y_c = (y_{c,1}, y_{c,2}, \ldots, y_{c,j}, \ldots, y_{c,J})$, respectively) for each student in the sample $(1, 2, \ldots, C)$. Characterizing the elements of the integer activity vector as $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_C)$ and the efficiency coefficient as α_c^{FDH}, the output-oriented FDH efficiency coefficient comes from the following linear program:

$$
\begin{aligned}
&\max_{\{\alpha_c^{FDH}, \lambda_1, \lambda_2, \ldots, \lambda_C\}} \alpha_c^{FDH}, \\
&\text{s.t.} \\
&\textstyle\sum_{s=1}^{C} \lambda_s x_{s,i} - x_{c,i} \geq 0, && i = 1, \ldots, I \\
&-\textstyle\sum_{s=1}^{C} \lambda_s y_{s,j} + \alpha_c^{FDH} y_{c,j} \geq 0, && j = 1, \ldots, J \\
&\textstyle\sum_{s=1}^{C} \lambda_s = 1, \\
&\lambda_s \in \{0, 1\}, && s = 1, \ldots, S
\end{aligned}
\tag{7}
$$

For each student c found to be FDH-inefficient, program(7) identifies another student in the sample with superior performance (more precisely, the student having a coefficient $\lambda_{s^*} = 1$); it

also estimates the increase in the output required to reach the nonconvex frontier ($\alpha_c^{FDH} > 1$), being $(1 - \alpha_c^{FDH})$ the required proportional increase in the output level, as illustrated in both subsection 2.1 and figures 1 and 2. For students declared FDH-efficient, program (7) offers an activity vector $\lambda_c = 1$ and an efficiency coefficient equal to the unity ($\alpha_c^{FDH} = 1$).

Some of the problems related to FDH estimations-say, the lack of statistical properties, the curse of dimensionality, or the effect of super-efficient units-can be rectified through recent extensions in the nonconvex efficiency framework. For instance, Cazals *et al.* (2002) and Simar (2003) introduce the order-m estimation, as it is an excellent tool for mitigating dimensionality problems, reducing the impact of extreme observations and, additionally, making statistical inference possible while maintaining the nonconvex and nonparametric nature. A brief description of the order-m assessment is provided in the following paragraphs.

Consider a positive fixed integer m. For a given level of input ($x_{c,i}$) and output ($y_{c,j}$), the estimation defines the expected maximum value of m random variables ($y_{1,j}, \ldots, y_{m,j}$), which are drawn from the conditional distribution of the output matrix $\mathbf{Y}$ observing the condition $y_{m,j} > y_{c,j}$.

Formally, the proposed algorithm used to compute the order-m estimator involves the execution of four steps:

1. For a given level of $y_{c,j}$, draw a random sample of size m with replacement among those $y_{m,j}$, such that $y_{m,j} \geq y_{c,j}$.

2. Compute program (7) and estimate $\widetilde{\alpha}_c$.

3. Repeat steps 1 and 2 B times and obtain B efficiency coefficients $\widetilde{\alpha}_c^b (b = 1, 2, \ldots, B)$. The quality of the approximation can be tuned by increasing B. (In most applications, $B = 200$ seems to be a reasonable choice, but we decided to fix $B = 2000$).

4. Compute the empirical mean of B samples as:

$$\alpha_c^m = \frac{\sum_{b=1}^{B} \widetilde{\alpha}_c^b}{B} \tag{8}$$

As m increases, the number of observations considered in the estimation approaches the observed units that meet the condition $y_{m,j} > y_{c,j}$, and the expected order-m estimator in each of the b iterations $\widetilde{\alpha}_c^b$ tends to the FDH efficiency coefficient $\widetilde{\alpha}_c^{FDH}$. So, m is an arbitrary positive integer value, but it is always convenient to observe fluctuations among the $\widetilde{\alpha}_c^b$ coefficients,

depending on the level of m. For acceptable m values, normally α_c^m will present values higher than unity; this indicates that these units are inefficient, as outputs can be increased without modifying the allocated inputs. When $\alpha_c^m < 1$, the unit c can be labeled as being super-efficient, provided the order-m frontier exhibits lower levels of outputs than the unit under analysis.

As mentioned, the order-m estimation is an excellent tool for mitigating problems relating to dimensionality and the presence of extreme observations and outliers. However, this evaluation is of little use if part of the found inefficiency derives from a lack of resources and/or specific environmental situations a school can experience, and we do not consider these variables in the assessment. To adjust the evaluation process to this situation, as previously discussed in models 2 and 3, here we define a multilevel frontier assessment process that can estimate the impact of potential resources and the spillover effects that schools can have that could impact the students' efficiency levels. This multilevel estimation is made possible by adapting what Battese and Rao (2002), Battese *et al.* (2004), and O'Donnell *et al.* (2008) define as metafrontier production function. For the aforementioned model 2, this process involves the execution of the following steps:

(a) Classify students $(1, 2, \ldots, C)$ according to the school in which they are enroled $(1, 2, \ldots, D)$.

(b) Complete steps 1 to 4 to estimate the efficiency coefficients that correspond to each student in the specific school in which he or she is enrolled (α_c^m) (i.e., consider the school frontier point represented by y' in Figure 3, in order to estimate STE). To facilitate the cross-comparison of results, irrespective of the number of students classified in each school, the same m value is assigned in all the estimations. In doing so, dimensionality problems and the potential impact of outliers are neutralized.

(c) After completing the conditional frontiers, add new input variables (i.e., the resources allocated and the students' capabilities, corresponding to each school d), apply again steps 1-4 of the order-m estimation to the complete sample, and estimate the efficiency coefficients with respect to the metafrontier ($\alpha_{c,1}^m$). These new coefficients provide an assessment of a student's efficiency with respect to the overall metafrontier, taking into account only those schools that operate with no more resources and no better environment than the school where the student had been enrolled—precisely what is represented by point y_1 in Figure 3.

(d) Estimate the resource endowment's effect (REE) as the technology gap ratio contained in ($\alpha^{m''}/\alpha_1^{m'}$).

(e) Estimate the school's effect (SCE) as the technology gap ratio that separates the local and the metafrontier through the ratio ($\alpha_{c,1}^{m}/\alpha_c^{m'}$).

With regard to model 3, one follows a similar process to estimate the spillover effect SPE. This requires that one define the additional steps needed to estimate ($\alpha_{c,2}^{m}$) (REE_3) and ($SPE_3 = \alpha_{c,1}^{m}/\alpha_{c,2}^{m}$). For the sake of brevity, we have not reproduced here the specific algorithm for this model.

Figures 3 and 4 illustrate the advantages of the order-m multilevel efficiency assessment. It is worth noting that in the previous literature, the relationship between socioeconomic factors and academic achievement is well-known; it is also known that at the student level, this relationship is not linear. In other words, against the odds, there exist a significant number of students who obtain exceptional marks, in spite of their socioeconomic level. Nonparametric methodologies are very sensitive to this situation, but the elimination of these students from the reference technology is probably not the optimal way to proceed, as they form part of the phenomena under study and their elimination would conceal part of the reality. Accordingly, the question to answer is: How can we establish, as best as we can, the representative frontier, but without distorting the behavior we can expect from the other students?

To illustrate this situation, let us consider Figure 3, which was built through the use of the overall sample. In this figure, we can see that for a socioeconomic level of -0.5 units of standard deviation, in the mathematics assessment, the average level we can expect to correspond to the schools is approximately 220 points. However, at the student level, we can see that there exists an important number of students obtaining maximum marks. Should we eliminate these students from the analysis? At what point should some students be considered outliers?

The order-m method implies real progress, as it does not require the elimination of any unit. This is possible because in the assessment of each student, a random sample of m observations is chosen, each of which produces at least the same level of output with equal or lower input levels. This process is continued, depending on the level fixed to parameter B. As a result, the efficiency assessment is transformed into a statistically robust process where outliers do not appear to have any impact.

Figure 4 serves to exemplify the situation. Two very different performance situations can

be seen, although the students do come from a similar socioeconomic level. Overall, school number 117 was found to be performing better than school number 45. The marks that the students were achieving are similar, and were even better than what could be expected from the total sample. This is corroborated by the STE coefficient ($\alpha_c^{m'}$)—which is, on the whole, better than those expected for the total sample bearing similar characteristics. As a consequence, they exhibit a super-efficient behavior that suggests that the expected output level inside the school is higher than that which corresponds to the total sample.

On the other hand, for school number 45, the expected value for the marks was higher with respect to the metafrontier than in the interior of the local school frontier. For this reason, the internal inefficiency level will be lower inside the school level ($\alpha_c^{m'}$) than with respect to the global sample level ($\alpha_c^{m''}$). As a consequence, this presents an important technology gap that indicates the extent of a school's inefficiency $(\alpha_c^{m''})/(\alpha_c^{m'} \geq 1)$.

As mentioned, when defining the school's effect, the multilevel assessment requires the introduction of additional variables; the inputs allocated to each school, after all, should be considered. By considering these variables, one implies in the course of the assessment that not all the schools are operating with an optimal level of resources—a difference which presents as a gap due to differences in input allocations (the so-called REE). These additional input variables have no impact on either the student's effect (provided all the students in the same school are exposed to the same variable) or on the school's effect. In the same way, starting from the overall efficiency coefficient ($\alpha_c^{m''}$) and introducing sequentially the variables that correspond to the average socioeconomic level and the average level of marks that each school has, the gaps corresponding to the so-called socioeconomic and spillover factors can be estimated.

3. Data description and sources

The data used in the study were obtained from the results of standardized tests undertaken by the national system for evaluating education quality in Chile (i.e., the Sistema de Evaluación de la Calidad de la Educación, or SIMCE), which since the mid-1990s has collected information on student characteristics and academic performance. It is a relevant system, not only because its results are widely disseminated (i.e., the schools' performance levels are listed in major newspapers and by the media), but also because the government has started to use SIMCE scores to allocate resources, as well as to promote accountability and transmit incentives (Mizala *et al.*, 2007). In Chile, there are four types of schools: private nonvoucher

schools, fee-charging private voucher schools, free private voucher schools, and public schools; all of them are subject to the standardized testing system. For a painstaking description of the Chilean school sector see, for instance, Anand *et al.* (2009) or Mizala and Romaguera (2004). Some basic information on the schools in our sample is reported in Table 1. For convenience, fee-charging private voucher schools and free private voucher schools have been merged into one category—namely, privately owned subsidized schools.

Our data correspond to the scores of fourth-grade students of basic education who took these standardized tests in 2008. In order to have a reasonable minimum number of students per school-and also to constrain the sample to schools that carry out standardized organizational procedures-we selected a sample comprising 176 schools within urban areas, each of which has more than 30 students who participated in the standardized tests, and each of which had existed for more than three years previous. Additionally, this group of schools was also subject to a survey, in which a minimum of five teachers in each school responded to questions relating to the quality and quantity of resources at their disposal. In total, the sample consisted of 11,319 students from a variety of schools, socioeconomic levels, and regions across the country. Table 2 provides information on both the students and schools included in the sample.

More specifically, the results corresponding to each student in the language and mathematics tests were obtained from the SIMCE database of student results (SIMCE, "Resultados por alumno" database). The socioeconomic level of each student's family corresponds to a latent variable constructed through the use of confirmatory factor analysis. The variables included were: (i) father's years of schooling, (ii) mother's years of schooling, and (iii) family's monthly income. The variables were obtained using the questionnaire answered by the parents of students participating in the SIMCE. At the school level, the variable "school's average socioeconomic and cultural leve" corresponds to the arithmetic mean of the variable "socioeconomic and cultural level of the student's family." The variable "availability of quality teaching resources in the school" corresponds to a latent variable comprising five questions drawn from the questionnaire that had been answered by the teachers. Additionally, the variable was constructed using confirmatory factor analysis.

When computing the latent variable score for quantifying the latent variables used in this application, we obtained normalized results that contain both negative and positive values. Since nonparametric frontier models cannot handle negative values, the latent variable scores were transformed so that we had only positive values (Pastor, 1996).

Table 2 reports summary statistics on the three selected variables at the student level (i.e., two outputs and one input), and the two selected inputs at the school level. We consider three models, each with different mixes of inputs and outputs. Specifically, model 1 has one input and two outputs, model 2 has two inputs and two outputs, and model 3 has three inputs and two outputs, where x_1, x_2, and x_3 are the inputs and y_1 and y_2 are the outputs.

4. Results

Table 3 reports order-m results for model 1, which considers only student-level data. The first row in the table indicates that, on average (i.e., geometric mean), the overall inefficiency (α'') obtained by maximizing outputs that corresponded to the 11,391 students from 176 schools in the sample was 1.2336; this is higher than the value of 1.1827 that corresponds to the student's effect in equation (2), which is presented in the second row of Table 3, α'. As a result, the inefficiency attributable to the school (α''/α') is 1.0423. Therefore, on average, the contribution of the student's effect to overall inefficiency is much higher than that attributable to the school.

Table 3 provides some additional summary statistics (i.e., median, maximum, minimum, and standard deviation, along with first- and third-quartile values). That table shows that 100% of the schools are globally (i.e., overall) inefficient. However, when considering inefficiency attributable to a school, 29.5% of schools in the sample were found to be efficient, and the inefficiency of the remainder is, on average, 7.51%.

In addition to these summary statistics, in order to complement the information reported on Table 3, we considered some tools that allow for a fuller view of the distributions of α'' (OE), α' (STE), and α''/α' (SCE). Specifically, using kernel methods, we made estimations of the densities corresponding to each indicator; these results offer much more detailed information than do the summary statistics. This information is reported in Figure 5, where the contributions of each component to the overall efficiency are added sequentially. The vertical lines correspond to the average of each effect. Figure 5.a displays the density corresponding to the student's effect, STE, which exhibits a certain amount of bimodality in the vicinity of 1.15. The school's effect, SCE, as indicated in Table 3, offsets the student's effect on average. Figure 5.b illustrates this fact, and we can see visually how the distributions corresponding to SCE and STE differ remarkably. Due to this discrepancy, the emerging picture is of a much flatter distribution that corresponds to the overall efficiency effect, OE, as shown in Figure 5.c; this finding indicates that its components contribute in different ways to the global effect, resulting

in a bimodal distribution.

Results corresponding to the tripartite decomposition of the overall efficiency included in model 2—which differs from model 1 on account of its inclusion of an additional effect at the school level (i.e., resource endowments effect)-are reported in Table 4. The presentation is analogous to that in Table 3 for model 1. In this case, the former school's effect (SCE_1) is decomposed into a net school's effect (SCE_2) and a resource endowments effect (REE_2). The impact of the latter has a positive effect on overall inefficiency. However, *on average*, its magnitude (1.0247) is still much lower than that corresponding to the student's effect, STE (1.1827).

Analogous to the analysis performed above for the bipartite decomposition (model 1), Figure 6 illustrates how the inclusion of this effect influences the relative contributions of each component of overall efficiency, while considering the full distributions of the effects. Figure 6.a reports the same information as Figure 5.a, although it is inserted again in Figure 5 to facilitate comparison. This is also convenient, because due to the new effect, REE, the OY axis scale is different. As shown in Figure 6.c, the impact of the resource endowments is much closer to the school's effect, SCE, than to the student's effect, STE, contributing modestly to overall efficiency. Estimates of densities are useful, because we can visually see how the SCE and REE differ. In the case of the resource endowments effect, the distribution is much tighter, pointing to a very homogeneous effect across schools. Finally, in Figure 6.d, it is apparent that the three effects combined (STE, SCE, and REE) yield a very flat (i.e., dispersed) overall efficiency, with the probability mass spread evenly spread in the $]1, 1.6[$ range.

Results for the quadripartite decomposition of overall efficiency in model 3 are reported in Table 5 (i.e., summary statistics). In this model, an additional noncontrollable input, x_3, is included to control for the peer effect at the school level caused by the socioeconomic characteristics of its students—characteristics that we have collectively labeled as a spillover or contagion/emulation effect, SPE. Compared to the tripartite decomposition (model 2), in this model it is the previous school-effect magnitude (SCE_2) that is split into the spillover effect (SPE_3) and the net (residual) school's effect (SCE_3). With the overall efficiency and the student's effect being constant, the inefficiency attributable to the school's management is only 0.55% ($\alpha_1/\alpha' = 1.0055$)—corresponding to 5.04% of the inefficiency, as indicated in the third row. The inefficiency of the spillover effect SPE_3 (i.e., 1.0189) is higher, on average, than the inefficiency due to the resource endowments .

Analogous to both the bipartite and tripartite decompositions of overall efficiency, Figure

7 displays the sequential densities corresponding to the relative contributions to overall efficiency. The information in Figures 7.a and 7.b, although already reported in Figure 6, has been included here so as to facilitate comparison—and also because the scale of the OY axis changes due to the SPE effect. The relative contribution of the entire distribution of the spillover effect is shown in Figure 7.d; the magnitude of the effect is very similar to that of the resource endowments effect. Indeed, Figures 7.c and 7.d are very difficult to distinguish, because the densities corresponding to and overlap to a large degree. Again, the four effects, when combined, yield a flat (i.e., dispersed) distribution in which the efficiency corresponding to each school is generated from a variety of causes.

In summary, according to these results, one may conclude that, regardless of the decomposition considered, overall inefficiency (α'') is primarily caused by the student's effect (α'), followed by the impact of the spillover effect (α''/α_1) and the impact of the resource endowments (α_1/α_2), and only to a lesser degree by the net school's effect (α_2/α').

When considering the different types of school listed in Table 6, results differ remarkably among them. On average, the overall inefficiency indicator (α'') for public schools is the highest (1.3008), followed by privately owned subsidized schools (1.1699). Privately owned fee-paying schools, meanwhile, are the most efficient (1.1063). However, these remarkable discrepancies are not attributable to school performance (α_2/α'), but to inefficiency at the student level (α'). Indeed, the student inefficiencies (α') are, on average, 1.2123, 1.1535, and 1.1398 for public, privately owned subsidized, and privately owned fee-paying schools, respectively. In addition, when analyzing the average inefficiency of inefficient schools, the (α_2/α') parameter takes values of 1.0594, 1.0362, and 1.0349 for public, privately owned subsidized, and privately owned fee-paying schools, respectively. However, although both types of privately owned schools show similar levels of performance, there exist larger discrepancies among the subsidized ones (42.50% of which are inefficient), whereas for fee-paying schools this percentage is only 14.29%. Regarding the SPE and REE effects, their impact on public schools is virtually generalized, since 95.51% and 97.75% of schools are inefficient for REE and SPE, respectively. The percentages of inefficient schools decrease for both privately owned subsidized (88.75% and 93.75% for REE and SPE effects, respectively) and privately owned fee-paying schools (85.71% and 14.29% for REE and SPE effects, respectively). As expected, peers and socioeconomic level impact to a lesser degree the latter category of schools; however, the effect of suboptimal resource endowments is generalized among Chilean schools, regardless of type. In addition, it is higher than the school's effect and lower than the student's effect.

However, as suggested, the differences referenced in the previous paragraphs are based on descriptive comparisons of averages for the different school types. Testing whether or not these differences for the means are statistically significant is possible, through the use of well-known instruments such as the Wilcoxon test. There have been some advances in the field of nonparametric statistics that enable one to perform statistical tests, in order to compare whether entire *distributions* show significant differences-differences that are not restricted to a few moments within the distributions. These tests were introduced by Li (1996) and by several applications, such as those in Murillo-Melchor *et al.* (2010), Pastor and Tortosa-Ausina (2008), and Balaguer-Coll *et al.* (2010).[2] Results are displayed in Table 7, which reports the results of testing the null hypothesis that the distributions of each of the variables of interest is the same for pairs of types of schools compared. For instance, the T-statistic (which does not correspond to the Student's t), yielded by comparing the OE distributions of public vs. privately owned subsidized schools, is 16.6565; this is significant at the 1% level. The differences between these types of schools are significant at this level for all effects; however, when the privately owned fee-paying schools enter the analysis, this result no longer holds. Rather, this result is jeopardized by the few observations available for this type of school.

Figure 8 provides an illustration of the student's (α') and the school's inefficiency (α''/α'), corresponding to the bipartite decomposition in equation (2). Drawing a vertical line and a horizontal line allows for the creation of a four-quadrant matrix that corresponds to the four groups of schools according to their performance mix, since the means of improving differ across the four types. Each group can be described as follows.

- The first group comprises those schools that are either efficient or super-efficient ($\alpha''/\alpha' \leq 1$), combined with a lower-than-average student efficiency (α'). This group corresponds to the lower-left quadrant of Figure 8. This group consists of 56 schools, representing 31.80% of the schools in the sample. Of these 56 schools, 17.85% are public, 71.42% are privately owned subsidized, and 10.71% are included in the privately owned fee-paying category. Most of the schools in the last of these categories (85.70%) are included in this group, whereas the same is true for only 50.00% of subsidized schools and 11.23% of public schools.

- The second group comprises those schools that are, on average, either efficient or super-efficient ($\alpha''/\alpha' \leq 1$), but show a student's inefficiency that is higher than the average

[2]The technical details of this test can be found in any of these articles, and also in Kumar and Russell (2002).

($\alpha' \geq \overline{\alpha'}$). This group accounts for 18.75% of the sample, and it constitutes 75.75% of public schools and 24.25% of privately owned subsidized schools.

- The third group comprises inefficient schools ($\alpha_1/\alpha' \geq 1$), although they do tend to show have lower-than-average student inefficiency ( ). In this group there are 35 schools of all types, corresponding to 19.90% of the total sample-most of which are privately owned subsidized (51.43%) and public (45.71%) schools.

- The fourth group comprises those schools with the worst indicators. They are inefficient schools and, in addition, their student's inefficiencies are higher than average ($\alpha' \geq \overline{\alpha'}$). It contains 52 schools (representing 29.50% of the sample), most of which are public (73.07%) and privately owned subsidized (26.93%) schools.

5. Conclusions

This article presented a methodological contribution by which one can evaluate educational performance in a multilevel context. Unlike previous studies that consider regression approaches in measuring student and school attainment, we consider frontier techniques which, in their nonparametric form, do not require the a priori specification of the functional form and which allow one to measure the performance of each individual (i.e., a student) in terms of best-practice performance. Likewise, some recent but scarce contributions—such as De Witte *et al.* (2010)—consider order-m techniques so that both the curse of dimensionality and the influence of outliers are largely alleviated, resulting in statistically robust results. In contrast with the proposal of De Witte *et al.* (2010) and those of Silva Portela and Thanassoulis (2001); Thanassoulis and Silva Portela (2002), we consider in the models we propose school-level variables. Both the literature pertaining to multilevel models and the results obtained in the application carried out in this paper for a sample of 176 primary-education schools in Chile suggest that it is convenient and necessary to include these types of variables; they allow for the inclusion of relevant extra information in examinations of the analyzed phenomenon.

Our application allowed for corroboration not allowed with a preliminary model that considers only student-level variables; it shows large global and school inefficiencies that are attributable not to the management of the school, but rather to policy that results in inadequate resource endowments. If school-level variables were not considered, we would undervalue the performance of those schools that either enroll students with unfavorable socioeconomic situations or that have suboptimal resource endowments.

Similarly, our analysis by type of school (i.e., public, privately owned subsidized, or privately owned fee-paying) indicates that the large discrepancies among the different types diminish sharply when school-level variables are included in the analysis; these discrepancies virtually fade away when confining the analysis solely to privately owned schools. However, privately owned schools perform better than public schools. One may expect that the inclusion of other variables that proxy for resources would contribute to a further reduction in this gap.

In comparing public schools and privately owned fee-paying schools-both of which compete for similar students-it becomes clear that the global efficiency estimator for public schools is, on average, 1.3008; this is considerably higher than that of privately owned subsidized schools, whose average is 1.1699. However, this difference is not caused by management differences among each school type (1.0245 vs. 0.9880 for public and privately owned subsidized schools, respectively), but rather to remaining factors, especially those relate to student characteristics. Indeed, the average negative effect of REE is, in public schools, roughly twice that of subsidized schools, and the spillover effect attributable to student characteristics is also considerably higher in public schools than in privately owned subsidized schools.

The results of the current study roughly coincide with those of other studies in the field of education research, most of which point out that fewer than 30% of variance in the results of students' educational achievement are due to the school's effect. The 5.04% that we obtained as the average level of inefficiency among inefficient schools, for example, is not far from the 6.80% found by Mizala *et al.* (2002).

The gap in inefficiency between public schools and privately owned schools diminishes gradually when one includes in the model school-level variables. However, privately owned schools demonstrated a much higher level of performance, compared to public schools. One could reasonably expect that the inclusion of other variables that proxy for different types of resources might contribute to bridging the gap even further. This would reinforce the hypothesis that it would be best to endow higher subsidies to schools with more students at low socioeconomic levels, in order to offset the negative effect of a suboptimal mix of educational resources.

References

Aitkin, M. and Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, 149(1):1–43.

Anand, P., Mizala, A., and Repetto, A. (2009). Using school scholarships to estimate the effect of private education on the academic achievement of low-income students in Chile. *Economics of Education Review*, 28(3):370–381.

Balaguer-Coll, M. T., Prior, D., and Tortosa-Ausina, E. (2010). Devolution dynamics of Spanish local government. *Environment and Planning A*, 42(6):1476–1495.

Battese, G. E. and Rao, D. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business*, 1(2):87–93.

Battese, G. E., Rao, D., and O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, 21(1):91–103.

Bessent, A., Bessent, W., Kennington, J., and Reagan, B. (1982). An application of mathematical programming to assess productivity in the Houston independent school district. *Management Science*, 28(12):1355–1367.

Bifulco, R. and Bretschneider, S. (2001). Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review*, 20(5):417–429.

Blanco, E. (2010). Eficacia escolar y desigualdad: aportes para la política educativa. *Perfiles Latinoamericanos*, 17(34).

Bosker, R. J. and Witziers, B. (1996). The magnitude of school effects, or: Does it really matter which school a student attends. In *Annual Meeting of the American Educational Research Association*, New York.

Bryk, A. S. and Raudenbush, S. W. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Newbury Park, CA.

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

Cervini, R. A. (2009). Class, school, municipal, and state effects on mathematics achievement in Argentina: a multilevel analysis. *School Effectiveness and School Improvement*, 20(3):319–340.

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

Cook, W. D. and Seiford, L. M. (2009). Data envelopment analysis (DEA)-Thirty years on. *European Journal of Operational Research*, 192(1):1–17.

De Witte, K., Thanassoulis, E., Simpson, G., Battisti, G., and Charlesworth-May, A. (2010). Assessing pupil and school performance by non-parametric and parametric techniques. *Journal of the Operational Research Society*, 61(8):1224–1237.

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labor-efficiency in post offices. In Marchand, M., Pestieau, P., and Tulkens, H., editors, *The Performance of Public Enterprises: Concepts and Measurement*, chapter 10, pages 243–267. North-Holland, Amsterdam.

Elacqua, G., Schneider, M., and Buckley, J. (2006). School choice in Chile: Is it class or the classroom? *Journal of Policy Analysis and Management*, 25(3):577–601.

Emrouznejad, A., Parker, B. R., and Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, 42(3):151–157.

Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors (1993). *The Measurement of Productive Efficiency: Techniques and Applications*. Oxford University Press, Oxford.

Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors (2008). *The Measurement of Productive Efficiency and Productivity Change*. Oxford University Press, Oxford.

Goldstein, H. (1995). *Multilevel Statistical Models*. Wiley, London.

Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8(4):369–395.

Goldstein, H. and Woodhouse, G. (2000). School effectiveness research and educational policy. *Oxford Review of Education*, 26(3):353–363.

Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Kneip, A., Park, B. U., and Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, 14:783–793.

Kumar, S. and Russell, R. R. (2002). Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence. *American Economic Review*, 92(3):527–548.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews*, 15:261–274.

Loader, C. R. (1996). Local likelihood density estimation. *The Annals of Statistics*, 24(4):1602–1618.

Mancebón, M. J. and Mar Molinero, C. (2000). Performance in primary schools. *Journal of the Operational Research Society*, 51(7):843–854.

Mancebón, M. J. and Muñiz, M. A. (2007). Private versus public high schools in Spain: disentangling managerial and programme efficiencies. *Journal of the Operational Research Society*, 59(7):892–901.

Mizala, A. and Romaguera, P. (2004). School and teacher performance incentives: The Latin American experience. *International Journal of Educational Development*, 24(6):739–754.

Mizala, A., Romaguera, P., and Farren, D. (2002). The technical efficiency of schools in Chile. *Applied Economics*, 34(12):1533–1552.

Mizala, A., Romaguera, P., and Urquiola, M. (2007). Socioeconomic status or noise? Tradeoffs in the generation of school quality information. *Journal of Development Economics*, 84(1):61–75.

Murillo, F. J. (2010). ¿Quiénes son los responsables de los resultados de las evaluaciones? Hacia un planteamiento de valor agregado en educación. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 8(4):3–9.

Murillo-Melchor, C., Pastor, J. M., and Tortosa-Ausina, E. (2010). A bootstrap approach to analyse productivity growth in European banking. *Journal of the Operational Research Society*, 61:1729–1745.

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

Oliveira, M. A. and Santos, C. (2005). Assessing school efficiency in Portugal using FDH and bootstrapping. *Applied Economics*, 37(8):957–968.

Opdenakker, M. C. and Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal*, 27(4):407–432.

Ouellette, P. and Vierstraete, V. (2005). An evaluation of the efficiency of Québec's school boards using the Data Envelopment Analysis method. *Applied Economics*, 37(14):1643–1653.

Park, B. U., Simar, L., and Weiner, C. (2000). The FDH estimator for productivity efficiency scores. *Econometric Theory*, 16(6):855–877.

Pastor, J. M. and Tortosa-Ausina, E. (2008). Social capital and bank performance: An international comparison for OECD countries. *The Manchester School*, 76(2):223–265.

Pastor, J. T. (1996). Chapter 3 Translation invariance in Data Envelopment Analysis: A generalization. *Annals of Operations Research*, 66(2):91–102.

Patrinos, H. A. (1995). Socioeconomic background, schooling, experience, ability and monetary rewards in Greece. *Economics of Education Review*, 14(1):85–91.

Phillips, M. (1997). What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school. *American Educational Research Journal*, 34(4):633–662.

Ruggiero, J., Duncombe, W., and Miner, J. (1995). On the measurement and causes of technical inefficiency in local public services: With an application to public education. *Journal of Public Administration Research and Theory*, 5(4):403–428.

Sammons, P., Thomas, S., and Mortimore, P. (1997). *Forging links: Effective schools and effective departments*. Paul Chapman Publishing, London.

Silva Portela, M. C. A. and Camanho, A. S. (2010). Analysis of complementary methodologies for the estimation of school value added. *Journal of the Operational Research Society*, 61(7):1122–1132.

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

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.

Simar, L. and Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis*, 13(1):49–78.

Simar, L. and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In Fried, H., Lovell, C. A. K., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency*, chapter 4, pages 421–521. Oxford University Press, Oxford, 2nd edition.

Thanassoulis, E. and Silva Portela, M. C. A. (2002). School Outcomes: Sharing the Responsibility Between Pupil and School1. *Education Economics*, 10(2):183–207.

Thanssoulis, E. (1999). Setting achievement targets for school children. *Education Economics*, 7(2):101–119.

Wheelock, D. C. and Wilson, P. W. (2009). Robust nonparametric quantile estimation of efficiency and productivity change in U.S. commercial banking, 1985–2004. *Journal of Business and Economic Statistics*, 27(3):354–368.

Table 1: Number of students and schools in the sample

		Students		Schools	
		Number	%	Number	%
Type of school	Public	5,361	47.36	89	50.57
	Privately-owned subsidized	5,433	48.00	80	45.45
	Privately-owned fee-paying	525	4.64	7	3.98
	Total	11,319	100.00	176	100.00
Region	Metropolitan region	5,892	52.05	81	46.02
	Other regions	5,427	47.95	95	53,98
	Total	11,391	100.00	176	100.00

Source: own ellaboration.

Table 2: Selection of inputs and outputs, description and summary statistics

Level	Variable	Description	N	Minimum	Maximum	Mean	Std. dev.
Student level	y_1	Language scores	11,319	126.17	382.50	265.42	53.27
	y_2	Mathematics scores	11,319	104.87	377.54	254.19	54.16
	x_1	Socioeconomic and cultural level, student's family	11,319	3.24	10.00	6.71	0.86
School level	x_2	Quality teaching resources at school	11,319	7.52	10.00	8.95	0.51
	x_3	Socioeconomic and cultural level, school	11,319	7.55	10.00	8.41	0.45

Model 1: y_1, y_2, x_1
Model 2: y_1, y_2, x_1, x_2
Model 3: y_1, y_2, x_1, x_2, x_3

Table 3: Bipartite decomposition of overall efficiency (model 1, student level data), summary statistics

Inefficiency component	Mean	Median	Maximum	Minimum	Percentile 25	Percentile 75	Std. Dev.	% inefficient	Mean inefficient
Overall efficiency, OE (α'')	1.2336	1.2251	1.5977	1.0370	1.1364	1.3137	0.1155	100%	1.2336
Student's effect, STE (α')	1.1827	1.1796	1.4537	1.0713	1.1337	1.2245	0.0617	100%	1.1827
School effect, SCE (α''/α')	1.0423	1.0330	1.3536	0.9155	0.9899	1.0883	0.0723	70.50%	1.0751

Inputs: x_1; outputs: y_1, y_2.

Table 4: Tripartite decomposition of overall efficiency (model 2, student level data and controllable input), summary statistics

Inefficiency component	Mean	Median	Maximum	Minimum	1st quartile	3rd quartile	Std. Dev.	% inefficient	Mean inefficient
Overall efficiency, OE (α'')	1.2336	1.2251	1.5977	1.0370	1.1364	1.3137	0.1155	100%	1.2336
Student's effect, STE (α')	1.1827	1.1796	1.4537	1.0713	1.1337	1.2245	0.0617	100%	1.1827
School effect, SCE (α_1/α')	1.0247	1.0196	1.3388	0.8857	0.9746	1.0600	0.0695	61.40%	1.0646
Resource endowments' effect, REE (α''/α_1)	1.0173	1.0124	1.1164	0.9996	1.0063	1.0170	0.0210	92.00%	1.0188

Inputs: x_1, x_2; outputs: y_1, y_2.

Table 5: Quadripartite decomposition of overall efficiency (model 3, student level data and non-controllable inputs), summary statistics

Inefficiency component	Mean	Median	Maximum	Minimum	1st quartile	3rd quartile	Std. Dev.	% inefficient	Mean inefficient
Overall efficiency, OE (α'')	1.2336	1.2251	1.5977	1.0370	1.1364	1.3137	0.1155	100%	1.2336
Student's effect. STE (α')	1.1827	1.1796	1.4537	1.0713	1.1337	1.2245	0.0617	100%	1.1827
School effect. SCE (α_2/α')	1.0055	1.0019	1.2997	0.8854	0.9625	1.0401	0.0621	51.70%	1.0504
Spillover effect. SPE (α''/α_1)	1.0189	1.0140	1.1062	0.9996	1.0059	1.0246	0.0188	92.60%	1.0204
Resource endowments' effect. REE (α_1/α_2)	1.0173	1.0124	1.1164	0.9996	1.0063	1.0170	0.0210	92.00%	1.0188

Inputs: x_1, x_2, x_3; outputs: y_1, y_2.

Table 6: Global decomposition by school type, summary statistics

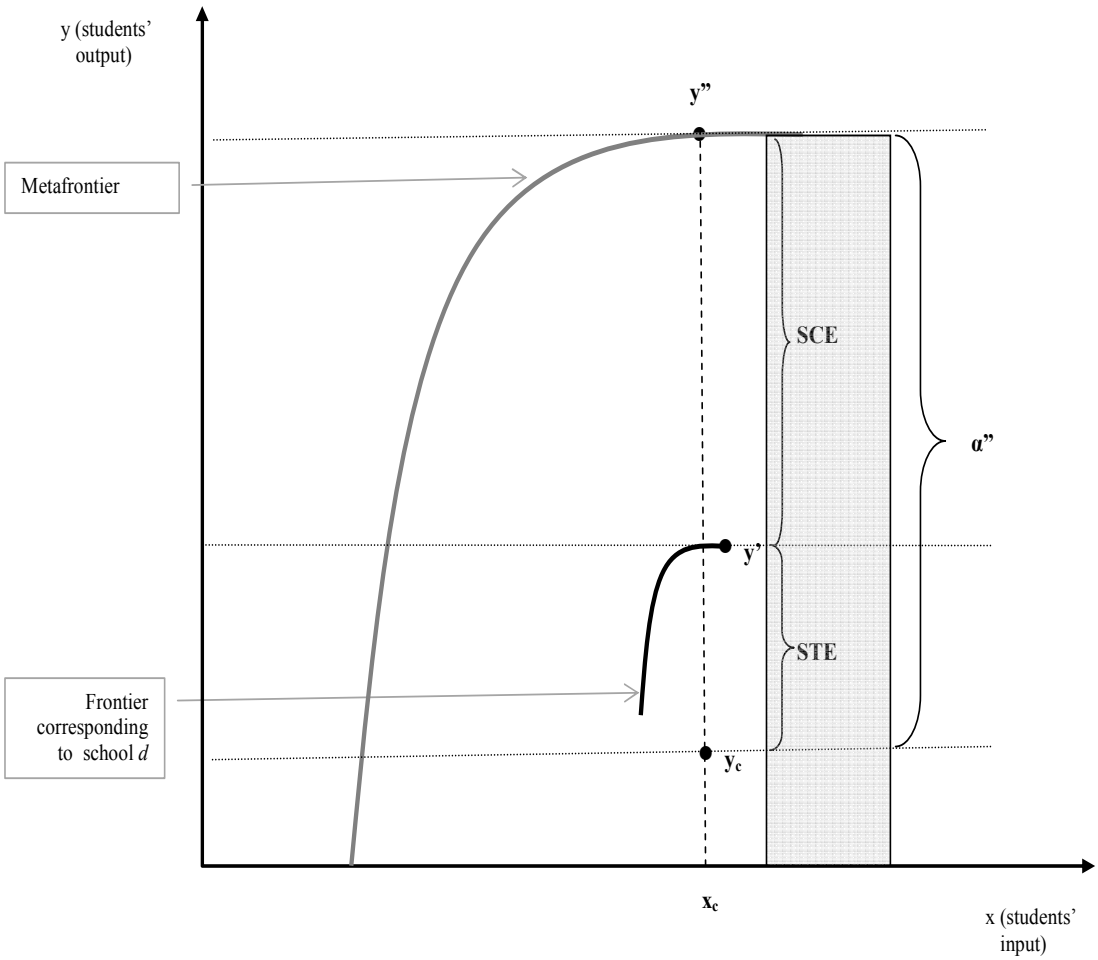
	Overall effect	Student's effect	School effect			Spillover effect			Resource endowments effect		
Type of school	α''	α'	α_2/α'	Mean (inefficient)	Inefficient schools (%)	α_1/α_2	Mean (inefficient)	Inefficient schools (%)	α''/α_1	Mean (inefficient)	Inefficient schools (%)
Public	1.3008	1.2123	1.0245	1.0594	62.92	1.0259	1.0264	97.75	1.0216	1.0227	95.51
Privately-owned subsidized	1.1699	1.1535	0.9880	1.0362	42.50	1.0129	1.0137	93.75	1.0132	1.0149	88.75
Privately-owned fee-paying	1.1063	1.1398	0.9630	1.0349	14.29	1.0003	1.0008	14.29	1.0080	1.0093	85.71
Total	1.2336	1.1827	1.0055	1.0504	51.70	1.0189	1.0204	92.61	1.0173	1.0188	92.05

Table 7: Distribution hypothesis tests by school type (Li, 1996)

Effect		Li (1996) test results	Public vs. privately-owned subsidized (1)	Public vs. privately-owned fee-paying (2)	Privately-owned subsidized vs. privately-owned fee-paying (3)
Overall effect	OE	T-statistic	16.6565	3.5769	0.2620
		p-value	0.0000	0.0002	0.3967
Student effect	STE_1	T-statistic	10.7097	1.7643	–0.6843
		p-value	0.0000	0.0388	0.7531
School effect	SCE_1	$t-$statistic	6.4774	2.1523	0.4534
		p-value	0.0000	0.0157	0.3251
	SCE_2	T-statistic	3.0585	1.4661	–0.1962
		p-value	0.0000	0.0713	0.5778
Spillover effect	SPE_3	T-statistic	2.3794	0.0622	–0.7525
		p-value	0.0087	0.4752	0.7741
Resource endowments' effect	REE_2	T-statistic	6.5077	3.1242	1.4237
		p-value	0.0000	0.0009	0.0773
	REE_3	T-statistic	7.0358	2.5676	2.3086
		p-value	0.0000	0.0051	0.0105

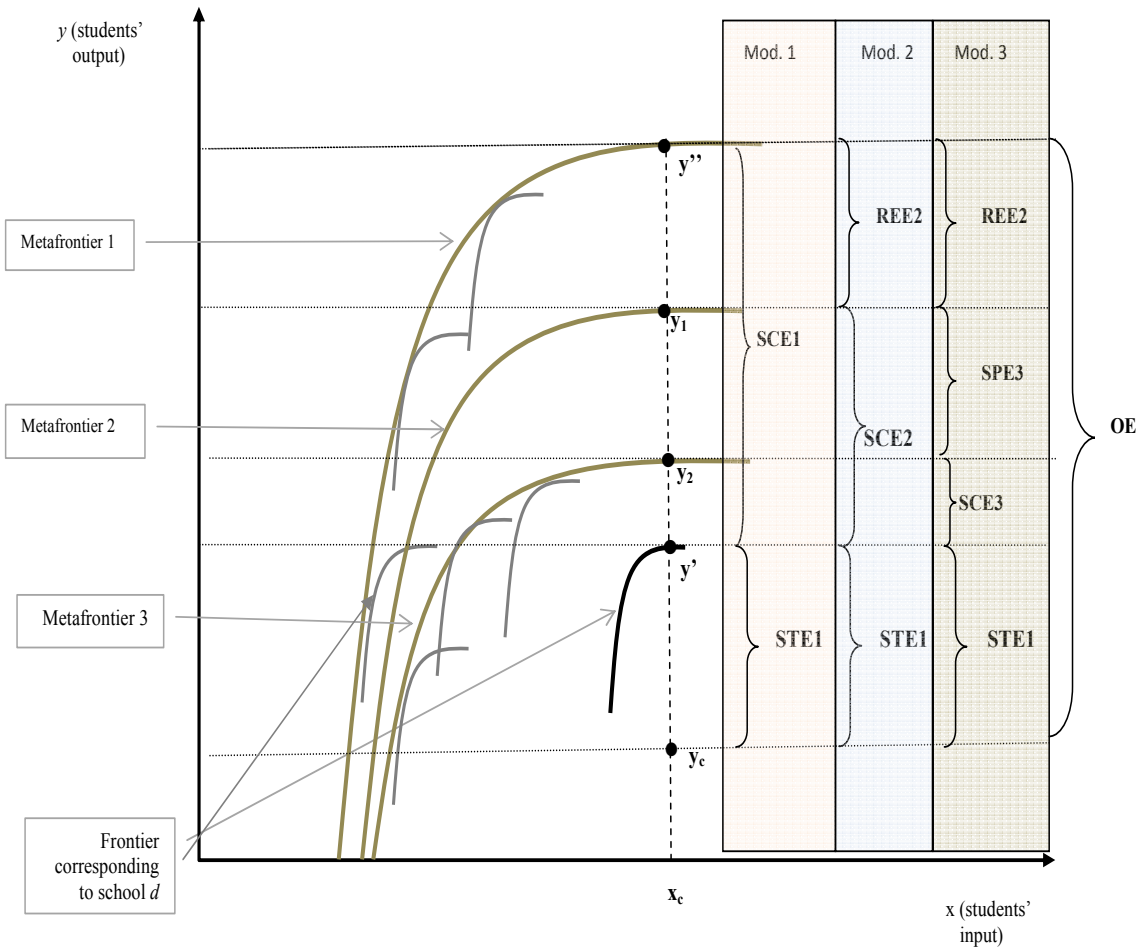
(1) $H_0 : f(\text{public}) = f(\text{privately-owned subsidized})$; $H_1 : f(\text{public}) \neq f(\text{privately-owned subsidized})$.
(2) $H_0 : f(\text{public}) = f(\text{privately-owned fee-paying})$; $H_1 : f(\text{public}) \neq f(\text{privately-owned fee-paying})$.
(3) $H_0 : f(\text{privately-owned subsidized}) = f(\text{privately-owned fee-paying})$; $H_1 : f(\text{privately-owned subsidized}) \neq f(\text{privately-owned fee-paying})$.

Figure 1: Silva Portela and Thanassoulis (2001) decomposition of student's inefficiency



Metafrontier: assumes the optimal level of resources and environmental factors to help student *c* to chieve the maximum output.

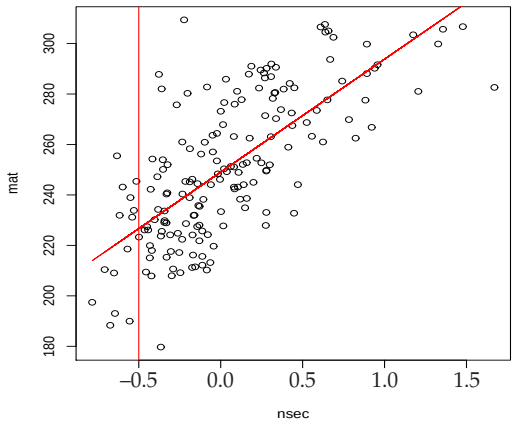
Figure 2: Proposed multilevel decomposition of inefficiency



Metafrontier 1 assumes that school *d* has the optimal level of resources and environmental factors to help student *c* to achieve the maximum output (this is the implicit assumption taken in Silva Portela and Thanassoulis, 2001).
Metafrontier 2 considers the observed level of resources corresponding to school *d* but the optimal level of environmental factors.
Metafrontier 3 considers the observed levels of resources and environmental factors corresponding to school *d*.

Figure 3: Educational attainment vs. socioeconomic level at school and student's level

(a) School



(b) Student

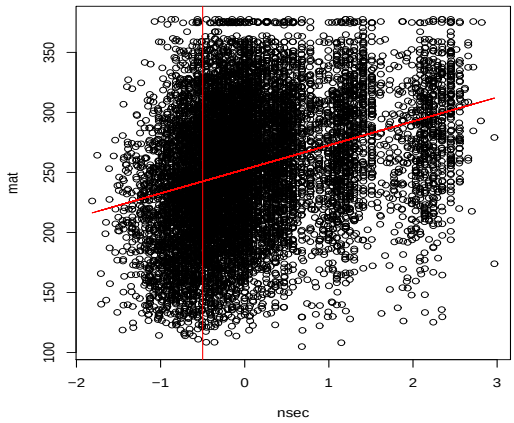
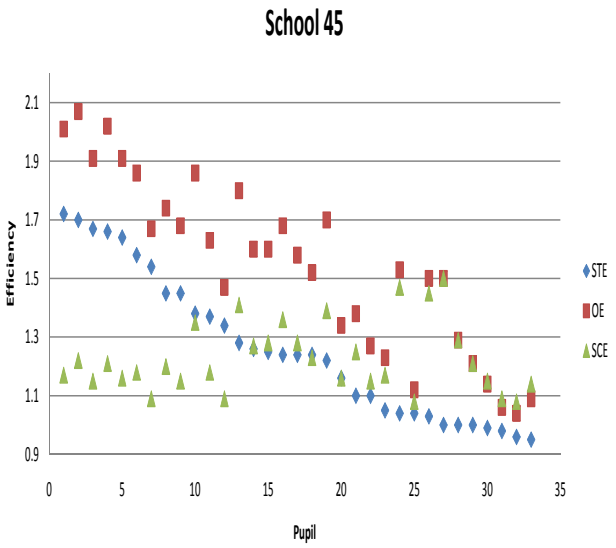


Figure 4: Educational attainment vs. socioeconomic level, schools 45 and 117

(a) School 45



(b) School 117

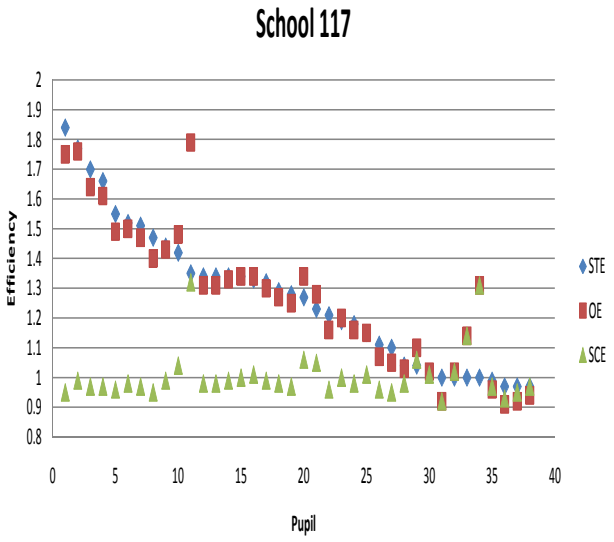
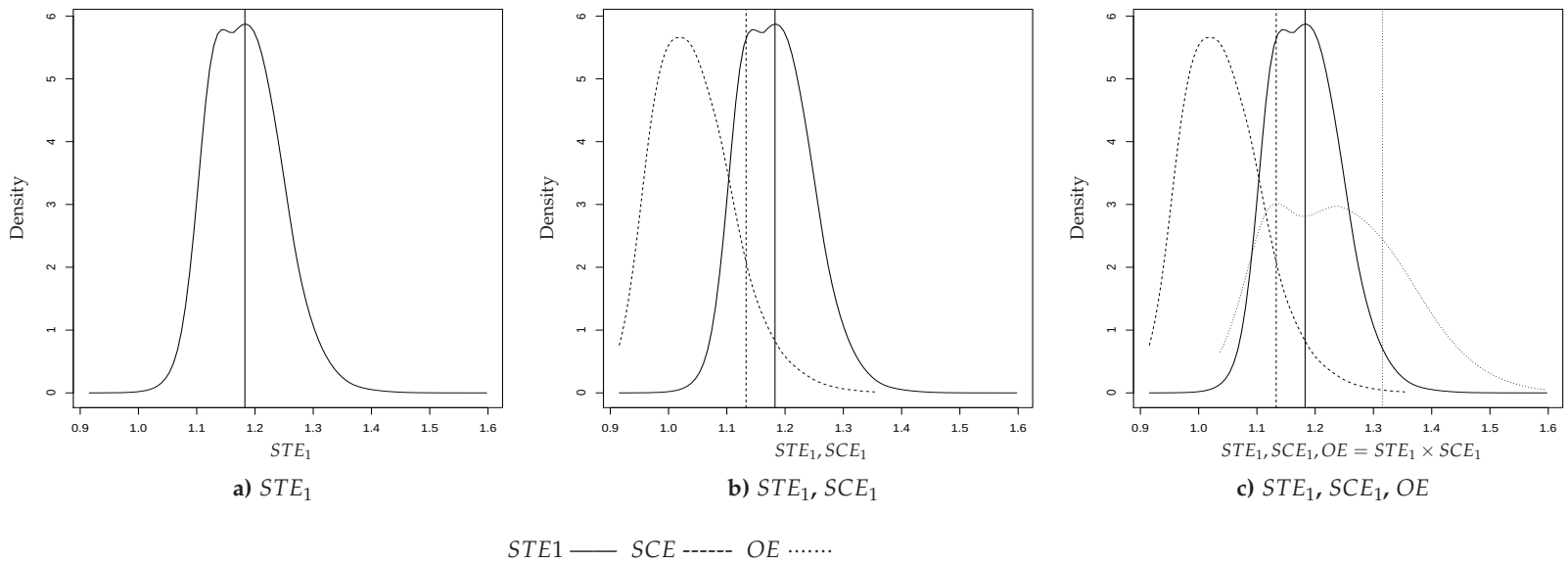


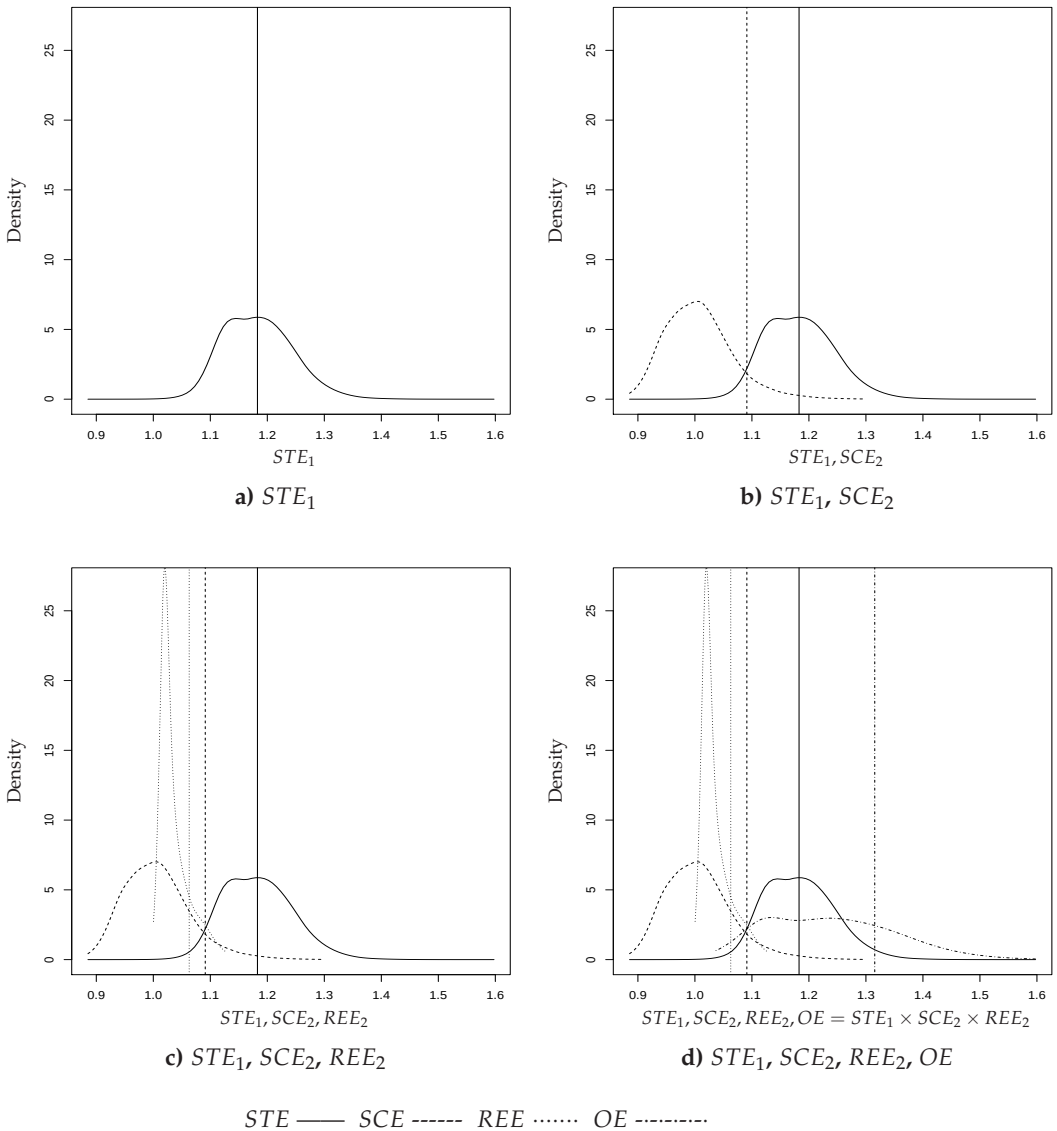
Figure 5: Kernel density plots of the bipartite decomposition of overall efficiency (model 1)



a) STE_1 **b)** STE_1, SCE_1 **c)** STE_1, SCE_1, OE

Notes: All figures contain densities estimated using kernel density estimation for the different components of the bipartite decomposition in expression (1). The vertical lines in each plot represents the average for each component of the decomposition. Densities have been estimated using local likelihood methods (Loader, 1996), and a Gaussian kernel has been chosen.

Figure 6: Kernel density plots of the tripartite decomposition of overall efficiency (model 2)



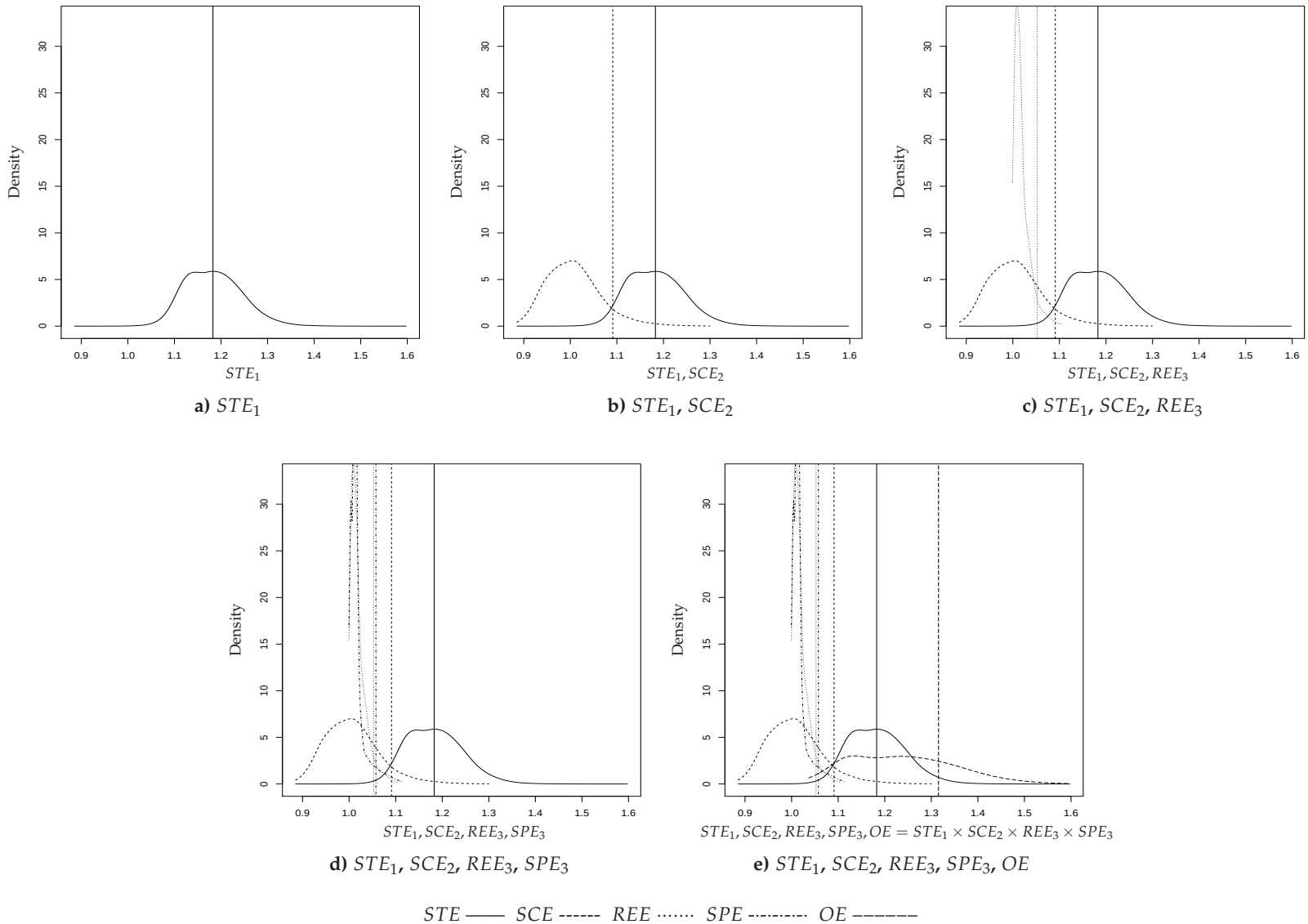
a) STE_1 **b)** STE_1, SCE_2

c) STE_1, SCE_2, REE_2 **d)** STE_1, SCE_2, REE_2, OE

STE —— SCE ------ REE ····· OE -·-·-·-

Notes: All figures contain densities estimated using kernel density estimation for the different components of the bipartite decomposition in expression (1). The vertical lines in each plot represents the average for each component of the decomposition. Densities have been estimated using local likelihood methods (Loader, 1996), and a Gaussian kernel has been chosen.

Figure 7: Kernel density plots of the quadripartite decomposition of overall efficiency (model 3)



a) STE_1 **b)** STE_1, SCE_2 **c)** STE_1, SCE_2, REE_3

d) STE_1, SCE_2, REE_3, SPE_3 **e)** STE_1, SCE_2, REE_3, SPE_3, OE

Notes: All figures contain densities estimated using kernel density estimation for the different components of the bipartite decomposition in expression (1). The vertical lines in each plot represents the average for each component of the decomposition. Densities have been estimated using local likelihood methods (Loader, 1996), and a Gaussian kernel has been chosen.

Figure 8: Internal (student) inefficiency (*STE*) vs. school inefficiency (*SCE*)