UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

TESIS DOCTORAL

# Essays on the estimations of educational technical efficiency under endogeneity

**(Ensayos sobre la estimación de la eficiencia técnica educativa bajo la presencia de endogeneidad)**

MEMORIA PARA OPTAR AL GRADO DE DOCTORA
PRESENTADA POR

**Gabriela Sicilia**

Director

Daniel Santín

**Madrid, 2015**

# Essays on the estimation of educational technical efficiency under endogeneity

(Ensayos sobre la estimación de la eficiencia técnica educativa
bajo la presencia de endogeneidad)

Gabriela Sicilia

**Ph.D. dissertation**

Supervisor:
Daniel Santín
Complutense University of Madrid

Department of Applied Economics VI: Public Finance and Tax System
Faculty of Economics and Business Administration
Universidad Complutense de Madrid
Campus de Somosaguas, 28223 Madrid (Spain)

December, 2014

*To my family, for their*
*constant love, support*
*and understanding*

# Agradecimientos

Este proyecto de Tesis Doctoral ha significado un gran esfuerzo durante muchos años, el cual naturalmente, no podría haber llevado a cabo sin el apoyo, consejo, colaboración y comprensión de muchas personas que forman parte de mi vida. Ahora quiero agradecerles y compartir con ustedes mi satisfacción y alegría de haber llegado hasta aquí.

Los primeros responsables de mi éxito académico son mis padres Alba y Héctor. Decir gracias mil veces no alcanzaría para expresar todo mi agradecimiento hacia ellos. He tenido la suerte de tener padres muy motivados, que valoran enormemente la educación de sus hijos y que siempre, hasta en las circunstancias más adversas, han priorizado nuestra educación y han dedicado muchos esfuerzos a buscar las mejores oportunidades. Más allá de su motivación, quiero agradecerles porque me han enseñado que las oportunidades hay que aprovecharlas y que nuestro éxito depende de nuestro esfuerzo. Pero sobre todo, quiero darles las gracias, porque siempre, pero siempre, me han apoyado en mis decisiones y pese a la distancia, me han acompañado en todo momento en este largo viaje. Y no sería justa si no incluyera en esta última frase a mi hermano Fede, que sin duda ha sido el más afectado por mis ausencias durante este tiempo y sin embargo tus palabras siempre fueron de ánimo y afecto. Sin tu apoyo incondicional hoy no estaría aquí, gracias. Y gracias a vos Guille, con quien cuatro años atrás decidimos emprender juntos este proyecto sabiendo que iba a ser un gran desafío, de mucho esfuerzo y trabajo. Gracias por tu constante apoyo y comprensión, sobre todo en momentos críticos, y por recordarme que trabajando duro las cosas finalmente salen. Espero que la alegría que sentimos en este momento compense tanto esfuerzo y te motive para acabar pronto tu proyecto también.

Pero sin dudas, la persona que más crédito tiene en el desarrollo de este trabajo, y a quien hace mucho quiero escribirle estas palabras, es mi director y amigo Dani. Gracias, gracias y gracias. Gracias por tu entusiasmo, dedicación y por tu confianza en mí desde el primer momento, que han sido los responsables de que hace tres años emprendiera este largo y gratificante camino. Gracias por las incalculables horas de trabajo dedicadas, por las discusiones eternas que parecían no acabar, por los consejos que a veces no supe escuchar, por el ánimo en momentos críticos, por las risas compartidas, en fin, por tu inmenso compromiso sin el cual esta tesis nunca se hubiera realizado. Pero sobre todo quiero darte las gracias por confiar en mí, por dejarme hacer y por tu paciencia y comprensión, que aunque sabiendo muchas veces que tenías razón desde el inicio has dejado que me equivocara para que finalmente tomara el camino adecuado y realmente aprehendiera.

Quiero también agradecer al Departamento de Hacienda Pública y Sistema Fiscal de la Universidad Complutense de Madrid y en especial a los integrantes de la Comisión de Doctorado: Aurelia Valiño, Raquel Paredes, Rosa Urbanos y Jorge Onrubia por su respaldo y valiosos comentarios a lo largo de todo el proyecto. Aurelia, en especial quiero agradecerte por tu colaboración incondicional recibida desde el inicio como coordinadora del programa del doctorado. Gracias Miriam por el apoyo material que me brindaste de forma desinteresada y por tus palabras de ánimo. Gracias Jorge y Jesús por sus valiosas críticas constructivas a este trabajo y también por las constantes palabras de ánimo tan necesarias sobre todo al final de este largo proceso.

Mis más sinceras gracias a José Manuel Cordero por su inestimable contribución a esta tesis, que en parte es también suya. Gracias Jose no sólo por tu aporte como investigador sino también como amigo. Gracias por tu paciencia, tu confianza en mi trabajo y tu ánimo constante durante todo este tiempo. También quiero agradecer a Juan Palomino por sus valiosas aportaciones y por las innumerables discusiones que hemos compartido a lo largo de todo el proceso. Desde tu rol de ejecutor de política educativa me has ayudado a entender mejor un proceso tan complejo como es la educación. Estaré también siempre agradecida por tu excelente revisión de la versión final de este trabajo que sin duda ha mejorado el producto final.

Y gracias a mis amigos y familia. A "los de de allá" (a quienes no nombro uno a uno porque afortunadamente son muchos), gracias porque lograron que nunca me sintiera fuera de casa. Siempre me demostraron su cariño y constante apoyo a la distancia. A "mis amigos de acá", porque han hecho que aún estando tan lejos me sienta como en casa. En especial quiero agradecer a Mar, Eduard, Pi, Albert, Juan, Mateo y Umpi. De una u otra forma todos han aportado su granito de arena a este proyecto y les estaré siempre agradecida. Fueron muchas las experiencias que compartimos juntos estos años, pero sobre todo, fueron muchas las horas que me escucharon con interés hablar de mi tesis, que discutimos muchas de las ideas aquí contenidas o que simplemente compartimos largas mañanas y tardes en alguna biblioteca madrileña trabajando hasta que nos echaran. A los siete, mil gracias.

Finalmente, quiero agradecer también a diversas instituciones que me han brindado apoyo económico durante estos años y que me han permitido llevar a cabo este proyecto. Agradezco (en orden cronológico) a la Fundación Carolina y a la Universidad de la República (Uruguay), a la Agencia Nacional de Investigación e Innovación (Uruguay) y a la Fundación Ramón Areces por la financiación recibida a través de sus diversos programas de Becas. Paralelamente quiero agradecer a la Cátedra de Familia de la UCM, y en especial a Teresa López quien ha confiado en mi trabajo para diversos proyectos a lo largo de estos años. Asimismo agradezco al Instituto Nacional de Evaluación Educativa (MECD) y a la Fundación Europea Sociedad y Educación por brindarme la oportunidad de participar en sus proyectos, y en particular a Mercedes Esteban por confiar en mi trabajo siempre.

*"Life is not easy for any of us. But what of that?
We must have perseverance and above all
confidence in ourselves. We must believe that
we are gifted for something, and that this thing,
at whatever cost, must be attained."*

Marie Curie

# Contents

# List of Figures

# List of Tables

# Introduction

*"Efficiency is doing better what*
*is already being done"*

Peter F. Drucker

The present Ph.D. dissertation aims to contribute theoretically and empirically to understand the extent to which the endogeneity problem, a major concern frequently observed in educational production processes, affects the estimation of technical efficiency using the Data Envelopment Analysis (DEA) technique. Furthermore, this research combines insights from impact evaluation literature and nonparametric frontier techniques in order to provide potential solutions to deal with this problem in educational empirical applications and obtain more accurate efficiency estimates. To do that, three chapters are developed. Although they are closely related, they have their own internal structure as they intend to be free-standing (in the sense that each one can be read and understood independently). Still, some common concepts, definitions and methodologies are exposed whenever required.

The evaluation of technical efficiency in the Public Sector has gained growing attention over the last decades. Public services providers have a natural interest in efficiency assessments since they face up increasing demands of quantities and quality together with financial constraints. Within this framework, the measurement of educational technical efficiency is one of the current major concerns as the education expenditure is one of the largest public budget items and the public sector is usually the main provider of education in most modern countries.

Given that the investment in quality education is essential to ensure sustainable development and economic growth (Barro and Lee, 1996, 2012; Hanushek and Kimbo, 2000; De la Fuente, 2011; Hanushek and Woessmann, 2012a, 2012b), several countries in the last decades have significantly increased their public educational budget. However, these efforts have not always been translated into better academic achievements. This fact has led researchers and policy-makers to wonder why these additional investments in educational resources do not lead to improvements in the quality of education. Although the answer is not evident, this fact alerts about the presence of great inefficiencies in schooling production and has spurred the interest in measuring these inefficiencies and explaining their main sources, with the ultimate goal of correcting these behaviours.

The educational production has, like most public sector production processes, some special characteristics that complicate the estimation of accurate efficiency measures (*i.e.* the completely unknown production technology, the lack of prices information or the frequent use of multiple proxy variables to approximate the real output). In this sense, nonparametric techniques and particularly the DEA model proposed by Charnes, Cooper and Rhodes (1978) and Banker, Charnes and Cooper (1984) are the most commonly applied methods for measuring educational technical efficiency (Worthington, 2001). This is mainly because of two reasons: its flexibility allows to adapting it to the stated particularities of this sector, and the results of this technique can be easily translated to stakeholders and politicians.

However, there is a major concern frequently observed in educational production processes which has been overlooked in the context of the technical efficiency estimation: the endogeneity problem. In statistical terms, this phenomenon implies the presence of a significant correlation between one input and the error term, and it can arise as the result of multiple sources (e.g. measurement errors, unobserved heterogeneity, the omission of relevant variables in the model specification or the presence of simultaneity). In the context of the estimation of technical efficiency with frontier techniques, this problem of endogeneity implies the presence of a significant correlation between at least one input and the efficiency term (Peyrache and Coelli, 2009).

In the education provision framework, the most common source of endogeneity is the educational self-selection. Students are not exogenously assigned to schools but their allocation depends on decisions made by parents, teachers and schools' principals. Indeed, this problem has been one of the focuses of attention in econometrics along the last three decades. Endogeneity has been argued to be the basis for multiple theoretical and empirical critiques of traditional findings and multiple methods have been developed in the literature to deal with this problem (Webbink 2005, Schlotter *et al.* 2011).

However, this widespread acknowledgement in the context of econometrics of the existence of the self-selection or the endogeneity problem is ignored when we move into the world of the efficiency estimation. There are only a handful of studies that using alternative simulation strategies have tested the performance of DEA under some kind of endogeneity (Gong and Sickles, 1992; Orme and Smith, 1996; Bifulco and Bretschneider, 2001, 2003; Ruggiero, 2003, 2004). Consequently, this problem is still an unknown and incipient issue in the literature of the estimation of frontiers using DEA and thus it is frequently overlooked when practitioners apply this technique.

Based on this background, the following three chapters of this Ph.D. dissertation address the endogeneity problem, measure its effects on the estimation of technical efficiency and provide different strategies to deal with it.

Chapter 1 analyses theoretically to which extent does the presence of endogeneity in the production process affect DEA estimates in finite samples, so practitioners performing this technique can be aware of the accuracy of their estimates. To do this, we firstly illustrate the endogeneity problem and its implications for the efficiency estimation from a conceptual perspective. Secondly, using synthetic data generated in a Monte Carlo experiment we evaluate how different levels of positive and negative endogeneity can affect DEA performance. We conclude that, although DEA is robust to negative endogeneity (Bifulco and Bretshneider, 2001, 2003 and Ruggiero, 2003), estimates could be severely impaired under the presence of a significant positive endogeneity, that is, when one input in the production process is highly positively correlated with the true efficiency term. This decline in DEA performance is further driven by the misidentification of the most inefficient DMUs with low levels of the endogenous input.

From these findings, the question that arises is: how can we deal with this problem in empirical research? In this direction, based on the Monte Carlo experiment results we propose a simple heuristic to detect this phenomenon in empirical applications. In addition, we get insights from causal inference literature, and particularly, from the Instrumental Variable (IV) approach developed in econometrics, to provide a potential solution to deal with this problem: the '*Instrumental Input DEA*' (II-DEA) strategy. Again, using a Monte Carlo experiment we test the performance of this proposal in finite samples.

Building upon this evidence, Chapter 2 implements these strategies to deal with the endogeneity problem in applied research. Using data from Uruguayan public secondary schools we use the proposed heuristic method to identify potential endogenous inputs. We actually found that the school's average socio-economic level (peer group) is highly correlated with schools' efficiency. Given this result, we tackle this problem by applying the II-DEA strategy proposed in Chapter 1 to obtain reliable technical efficiency estimates. We compare these results with those that arise from the conventional DEA to empirically investigate the impact of not controlling for the presence of endogeneity. Beyond estimating the efficiency potential improvements for each school and identifying the better and the worst performers, we aim to explore the explanatory factors of the efficient behaviours. Thus, once we have estimated the II-DEA efficiency scores, we regress them on several contextual variables related to students and schools characteristics. The results of this second stage allow us to draw conclusions about which educational policies and practices would be desirable to design and promote in order to improve the quality of education.

The II-DEA strategy proposed in the first chapter and implemented in Chapter 2 requires finding a good instrument. This is not an easy task and, in some contexts, it may not even be possible to find one. In the third chapter, taking again insights from the impact evaluation literature we provide an alternative strategy to deal with the endogeneity problem in the estimation of educational technical efficiency.

Chapter 3 focuses on the estimation of teachers' technical efficiency and its effect on students' academic results taking into account the presence of self-selection. To tackle this problem we take advantage of a database for Spanish primary schools where we can identify those schools where two classrooms were evaluated and where students were randomly assigned into these classrooms. This implies that, on average, students in both classrooms are similar (both in observable and unobservable characteristics), since parents can self-select into schools but they cannot choose the classroom inside the school. Therefore, the only difference between classrooms in each school is the teacher who was randomly assigned. This randomization produces a natural experiment where by chance one classroom has been assigned to the most efficient teacher and the other one to the least efficient teacher.

In sum, the strategy proposed in Chapter 3 consists on estimating the efficiency level for each classroom within schools with random assignment, and on exploiting the exogenous efficiency variation between classrooms within schools. This strategy allows us to obtain an unbiased measure of the true teacher's effect on students' achievement and to explore the main drivers of teachers' efficiency. As in the previous chapter, we also perform the analysis without taking into account the presence of self-selection to empirically quantify the effect of this problem in terms of educational public policy recommendations.

To conclude, for conducting a research it is evident that having a novel and relevant motivating question is crucial if we aim to contribute to better understand a specific problem or to scientific progress. But it is not a sufficient condition. The best question in the world becomes useless if we answer it with an inappropriate technique. Both a relevant question and an accurate method to answer it are necessary. In this regard, this Ph.D. dissertation attempts to be a helpful methodological contribution, which we expect it could be applied in the near future to answer pertinent questions not only in the context of the measurement of educational technical efficiency, as we do here, but also in other fields where the endogeneity problem is present.

# Chapter 1

# Dealing with the endogeneity problem in Data Envelopment Analysis

## 1.1 Introduction

The evaluation of technical efficiency of Decision Making Units (DMUs) is basic for adopting organizational decisions in order to save resources, monitoring DMUs activity to detect best and worst performers and for improving results. Public services providers have a natural interest in efficiency assessments since they face both increasing demands of quantities and quality and financial constraints. However, the special characteristics of the public sector production, *i.e.* the lack of profit maximization behaviours, the completely unknown production technology or the frequent use of multiple proxy variables to approach the real output complicate the estimation of accurate efficiency measures (Bowlin, 1986).

In these contexts, nonparametric techniques, and especially Data Envelopment Analysis (DEA), are the most commonly applied methods for measuring technical efficiency relative to an estimate of an unobserved true frontier in multiple frameworks (Gattoufi *et al.*, 2004). The main reason is its flexibility and the few assumptions needed about the implicit technology that relates inputs with outputs. Therefore, this technique does not assume *a priori* a particular functional form for the underlying production technology (only some axiomatic assumptions, *i.e.* monotonicity and concavity) or the inefficiency distribution. Thus the frontier is drawn by the observed data resulting from an underlying and unknown data generating process. By contrast, the most important and traditional limitation of this technique has been the lack of statistical foundations and the inability to perform statistical inference. However, Banker (1993) and Korotelev *et al.* (1995) were the first who demonstrated that, under certain assumptions, DEA estimators are statistically consistent and have a known rate of convergence. Likewise, the asymptotic distribution of DEA estimators has also been derived and different bootstrap methods have been proposed for conducting valid inference about the true efficiency from the DEA estimates in a multivariate framework (Gijbels *et al.*, 1999; Kneip *et al.*, 1998, 2008, 2011; Simar and Wilson, 2008).

Within this framework, selecting the appropriate input and output variables to include in the model is one of the most critical choices that practitioners will have to undertake in order to obtain reliable efficiency scores. This point has also received a lot of attention in the DEA literature over the past decades, where several works have analysed the effects of misidentification on DEA estimates (Smith, 1997; Pedraja-Chaparro *et al.*, 1999; Dyson *et al.* 2001; Simar and Wilson, 2001; Galagedera and Silvapulle, 2003; Ruggiero, 2005; Morita and Avkiran, 2009; Nataraja and Johnson, 2011). In addition, several studies have analyzed using simulated data how the presence of random noise or measurement errors can affect the performance of DEA estimates (Banker *et al.*, 1993; Bojanic *et al.*, 1998; Ruggiero, 2004; Simar, 2007; Krüger, 2012). Moreover, different extensions of the technique have been developed in order to improve its robustness, for example to correct for the presence of outliers or to include non-discretionary inputs in the model[1].

[1]See Simar and Wilson (2011) for a detail review of multi-stage models.

However, there is another major concern, namely, the presence of endogeneity in the production process, which is frequently overlooked when practitioners apply DEA. In statistical terms, this phenomenon implies the presence of a significant correlation between the error term and at least one explanatory variable. Peyrache and Coelli (2009) pointed out that in the estimation of technical efficiency with frontier techniques framework, the endogeneity arises when at least one input is correlated with the efficiency term. Although the potential distortions that this endogeneity can cause on the estimation of economic models have been widely studied in the econometrics literature, its effects on efficiency measures calculated using nonparametric frontier techniques like DEA have not been analysed in depth yet. There are only a handful of studies that using alternative simulation strategies have tested the performance of DEA under some kind of endogeneity (Gong and Sickles, 1992; Orme and Smith, 1996; Bifulco and Bretschneider, 2001, 2003; Ruggiero, 2003, 2004). However, these previous works do not allow drawing general conclusions about the potential distortions of this issue on DEA estimates.

Gong and Sickles (1992) compare Stochastic Frontier Analysis (SFA) with DEA using different Monte Carlo experiments based on panel data generated by a CRESH production function with three inputs and a single output considering different time periods. With regard to our aim, they examine the effect that a rather low negative correlation (from -0.21 to -0.37) between inputs levels and technical efficiency may have on both techniques and conclude that DEA measures are much closer to the true levels of efficiency than those estimated with SFA. Orme and Smith (1996) also conduct a Monte Carlo simulation to evaluate the performance of DEA under the presence of endogeneity in data. Their data generation process (DGP) relies on a Cobb-Douglas production function with constant returns to scale and they only consider a negative correlation between inputs and the efficiency. They conclude that the efficiency estimates generated by DEA in the presence of this negative endogeneity can be subject to bias, in the sense that inefficient units using low levels of the endogenous resource may be set tougher efficiency targets than equally inefficient units using more resources.

More recently, Bifulco and Bretschneider (2001) use simulated data with the aim of assessing the performance of two alternative methods (DEA and COLS) in different scenarios characterized by the presence of measurement error and a high level of negative correlation between inputs and the efficiency term (ranging from 0.78 to 0.92). For that purpose, they also use a log linear Cobb-Douglas production function and assume constant return to scale to generate data. They conclude that without measurement error the performance of DEA does not change substantially when negative correlation between inefficiency and one of the inputs is present (consistent result with Gong and Sickles, 1992). Their main contribution was meant to be the use of a production technology with two outputs and three inputs in an attempt to emulate the characteristics of educational production contexts. Unfortunately, this function was inconsistent with economic theory[2], since they were actually generating an increasing return to scale

[2] Essentially, the problem arises because the second output can actually be interpreted as the inverse of a fourth input, since inefficiency is modelled as an output reduction of the other output.

technology with one output and four inputs (Ruggiero, 2003). In a subsequent paper, Ruggiero (2003) uses a corrected DGP based also on a Cobb-Douglas production function with only one output and concludes that DEA provides decent measures of efficiency even in the presence of negative endogeneity if there is not measurement error. Afterwards, Bifulco and Bretschneider (2003) perform a new simulation study using the same corrected DGP and they conclude that the primary results of their study remain. Finally, Ruggiero (2004) using simulated data is the only work who analyses the effect of a positive correlation between true technical efficiency and one non-discretionary environmental variable, showing that in this case DEA efficiency estimates are biased upward. Although, naturally, non-discretionary variables differ from the inputs in the DEA model specification, these results provide a useful basis for comparison with our results in the Cobb-Douglas scenario.

Thus, the first aim of this research is to analyse more generally whether the presence of endogeneity can bias or not the results obtained with DEA, so that practitioners using this technique can be aware of the accuracy of their estimates. In this regard, we have attempted to overcome some of the limitations of previous works in order to obtain more general conclusions about the effect of endogeneity in the DEA estimates. Firstly, we focus the analysis only on DEA performance and on determining how the presence of endogeneity affects DEA estimates instead of comparing its performance with alternative methods to measure technical efficiency. Secondly, we incorporate a more flexible *Translog* production function in addition to the traditional Cobb Douglas, which fails to capture the potential nonlinear effects of inputs on the output variable[3]. Thirdly, we conduct our simulations by performing a Monte Carlo experiment to provide more robust results than most of previous studies that based their conclusions on a single replication. Finally, we simulate different intensities of both the negative and the positive endogeneity, whereas all previous studies only examine the effect of the negative correlation between the inputs and the true efficiency.

The second objective of this chapter is more ambitious and challenging. As practitioners we wonder how we can deal with the endogeneity issue in an empirical research. From this question two issues arise: how to identify the presence of an endogenous input and, how to tackle this problem in order to improve DEA estimations. In this direction, firstly, we propose a simple heuristic to identify the presence of correlation between an input and technical efficiency. Then, we propose an *Instrumental Input DEA* (II-DEA) strategy for dealing with this problem in empirical DEA applications. Again, it is important to stress the relevance of these two contributions for the nonparametric efficiency models literature, since although this issues have receive considerably attention in statistics and econometrics, there are almost no previous studies that have dealt with these issues in the context of efficiency models.

In this sense, Wilson (2003) explores a number of relative simple independence tests that can be used in the context of efficiency estimation, and provides some empirical examples to illustrate their use. However, his Monte Carlo results show that these tests have poor size

[3]This can be a significant weakness in complex production frameworks such as education or health provision.

properties and low power in moderate sample sizes. Based on this work, Peyrache and Coelli (2008) propose a semi-parametric Hausmann-type asymptotic test for linear independence and, using a Monte Carlo experiment, they show that it has good size and power properties in finite samples. However, the proposed test has a major limitation because it is based on the distribution of the true technical efficiency, which is, of course, unobservable. To solve this, the authors propose using the empirical distribution of the individual efficiencies estimated via nonparametric techniques (DEA) or Free Disposal Hull (FDH) assuming that these are consistent estimators of the true efficiency. However, for this consistency to hold true, inputs and efficiency must be uncorrelated, which is the same hypothesis that is being tested. In this sense, the heuristic to detect an endogeneity problem proposed in this chapter overcomes this limitation as it is based on the correlations coefficients between inputs and the estimated DEA efficiency scores, but does not require any previous assumption about the distribution of the true efficiency or the consistency of their estimates.

Finally, it is worth noting that the endogeneity problem has been also considered recently in the estimation of technical efficiency using parametric frontier techniques in empirical research. For example, Solís *et al.* (2007) employ a switching regression model to handle the selection bias in hillside farmers under different levels of adoption of soil conservation in El Salvador and Honduras. Greene (2010) proposes a simple way to extend the Heckman sample selection model to the stochastic frontier analysis framework and apply it to measure state health system performance. Perelman and Santín (2011) address the endogeneity problem of school choice in Spain using instrumental variables. Finally, Mayen *et al.* (2010); Bravo-Ureta *et al.* (2012) and Crespo-Cebada *et al.* (2013) apply propensity score matching to American dairy farms, farmers in Honduras and education in Spain, respectively.

The rest of the chapter is organized as follows. Section 1.2 conceptually illustrates the endogeneity issue and the potential effects that can arise on DEA estimates. Section 1.3 describes the methodology used to generate the synthetic data in our Monte Carlo experimental design and the main results obtained in the analysis. The fourth section is devoted to describe the proposed methods to tackle the endogeneity problem in empirical applications. The chapter concludes with a discussion of the main implications of our findings for practitioners using DEA to measure technical efficiency in different contexts, as well as with some directions for future research.

## 1.2 The endogeneity issue and its potential effects on DEA

### 1.2.1 The endogeneity issue

The analysis of data in the presence of endogeneity is one of the main recent contributions of econometrics to statistical science (Blundell and Powell, 2003). Consider the multiple-input single-output productive function:

$$y_i = f(x_i) + \epsilon_i \;\; i = 1, 2, ..., n \tag{1.1}$$

where $y_i$ is the level of observed output for DMU $i$, $f$ is an unknown production function to be estimated, $x_i \in \Re^m$ is the vector of observable inputs and $\epsilon_i$ represents the unobservable error term, which can also be identified as the distance to the true productive frontier. In fact, if we limit the estimation of $f$ to the non-stochastic frontier models, we can assume that all those deviations are due to technical inefficiency and therefore $\epsilon_i \leq 0 \;\; i = 1, 2, ..., n$.

In order to properly estimate the Equation 1.1 using a regression model, some crucial assumptions are required, including that the error term be uncorrelated with all the observed inputs $E(\epsilon|X) = 0$ ; *i.e.*, all regressors must be exogenous. In this context, the presence of endogeneity implies that $x_i$ and $\epsilon_i$ are correlated, thus the latter assumption cannot be hold in practice and $E(\epsilon|X) \neq 0$ . This phenomenon can arise as the result of multiple sources, such as measurement errors, unobserved heterogeneity or the omission of relevant variables in the model specification; although perhaps the most common cause is the presence of simultaneity or two-way causal relationships between the dependent and independent variables (Wooldridge, 2012). The idea behind this concept is that some inputs are not exogenous and are determined within the model.

The education sector is a good example to illustrate this issue (Mayston 2003), where the endogeneity problem is frequently observed. Actually, in this framework the presence of the self-selection problem has been argued to be the basis for multiple theoretical and empirical critiques of traditional findings using conventional econometric techniques and multiple methods have been developed in the literature to deal with this problem (Webbink 2005, Schlotter *et al.* 2011). For example, it is claimed that more motivated parents tend to devote more time and resources to choose the best schools (those with better peer group and academic outcomes) for their children than less motivated parents (see more example in Evans *et al.* 1992, Hoxby, 2000 or McEwan, 2003). But this parents' motivation, which is generally positive associated with families' socio-economic background at school level, is unobserved. As a result, groups of pupils from more advantaged backgrounds, and thus the school they attend, will tend to obtain better academic results for two reasons. Firstly, they have better average socio-economic level which is an essential input for producing educational output. Secondly, because these students are also more motivated and this fact positively affects school's efficiency. Consequently, we will observe that schools whose students come from a high socio-economic background are more prone to be fully efficient. To be fully efficient implies obtaining better results compared with other schools with similar inputs, so once again, these schools will attract more motivated parents reinforcing the endogeneity issue. This mechanism results in a positive correlation between the school's average socio-economic background (input) and technical efficiency: $E(\epsilon|X) \geq 0$

The same reasoning can be applied for the teacher's self-selection problem in many public education systems. Highly qualified and more motivated teachers tend to choose school first, self-selecting into smart schools with higher academic results, better facilities and students coming from higher income families. Again, this process derives in a positive correlation between the input level and the school efficiency. The intensity of this correlation will depend not only on the importance of parents, students or/and teachers motivation, but also on the correlation between these unobservable variables and the observed input (socio-economic level).

However, the endogeneity problem in the education sector can also arise in the opposite direction when a direct negative feedback from outputs to resources is observed (simultaneity). This applies for example when school funding systems operating compensatory policies allocate more resources to schools with poorer academic results in order to improve the performance of these schools (Orme and Smith 1996 and Levacic and Vignoles 2002). If poorer results are due to a high inefficiency, then the reverse causality problem implies allocating more resources to inefficient schools causing a negative correlation between resources (input) and the true efficiency: $E(\epsilon|X) \leq 0$.

The presence of correlation between inputs and technical efficiency can be also observed in many other production processes. For example, large firms can usually attract better managers (more qualified and motivated) as they can offer better salaries and conditions than small firms. As large firms use more inputs to produce outputs than small firms, again, one would also expect a positive correlation between the firm technical efficiency and the levels of input (Wilson 2003).

In short, endogeneity is a very frequent issue in production processes. It exceeds the scope of this work making an exhaustive analysis of all potential endogenous settings. Regardless of each specific endogeneity source, the target of this chapter is to address how this potential problem can affect DEA technical efficiency estimates.

### 1.2.2 The potential effects of endogeneity on DEA

In principle, it might seem that DEA should not be influenced by the presence of endogeneity, since it constructs a boundary around feasible combinations of inputs and outputs without assuming a parametric functional form (Orme and Smith 1996). However, if we apply insights from Kuosmanen and Johnson (2010) and interpret the DEA model as a constrained variant of the convex nonparametric least squares regression (Kuosmanen 2008), we can derive straightforward that the same problems of bias caused by the presence of endogeneity in econometrics explained above can also arise within this approach.

DEA is a mathematical programming approach that was originally proposed by Charnes, Cooper and Rhodes (1978) (DEA-CCR) and Banker, Charnes and Cooper (1984) (DEA-BCC) to measure the productive efficiency of a set of decision-making units (DMU) under constant and variable returns to scale respectively (CRS and VRS hereafter). The output-oriented problem under VRS can be specified using the following linear programming (LP) expression (multiplicative DEA efficiency measure):

$$\varphi_i = \max_{\lambda,\varphi}\{\varphi|\varphi y_{ri} \leq \sum_{i=1}^{n}\lambda_i y_i; x_k i \geq \sum_{i=1}^{n}\lambda_i x_1; \sum_{i=1}^{n}\lambda_i = 1; \lambda \geq 0 \;\; \forall i = 1,2,...,n\} \tag{1.2}$$

where $x_k$ denotes input $k$, $y_r$ stands for output $r$ and $i$ represents the production units. Multipliers $\lambda_i$ are referred as intensity weights of each DMU determined by the program solution. The technical efficiency score of the ith DMU is equal or greater than one, where $\hat{\varphi}_i = 1$ represents an efficient unit, whereas $\hat{\varphi}_i \geq 1$ indicates that the ith DMU is inefficient.

According to Banker (1993), the variable returns to scale DEA estimator of a production function f can be formally defined as:

$$f^{DEA}(x) = \max_{\lambda \in R^n_+} \{y | y = \sum_{i=1}^{n} \lambda_i y_i ; x \geq \sum_{i=1}^{n} \lambda_i x_i ; \sum_{i=1}^{n} \lambda_i = 1 ; \lambda \geq 0 \;\; \forall i = 1, 2, ..., n\} \tag{1.3}$$

Substituting f in the equation 1.1 by the DEA estimator, we can observe that the DEA efficiency scores for each unit can also be obtained as the optimal solution to the following LP problem (additive DEA efficiency measure):

$$\epsilon_i^{DEA} = \min_{\lambda, \epsilon} \{\epsilon | y = \sum_{i=1}^{n} \lambda_i y_i + \epsilon_i ; x \geq \sum_{i=1}^{n} \lambda_i x_i ; \sum_{i=1}^{n} \lambda_i = 1 ; \lambda \geq 0 \;\; \forall i = 1, 2, ..., n\} \tag{1.4}$$

Therefore, Kuosmanen and Johnson (2010) expose that the formulations 1.2 and 1.4 are equivalent in the single-output setting and thus: $\varphi_i^{DEA} = 1 - \frac{\epsilon_i^{DEA}}{y_i}$. In fact, the authors demonstrate in their work that the DEA problem can be interpreted as a nonparametric least-squares model under the assumption that $\epsilon_i \leq 0$ (Kuosmanen and Johnson 2010, p. 152). This connection between the nonparametric regressions and the mathematical programming approaches contributes to developing the statistical foundation of DEA. As a result, we can derive that DEA estimators will be consistent if all the assumptions in the least-squares regression model are fulfilled. However, in the case that $\epsilon_i$ is correlated with at least one input, the assumption $E(\epsilon | X) = 0$ does not hold and, therefore, efficiency estimates $\hat{\varphi}_i$ in Equation 1.2 can be biased. To better understand these ideas hereinafter we graphically illustrate this problem.

Figure 1.1 represents a single-input ($x$) / single-output ($y$) production setting in which true efficiency $\varphi_i$ is exogenously distributed, *i.e.*, $E(\epsilon | X) = 0$. In this scenario, the frontier estimated by DEA is very similar to the true one for the entire data range. Fully efficient DMUs are correctly identified, and efficiency is randomly spread along the production frontier.

However, as noted above, we may well find in real-world production processes some kind of correlation between the true efficiency and the level of input that is significantly different from zero: $E(\varphi | X) \neq 0$ . This correlation can be either positive or negative, as mentioned previously in the examples of different educational settings. Figure 1.2 illustrates the situation where endogeneity is positive, $E(\epsilon | X) > 0$.

In this case, although microeconomic theory establishes that the input level and the true efficiency are independently distributed, the existence of this positive endogeneity can break this assumption. According to Figure 1.2, DMUs with higher levels of input (and outputs), *e.g.*, dots C and D, are generally closer to the true frontier, whereas DMUs with lower input levels are less efficient. However, as DEA estimates efficiency scores are based on observed data, the frontier built by DEA will find and classify several DMUs that have low input level and are in fact highly inefficient as efficient. This is the case for dots A and B in Figure 1.2, which are actually far away from the true frontier but are identified by DEA as efficient units. Consequently, the frontier estimated by DEA will be far away from the true one in the lower input frontier region. This means that efficiency improvement targets will (incorrectly) be more demanding for observations with a higher input level than for those with a low input level. For example, while unit E is clearly closer than unit F to the true frontier in terms of output, both units appear to have a similar estimated technical efficiency because the actual production frontier is wrongly identified at low inputs levels. Since efficiency scores are relative measures, the misidentification of some DMUs distorts all efficiency estimates and the performance ranking. This result could have very important implications, particularly, if DEA is conducted for benchmarking and performance-based policy making.

On the other hand, the existence of a significant negative correlation between the input level and the true efficiency seems to just slightly affect DEA estimates. Figure 3 illustrates this context where in terms of the true production frontier more efficient units show low input levels and more inefficient units are those with high input levels.

It is worth to note, that at the region of high input level the estimated frontier shifts slightly down from the true one. But in this case the most inefficient DMUs (those with high input level) remain far enough away from the DEA estimated frontier to be still identified as the most inefficient producers compared with other DMUs. The main reason for conserving the high relative distance to the frontier for high input level DMUs is the monotonicity assumption. Monotonicity impedes DEA to pursue inefficient DMUs to drawing the estimated production frontier. In addition, the negative correlation between the input and the technical efficiency provides more information to DEA in order to correctly identify and estimate DMUs efficiency. The reason is that negative endogeneity reinforces the major microeconomic assumption behind the measurement of efficiency, *i.e.* for a constant output level, using higher level of inputs implies greater inefficiency. For instance, unit G is highly inefficient, and although the DEA frontier is closer to G that the true one, the distance between G and the estimated frontier is still large enough in terms of output to be recognized as one of the most inefficient producers. In general, inefficiency is correctly identified since all DMUs keep their relative position. Thus, it is expected that under the presence of negative endogeneity the efficiency scores estimated by DEA will better match the true relative positions, and thus the estimated ranking will not be significantly different from the true one. However, negative endogeneity could move upwards average efficiency scores because now the DEA frontier is closer to the most inefficient DMUs than the true one.

Now, these potential implications of the presence of different kinds of endogeneity have to be measured in quantitative terms. On this ground, we test DEA performance in finite samples using data from a Monte Carlo experiment by simulating endogeneity through a significant positive or negative correlation between the true technical efficiency and one input.

## 1.3 Monte Carlo experiment

### 1.3.1 MC experimental design and DGP

In order to illustrate the ideas developed above we perform a Monte Carlo experiment applied to seven scenarios. Firstly, we compute a baseline dataset without endogeneity (from now on, the *exogenous* scenario); then, six alternative settings were simulated taking into account the presence of correlation between the true efficiency ($\varphi_i$) and one observed input (from now on, the *endogenous* scenarios). Results from each endogenous scenario are then compared to the baseline one in order to measure the effects that endogeneity introduces on DEA estimations. All datasets were defined in a single output framework with three inputs. The first decision to be made in the DGP in order to carry out the experimental design was to choose the functional form for the production function.

#### 1.3.1.1 The production function

Almost all previous studies in the literature have simulated data using the Cobb-Douglas production function. Hence, in order to obtain comparable results we also draw data from a Cobb-Douglas with a single output and three inputs:

$$ln\ y_i = \alpha_1\ ln\ x_{1i} + \alpha_2\ ln\ x_{2i} + \alpha_3\ ln\ x_{3i} \tag{1.5}$$

where $y_i$ represents the output, and $x_1$ $x_2$ and $x_3$ are the observed inputs. The inputs weights assigned in this work where $\alpha_1 = 0.3$, $\alpha_2 = 0.35$ and $\alpha_3 = 0.35$, assuming constant returns to scale[4]. Although this functional form is the most commonly used in economics and operational research, it involves a significant drawback represented by the assumption of constant input-output elasticities. This means that no matter the scale of production, the marginal effects of inputs on outputs are the same; so it fails to capture potential non linear effects of those resources. Since the main aim of this work is testing the accuracy of DEA in an experimental setting that reproduces a more realistic context, we carried out our experimental design also considering a more flexible technology, the *Translog* production function introduced by Christensen *et al.* (1971).

$$ln\ y_i = \beta_0 + \sum_{k=1}^{K} \beta_k\ ln\ x_{ki} + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{K} \beta_{kj}\ ln\ x_{ki}\ ln\ x_{ji} \tag{1.6}$$

where $y$ denotes the output and $x_k$ $(k = 1, 2, 3)$ are the three inputs. We assume $\beta_0 = 3.5; \beta_1 = 0.5; \beta_2 = 0.3; \beta_3 = 0.5; \beta_{11} = -0.1; \beta_{22} = -0.05; \beta_{33} = -0.1; \beta_{12} = \beta_{13} = \beta_{23} = 0.01$. These parameters were defined in order to obtain a well-behaved production function within the bounds imposed by the inputs distribution that are uniformly distributed over the interval [5, 50]. Therefore, after having generated the data we checked for two desirable conditions at each simulated data point.

[4]Similar results were obtained using increasing returns to scale and decreasing returns to scale.

Firstly, we verify the monotonicity condition, where in a single output case requires that all marginal products must be non-negative $\partial y/\partial x_k \leq 0$. For the *Translog* production function this implies:

$$\partial y/\partial x_k = \frac{y}{x_k}.\frac{\partial\ ln\ y}{\partial\ ln\ x_k} = \frac{y}{x_k}.\{\beta_k + \sum_{j=1}^{K}\beta_{kj}\ ln\ x_j\} = \frac{y}{x_k}.s_k \geq 0\ \ \forall k \tag{1.7}$$

where $y/x_k$ is the average product and $s_k$ is the elasticity of $y$ with respect to $x_k$. As the average product $y/x_k$ is always positive, monotonicity implies that all input-output elasticities $s_k$ must be non-negative for all DMUs across all inputs range.

Secondly, we test for concavity in all inputs, which implies that all marginal products apart from being non-negative must be decreasing in inputs, *i.e.* the law of diminishing marginal productivity must be fulfilled (Coelli *et. al*, 2005). For the *Translog* production function this implies that all inputs must satisfy the following expression throughout all simulated data range:

$$\partial^2 y/\partial^2 x_k = \frac{y}{x_k^2}[\beta_{kk} + (\beta_k - 1 + \sum_{j=1}^{K}\beta_{kj}\ ln\ x_j)(\beta_k + \sum_{j=1}^{K}\beta_{kj}\ ln\ x_j)] = \frac{y}{x_k^2}[\beta_{kk} + (s_k - 1)s_k] < 0\ \ \forall k \tag{1.8}$$

Finally, the selected parameters and the distribution of inputs define the production scale elasticity. We perform the simulation assuming decreasing returns to scale (DRS), where scale elasticity ranges from 0.56 to 0.97, with a mean value of 0.69. These results are consistent with most complex production processes that take place in the public sector. For example, in the field of education if the initial endowments of all school inputs are doubled, it would be reasonable to expect an increase in students' test scores but in a less proportion than double, particularly at high levels of educational achievements (Essid *et.al.*, 2013).

#### 1.3.1.2 Data Generation Process

The baseline scenario represents the exogenous case, where all inputs are uncorrelated with the true technical efficiency, and it is simulated using the following procedure:

1. Generate randomly and independently three input vectors $x_1$, $x_2$ and $x_3$ using a uniform distribution over the interval [5, 50] for $N$ DMUs, $n = 1, 2, ..., N$.

2. Calculate the efficient level of output as $y_i = exp(lny_i)$ using $ln\ y_i = f(.)$, where $f(.)$ is Equation 1.5 or Equation 1.6 respectively.

3. Draw a random error term $v_i$ from a $N(0; 0.04)$, representing the random statistical perturbation in the production function. Since the main aim of this research is to test the performance of DEA under endogeneity, we do not simulate different magnitudes of the random shocks. As it has been demonstrated in previous studies, the larger the measurement error the poorer the performance of DEA (*e.g.* Bifulco and Bretshneider, 2001). Therefore, we choose a small measurement error in order to have, as in the real world, some noise but not so large that it distorts the analysis of endogeneity.

4. Randomly and independently generate N values of $u_i$ using a half-normal distribution $u_i \sim |N(0; 0.25)|$ and compute the vector $\varphi_i = exp(u_i)$. Then, compute the true technical efficiency level for each DMU $0 \leq \theta_i = \dfrac{1}{\varphi_i} \leq 1$.

5. Compute the observed output as: $\hat{y_i} = y_i.exp(v_i).\theta_i$.

The remaining six scenarios were developed through a similar DGP, but taking into account the existence of endogeneity, which was modelled through the Pearson's correlation coefficient between the true technical efficiency $\theta_i$ and one observed input. Therefore, in each dataset we substitute the exogenous input $x_3$ by an endogenous input $E$. In order to compute the latter with the same distribution as the exogenous inputs $(x_1, x_2, x_3)$ and with a specific level of correlation with $\theta_i$, for each endogenous scenario we follow this procedure:

1. Select the desired Pearson's correlation coefficient $\rho_{E,\theta}$ between the endogenous input $E$ and the true technical efficiency $\theta$.

2. Draw a random matrix $A = (a_1, a_2)$ from a multivariate normal distribution $N(0; \Sigma)$, where $\Sigma = \begin{bmatrix} 1 & \rho_{E,\theta} \\ \rho_{E,\theta} & 1 \end{bmatrix}$.

3. Compute an identification number variable (ID) from 1 to N.

4. Match the ID with the vector $a_1$ obtaining: $B = [ID \;\; a_1]$. Sort B by $a_1$ in an ascending order (the ID variable will be unsorted): $B' = [ID_{a1} \;\; a_1]$.

5. Generate an independent vector $x_{n\times1}$ from a uniform distribution over the interval [5, 50] and sort it in an ascending order obtaining $x_S$.

6. Compute a new C matrix by merging $B'$ with $x_s$: $C = [ID_{a1} \;\; a_1 \;\; x_s]$.

7. Sort C by the ID variable in an ascending order: $C' = [ID \;\; a_{1,ID} \;\; x_{ID}]$.

8. The latter vector of $C'$, $(x_{ID})$, will be defined as the endogenous input, $E = x_{ID}$.

9. Match ID with the vector $a_2$ obtaining: $D = [ID \;\; a_2]$. Sort D by $a_2$ in a descending order (the ID variable will be unsorted): $D' = [ID_{a2} \;\; a_2]$.

10. Randomly and independently generate N values of $u_i$ using a half-normal distribution $u_i \sim |N(0; 0.25)|$. Then, compute the vector $\varphi_i = exp(u_i)$ and sort this variable in an ascending order obtaining $\varphi_s$.

11. Compute a new H matrix by merging $D'$ with $\varphi_s$: $H = [ID_{a2} \;\; a_2 \;\; \varphi_s]$.

12. Sort H by the ID variable in an ascending order: $H' = [ID \;\; a_{2,ID} \;\; \varphi_{ID}]$.

13. The latter vector of $H'$ $(\varphi_{ID})$, is used to computed the true technical efficiency level for each unit, $\theta_i = 1/\varphi_{ID}$. The generated average true efficiency in each experiment ranges from 0.828 to 0.859 with standard deviations values between 0.097 and 0.116.

14. Using the exogenous inputs $x_1$ and $x_2$ generated in the baseline scenario and the endogenous input $E$, compute the efficient level of output as $y_{iend} = exp(lny_{iend})$ using $lny_{iend} = f(.)$, where $f(.)$ is Equation 1.5 or Equation 1.6 respectively.

15. Finally, calculate the observed output using the random term $v_i$ computed in the baseline dataset and the true efficiency level $\theta_i$ computed in step 13: $\hat{y}_{iend} = y_{iend}.exp(v_i).\theta_i$.

Two factors were allowed to vary in order to generate the six endogenous settings: the sign (negative or positive) and the intensity (high, medium or weak) of the correlation coefficient between the true efficiency and the endogenous inputs ($\rho_{\theta,E}$). Table 1.1 summarizes the main descriptive statistics of the correlation coefficients that have actually been obtained in each simulated scenario.

All scenarios were replicated using the Cobb-Douglas and the Translog production functions for a sample size of 100 DMUs[5]. Finally, for each dataset we estimate the efficiency scores $\hat{\theta}_i$, by running an output oriented DEA model under CRS and VRS. As a result, 28 scenarios were analysed (the exogenous scenario, six types of endogeneity with different intensities and signs, two production technologies and two types of return to scale). In order to make the results more reliable, we undertake a Monte Carlo experiment where $B$, the number of replicates is 1,000; consequently, all measures are computed in each replication and finally averaged to obtain the results presented in the next section[6].

### 1.3.2 MC experiment results

#### 1.3.2.1 Accuracy measures

In order to test the adequacy of DEA under endogeneity in finite samples we present a set of accuracy measures. Firstly, we are interested in measuring the ability of DEA to correctly rank observations. For this purpose, we compute Spearman's rho ($r_s$) correlation coefficients between the true efficiency and estimated scores pairs:

$$r_s = 1 - \frac{6 \sum\limits_{i=1}^{N} d_i^2}{n(n^2-1)} \tag{1.9}$$

where $d_i = rank(\hat{\theta}_i) - rank(\theta_i)$ is the difference between the rank assigned to the i-unit according to DEA estimations and the place that i-unit actually has when we rank observations by the true efficiency value (in an ascending order). The higher the correlation coefficient $r_s$, the better the ability of DEA to identify the true efficiency distribution. The first two columns of Table 1.2 contain these coefficients for DEA-CRS and DEA-VRS models under different

[5]We replicated the analysis for sample sizes 40 and 300 and results did not change significantly. Results are available under request.

[6]Simulations were carried out using MATLAB 7.9.0 software.

endogenous scenarios compared to the exogenous baseline assuming data coming from a Cobb-Douglas production function. Table 1.3 contains the analogous results for a *Translog* DGP.

Secondly, we are interested in testing the capability of the method to estimate the true level of efficiency. For this purpose, we average the estimated efficiency scores (mean estimated efficiency) to compare this value with the true mean efficiency. If the former is larger (smaller) than the second, DEA overestimates (underestimates) the true efficiency level. Finally, we also calculate the Mean Absolute Error (MAE).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{\theta}_i - \theta_i| \tag{1.10}$$

The MAE arises from computing the sum of absolute deviations of DEA estimated scores from the true efficiency level for each observation and averaging them. A low MAE implies that, on average, the estimates are near the true efficiency values; therefore, small values are preferred. All Monte Carlo results are provided in Table 1.2 and Table 1.3 for data generated from the Cobb-Douglas and the *Translog* production technology respectively; and under both CRS and VRS assumed.

Finally, following Bifulco and Bretschneider (2001), we present a performance measure based on a quintile analysis. Observations were first divided into quintiles according to their true efficiency score, and then we examined the ability of the technique to place observations in the appropriate quintile. This complementary measure allows us to evaluate the technique accuracy at different points of the distribution and hence, it is a helpful tool to locate the main drawbacks of the technique. For example, if the objective of the research is to identify the best practices, we will be especially interested in the percentage of top quintile observations that DEA assign correctly to the top quintile rather than in the overall ranking accuracy. Results from this analysis are presented in Table 1.4 and Table 1.5 for the Cobb-Douglas and the *Translog* DGP respectively, under CRS and VRS assumptions.

#### 1.3.2.2 Baseline scenario results

The results confirm that DEA performs reasonably well in the exogenous case regardless the production function or the returns to scale assumed. These findings are similar to those of previous studies (Bifulco and Bretschneider, 2001; Ruggiero, 2003; Krüger, 2012). However, as expected, CRS-DEA estimates outperform VRS-DEA when data was generated with a Cobb-Douglas production function, and vice versa for data derived from the *Translog* one. For example, under the Cobb-Douglas technology Spearman's correlation coefficients between the true and estimated efficiency is 0.78 under CRS and 0.67 under VRS; conversely for the *Translog* these figures are 0.67 and 0.73 respectively. This finding highlights the importance of making a correct choice of the returns to scale assumed before conducting a DEA efficiency analysis. Given this evidence, hereafter we will refer to DEA-CRS for the results estimated from the Cobb-Douglas under CRS and to DEA-VRS for the results estimated from the Translog scenarios under VRS.

Results from Table 1.4 and Table 1.5 also confirm the accurate performance of DEA in the exogenous case. Almost 50% of observations are placed in the correct quintile and about only one of every eight units is placed two or more quintiles away from the right one. From columns 5 to 12 it can be concluded that the major weakness of DEA lies in the ability to correctly identify the most efficient DMUs. While about three quarters of the most inefficient units are correctly assigned to the bottom quintile, this proportion drops to around 50% for units properly identified in the top quintile. Moreover, the percentage of units placed in the bottom quintile being actually in the first two is near to zero in the exogenous case; but this figure arises to 7% and to 12.3% for observations in the bottom quintile assigned to the top one under CRS and VRS respectively. This evidence should be taken into account specially if DEA is conducted with the purpose of performance-based reforms, for there would be some units identified as benchmarks when they actually are not.

#### 1.3.2.3 Endogeneity effects on DEA

The accuracy of DEA under endogeneity depends on the direction and intensity of the correlation between the endogenous input and the true technical efficiency. However, the overall effects on DEA-CRS and DEA-VRS estimates are similar, being on the latter more pronounced than on the former. For instance, in the baseline scenario the Spearman's correlation coefficient between the true efficiency and DEA estimates is 0.778 under CRS and 0.729 VRS; while when high and positive endogeneity is introduced these correlations fall down to 0.52 and 0.342 respectively. This performance can be explained by the fact that under VRS the technique is more sensitive to changes in the distribution of data than under the CRS assumption. Given that VRS is a more realistic and frequent assumption in real world applications, and that conclusions are similar under both type of returns to scale, hereafter we will comment more in depth results only for DEA-VRS (Table 1.3 and Table 1.5)[7].

The main finding that arises from our simulations is that positive and high endogeneity is the worst possible scenario, shattering DEA performance. As the intensity of this positive endogeneity decreases to medium, DEA improves its results and the errors mitigate progressively to the extent that for the case of only low positive endogeneity DEA estimations are very close to those in the baseline scenario. As Table 1.3 shows, the exogenous dataset simulations yield a Spearman's correlation coefficient of 0.729 between estimated and actual efficiency, which is reduced to 0.342 in the presence of high and positive endogeneity and to 0.612 in the case of medium positive endogeneity. The MAE remarks this result, which in the positive and high endogeneity scenario reaches a 0.116 value, significantly higher than 0.072 calculated in the exogenous baseline one. Another way to observe the effects over the estimated efficiency level is throughout the average estimated efficiency (column 4 of Table 1.3). It reveals that under both types of endogeneity (negative and positive) DEA overestimates the true mean technical efficiency, particularly when the input and the efficiency are highly correlated.

[7] The results are presented also under CRS for the most interested readers.

An alternative approach to evaluate the damage caused by the endogeneity is through the proportion of units assigned to the correct quintile by DEA. According to Table1.5 the DMUs correctly assigned to their quintile falls from almost 47% in the exogenous setting to 28% (40%) under high (medium) and positive endogeneity. Additionally, the proportion of units assigned two or more quintiles away from the correct one almost triples the baseline percentage (from 13% to 35% respectively) in the case of high positive endogeneity. The quintile analysis allows us to note that the decline in DEA performance is further driven by the fact that under positive and high endogeneity, the technique identifies as efficient several units that actually are some of the most inefficient ones. Only 40% of units assigned by DEA to bottom quintile were actually in the bottom quintile when high and positive correlation between true efficiency and one input is observed while in the exogenous scenario this percentage reaches 75%. In addition, the proportion of DMUs placed in the top quintile but which were actually in the two last quintiles is almost tripled compared to the baseline scenario. These results confirm what it was discussed in Figure 1.2: at low levels of the endogenous input the estimated frontier by DEA (which is driven by the data shape), is located further from the true one, identifying as very efficient such many inefficient units.

As we have exposed earlier DEA efficiency scores are relative measures, therefore the misidentification of the true frontier at low levels of input leads to inaccurate estimated scores for all observations. This implies that the ability of DEA to correctly identify the most efficient DMUs is also deteriorated under such endogeneity. For instance, the proportion of units properly assigned to the top quintile drops from 47% to 33%. Furthermore, while in the absence of endogeneity we cannot find units assigned to the bottom quintile that are actually ranked in the two first ones; under high positive endogeneity we observe that this happens for a 8% of DMUs.

Finally, under negative endogeneity Monte Carlo simulations evidence that DEA estimates remain robust. Only in the scenario where negative endogeneity is high, estimations seem to be slightly damaged. These results are similar to those obtained by Bifulco and Bretscneider (2003) and Ruggiero (2003) where they conclude that for the same measurement error of our simulation, the performance of DEA does not change substantially under negative endogeneity. This finding can be explained by the fact that the negative endogeneity correlates the input and the efficiency in the same way that DEA assumes to construct the frontier (*i.e.* the higher input level, the lower technical efficiency). In other words, endogeneity in this case reinforces the microeconomic assumption behind the DEA program, and therefore, estimates are unaffected by endogeneity.

In summary, our results allow us to conclude that DEA-CRS and DEA-VRS provide accurate efficiency measures in all scenarios except when there is a medium or high positive correlation between one input and the true efficiency. It should be highlighted again that DEA estimates will be far away from the actual efficiency values in the presence of a high positive endogeneity regardless of the assumed functional form. This is a very remarkable result since those endogenous scenarios are similar to those that are likely to be found in public sector efficiency analysis applications (due to a two-way causality or an omitted variable) and specially in sectors like education where school choice plays an important role. Therefore, this evidence suggests that in those cases, the estimation of the technical efficiency using DEA models, without taking into account the presence of endogeneity, could lead to misleading efficiency estimates; and thus inappropriate performance-based recommendations.

Drawing on these findings, two key issues arise now: how can we detect the presence of an endogenous input? And, how can we deal with this problem in DEA empirical applications to overcome this problem and improve estimations?

## 1.4 Dealing with the endogeneity in DEA estimations

In this section, we propose a simple heuristic method which allows practitioners to identify the presence of an endogenous input in an empirical research. In addition, we propose a potential solution to deal with this problem in order to improve DEA estimations: an '*Instrumental Input DEA*' strategy (II-DEA from now on). We evaluate the performance of both proposals in finite samples problems using synthetic data generated in a Monte Carlo experiment as in the previous section.

### 1.4.1 How to identify the endogeneity problem?

In this section we propose a simple heuristic method to identify the presence of an endogenous input in a DEA application. From the Monte Carlo experiment we observe that the distribution of the correlation coefficients between the inputs and the estimated efficiency scores $\hat{\theta}_i$ considerably differ in each simulated scenario (Figure 1.4 and Figure 1.5). From a microeconomic viewpoint and assuming that inputs are exogenous, the correlation coefficient between the inputs and the DEA estimated efficiency scores should be slightly negative and close to zero (or at least non-positive), as DEA assumes that for a given output, the higher input level the higher inefficiency. Then, our proposed heuristic method is based on these expected correlation coefficients in order to classify the nature of each input included in the DEA model. In practice, we proceed in six steps as follows:

1. From the empirical dataset $\chi = \{(X_i, Y_i)\ i = 1, ..., n\}$ randomly draw with replacement a bootstrap sample B=1,000 $\chi_b^* = \{(X_{ib}^*, Y_{ib}^*)\ i = 1, ..., n\}$

2. Compute the efficiency scores $\hat{\theta_{ib}^*} = \dfrac{1}{\varphi_{ib}^*} \leq 1\ \ i = 1, ..., n$ using the DEA-VRS LP

$$\varphi_i^* = \max_{\lambda,\varphi}\{\varphi|\varphi y_{ri} \leq \sum_{i=1}^{n} \lambda_i y_i; x_k i \geq \sum_{i=1}^{n} \lambda_i x_1; \sum_{i=1}^{n} \lambda_i = 1; \lambda \geq 0\ \ \forall i = 1, 2, ..., n\} \tag{1.11}$$

3. For each input $k = 1, ..., p$ compute the Pearson's correlation coefficient between the estimated efficiency score $\hat{\theta_{ib}^*}$ and the input $k$ $\rho_{kb}^* = corr(x_{ik}^*, \hat{\theta_i^*})\ \ i = 1, ..., n\ \ k = 1, ..., p$

4. Repeat steps 1-3 B=1,000 times in order to obtain for $k = 1, ..., p$ a set of correlations: $\{\rho_{kb}^*,\ \ b = 1, ..., B\}$

5. For each input $k$ compute $\gamma_k^* = \dfrac{1}{B}\sum_{b=1}^{B}[I_{[0,1]}(\rho_k^*)]_b \;\; for \;\; k = 1, ..., p$ where $I_{[0,1]}(\rho_k^*)$ is the Indicator Function defined by:

$$I_{[0,1]}(\rho_k^*) = \begin{cases} 1, & \text{if } 0 \leq \rho_k^* \leq 1; \\ 0, & \text{otherwise.} \end{cases} \tag{1.12}$$

6. Finally, classify each input using the following criterion:
   - If $\gamma_k^* < 0.25 \rightarrow$ Exogenous/Negative endogenous input $k$
   - If $0.25 \leq \gamma_k^* < 0.5 \rightarrow$ Positive LOW endogenous input $k$
   - If $0.5 \leq \gamma_k^* < 0.75 \rightarrow$ Positive MIDDLE endogenous input $k$
   - If $\gamma_k^* \geq 0.75 \rightarrow$ Positive HIGH endogenous input $k$

### 1.4.2 The *Instrumental Input DEA* strategy

In order to improve DEA estimates under the presence of a positive and significant correlation between one input and the true efficiency we propose a semi-parametric strategy that introduces the well-known Instrumental Variables (IV) approach (*e.g.* see Greene, 2003 or Wooldridge, 2012) into the conventional DEA model specification, which we call '*Instrumental Input DEA*'. The intuitive idea behind this proposal is the same as in the IV strategy, to include in the DEA specification only the exogenous part of the endogenous input. To do this, we propose replacing the endogenous input by an exogenous variable, which only contains the exogenous information of the original one, that is, that part which is uncorrelated with the technical efficiency.

Consider the single-output multi-input productive dataset $\chi = \{(X_i, Y_i) \; i = 1, ..., n\}$, where one input is significantly positive correlated with the true efficiency term (hereafter the endogenous input $E$). As in the classic IV approach, the first step is to find a good instrumental input $G$ which must satisfies at the same time two basic conditions:

i. Relevance: the instrument G must be significantly correlated with the endogenous input E, *i.e.* $E(E|G) \neq 0$;

ii. Exogeneity: the instrument G must be uncorrelated with the true efficiency term, *i.e.* $E(\varphi|G) = 0$

The first condition can be contrasted in empirical applications by testing the significance of the parameter $\tau$ in the following estimated regression $E = \alpha + \tau G + \xi$. If we do not reject$H_0 : \tau = 0$, we can assume that the instrument is relevant. However, the second condition cannot be directly tested because true efficiency is not observed in empirical settings. In this case, the exogeneity condition can be interpreted as the absence of a causal relationship between the instrumental input $G$ and the output variable Y. That is, $G$ should have no partial effect on Y (beyond the effect through the endogenous input). As Wooldridge sets '*...we must maintain this* *condition by appealing to economic behavior or introspection*' (Wooldridge, 2012 p.514). The II-DEA procedure is implemented following two simple steps:

1. The aim of the first step is to isolate the exogenous component of the endogenous input that is uncorrelated with the true efficiency. To do this, regress the endogenous input ($E$) over the instrumental input ($G$) and the rest of the exogenous inputs

$$E = \alpha + \delta_1 x_1 + ... + \delta_{k-1} x_{k-1} + \phi G + \mu_i \tag{1.13}$$

   where $x_{k-1}$ are the k-1 exogenous inputs, G is the instrumental input and $\mu_i$ is a random white noise component.

2. Secondly, in order to obtain the corrected DEA efficiency scores for each DMU replace the endogenous input ($E$) by the estimated exogenous variable $\hat{E}_i$ in the conventional DEA linear program 1.2.

### 1.4.3 Monte Carlo results

We test the performance of the II-DEA strategy to control for the presence of endogeneity in finite sample problems. To do this, we reproduce the experimental design presented in the previous section but, in this case, we additionally generate a new variable: an instrumental input G. Like the remaining inputs, G is uniformly distributed $U[5, 50]$ and is uncorrelated with the true efficiency level $E(\varphi|G) = 0$ and moderately correlated with the endogenous input $E(E|G) \approx 0.25$[8].

After generating the dataset we estimate the efficiency scores using the conventional DEA and the II-DEA model proposed in order to compare their performance. Results from the Monte Carlo simulations are reported in Table 1.6 and Table 1.8 for the Cobb Douglas DGP and in Table 1.7 and Table 1.9 for the *Translog* technology.

The first conclusion is that under both specifications results do not show significant differences. For this reason, hereafter we will only discuss the results for DEA and II-DEA under VRS from the *Translog* DGP, because it is a more realistic assumption in real educational applications[9].

From the results showed in Table 1.7 we find that in the worst scenario, *i.e.* under high and positive correlation between efficiency and one input, the II-DEA model outperforms the conventional DEA not only in terms of the Spearman's correlation coefficient but also in terms of MAE. In fact, in terms of Spearman's correlation the results from the II-DEA are very similar to those observed from the conventional DEA in the baseline scenario under the exogeneity assumption. However, in the case of the MAE the proposed method outperforms the results of

[8] In real data we seldom find instruments with greater correlation, previous research found similar correlations (Wooldridge 2012, pp. 519-520).

[9] However, results are also reported for the Cobb-Douglas DGP under CRS.

the conventional DEA in the presence of high positive correlation, but it shows not as good results as under the assumption of exogeneity. On the other hand, we confirm that conventional DEA is robust under the presence of a negative correlation or a low positive one, and instrumenting the endogenous inputs in these cases conducts to considerably detrimental results.

Following the previous Monte Carlo simulations, now we also test the ability of the proposed II-DEA method to place observations in the appropriate quintile in order to know what are the main improvements of the technique over the situation in which endogeneity is not corrected. As we have mentioned before, it only has sense to apply the II-DEA when conventional DEA estimations are damaged, therefore, hereafter we only discuss the results for the scenarios where $\rho = 0.4$ and $\rho = 0.8$.

The results show that the outperformance of the II-DEA method in the worst scenario, *i.e.* when $\rho = 0.8$, is further driven by its ability to correctly identify the most inefficient units. In this case, the percentage of units correctly assigned to the bottom quintile considerable increases from 41% to 76% under the VRS assumption. Recall that in this scenario, the most inefficient units are those with low input level, so this finding confirms that the proposed method can deal with the misidentification of the true frontier at this region. In addition, the percentage of units actually in the two last quintiles wrongly assigned to the top one is halved when II-DEA is applied instead of the conventional DEA method. As the DEA estimates are relative measures, this improvement also affects the technique ability to correctly identify the most efficient units. In this sense, we observe a substantial reduction in the percentage of efficient units assigned to the bottom quintile by the II-DEA under significant endogeneity, which drops from 8% to almost zero.

Under the assumption of a moderate correlation between the true efficiency and one input ($\rho = 0.4$) both techniques DEA and II-DEA show similar results, which implies that the proposed method is not powerful enough to overcome the damage caused by such endogeneity at low input levels. In this case, the exogenous part of the endogenous input that is included in the II-DEA specification through $\hat{E}_i$ does not seem to provide enough information to correctly identify the true frontier at low inputs levels. From these results it seems that the final decision about whether or not to instrument the endogenous input in the case of a moderate positive correlation will depend on the empirical application aims. In terms of Spearman's correlation the II-DEA dominates the conventional DEA and *vice versa*, the latter dominates in terms of MAE. For instance, in many educational applications the main purpose is focused on benchmarking schools and then analysing which are the main drivers of efficiency, rather than on correctly estimate the true mean efficiency. In these cases, it would be preferable to apply the II-DEA.

## 1.5 Concluding remarks

Endogeneity, and the distortions that it causes on the estimation of economic models, is a usual problem in the econometrics literature. As a result, some empirical research is starting to apply conventional econometric approaches to deal with this problem in the estimation of production frontiers using parametric techniques. However, the effects of endogeneity on efficiency estimates obtained with nonparametric methods like DEA have received less attention in the literature so far.

In this chapter we analyze to which extent can the presence of endogeneity in the production process affect DEA estimations in finite samples. For this purpose, we simulate different intensities of negative and positive endogeneity through the correlation between one input and the true efficiency using synthetic data generated in a Monte Carlo experiment. In line with previous studies, we find that DEA is robust to the presence of negative endogeneity. However, a significant positive endogeneity, *i.e.* a significant positive correlation between one input and the true efficiency level, severely biases DEA performance. In addition, we find that this decline in DEA performance is further driven by the misidentification of the most inefficient DMUs with low levels of the endogenous input. These findings take greater significance since high positive endogenous scenarios are similar to those that are likely to be found in public sector production processes (due to a two-way causality or an omitted variable problem) and specially in sectors like education. In this context, the estimation of the technical efficiency using DEA models without taking into account the presence of endogeneity leads to inaccurate efficiency estimates. The main reason behind this result is that many of the most inefficient DMUs are identified as benchmarks, which will lead to inappropriate performance-based recommendations.

In this sense, we propose a simple heuristic method which allows practitioners to identify potential endogenous inputs in empirical research. In addition, getting insights from econometrics, we provide a potential solution to deal with this problem in order to improve DEA estimations: the '*Instrumental Input DEA*' strategy. Monte Carlo simulations show that the II-DEA outperforms the conventional DEA model when one input exhibits a high and positive correlation with efficiency. Furthermore, we can conclude that the II-DEA strategy can deal with the misidentification of the true frontier at low inputs levels and hence, it can correctly identify the most inefficient units located in this frontier region.

To summarize, this study provides new insights about a major concern in economics and alerts DEA practitioners about the accuracy of their estimates when they suspect that there might be some significant positive endogeneity in their data, providing potential solutions to deal with this problem in empirical applications. More research is still needed in different directions but that exceeds the scope of the present work. Although the experimental Monte Carlo design tries to replicate a general production setting and is in line with several previous studies, results must be cautiously interpreted as they depend on the parameters and functional forms assumed and cannot be generalized to all contexts. In this sense, we think that deriving the asymptotic properties of the proposed II-DEA estimator, extending the analysis to multi-output settings and researching how other nonparametric efficiency techniques (Free Disposal Hull, order-m, order-alpha, total factor productivity indexes based on DEA, conditional efficiency models and so on) can be affected by the endogeneity could be three of the most promising contributions.

## 1.6 References


Banker, R. D. (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. Management Science, 39(10), 1265-1273.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-1092.

Banker, R. D., Gadh, V. M., and Gorr, W. L. (1993). A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis. European Journal of Operational Research, 67(3), 332-343.

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

Bifulco, R., and Bretschneider, S. (2003). Response to comment on estimating school efficiency. Economics of Education Review, 22(6), 635-638.

Blundell, R., and Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, Hansen, L.P. and Turnovsky, S. (Ed.), Advances in Economics and Econometrics: Theory and Applications: Eight World Congress (pp. 312-357): Cambridge University Press.

Bojanic, A. N., Caudill , S. B., and Ford, J. M. (1998). Small Sample Properties of Ml, Cols and Dea Estimators of Frontier Models in the Presence of Heteroscedasticity European Journal of Operational Research, 136(2 ), 468-469.

Bowlin, W. F. (1986). Evaluating performance in governmental organizations. The Government Accountants Journal, 35(2), 50-57.

Bravo-Ureta, B. E., Greene, W., and Solís, D. (2012). Technical efficiency analysis correcting for biases from observed and unobserved variables: an application to a natural resource management project. Empirical Economics, 43(1), 55-72.

Coelli, T., Rao, D., O'Donnell, C., and Battese, G. (2005). An Introduction to Efficiency and Productivity Analysis: Springer, New York.

Crespo-Cebada, E., Pedraja-Chaparro, F., and Santín, D. (2013). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. Journal of Productivity Analysis, 1-20. doi: 10.1007/s11123-013-0338-y

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.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1971). Conjugate duality and the transcendental logarithmic function Econornetrica 39 (4): 255-6: July.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., and Shale, E. A. (2001). Pitfalls and protocols in DEA. European Journal of Operational Research, 132(2), 245-259.

Essid, H., Ouellette, P., and Vigeant, S. (2013). Small is not that beautiful after all: Measuring the scale efficiency of Tunisian High Schools using a DEA-bootstrap method. Applied Economics, 45(9), 1109-1120.

Evans, W. N., Oates, W. E., and Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. Journal of Political Economy, 966-991.

Galagedera, D., and Silvapulle, P. (2003). Experimental evidence on robustness of data envelopment analysis. Journal of the Operational Research Society, 54(6), 654-660.

Gattoufi, S., Oral, M., Kumar, A., and Reisman, A. (2004). Content analysis of data envelopment analysis literature and its comparison with that of other OR/MS fields. Journal of the Operational Research Society, 55(9), 911-935.

Gijbels, I., Mammen, E., Park, B. U., and Simar, L. (1999). On estimation of monotone and concave frontier functions. Journal of the American Statistical Association, 94(445), 220-228.

Gong, B.-H., and Sickles, R. C. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. Journal of econometrics, 51(1), 259-284.

Greene, W. (2003). Econometric Analysis. New Jersey, Prentice Hall.

Greene, W. (2010). A stochastic frontier model with correction for sample selection. Journal of Productivity Analysis, 34(1), 15-24.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation (No. w7867). National Bureau of Economic Research.

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(06), 783-793.

Kneip, A., Simar, L., and Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. Econometric Theory, 24(06), 1663-1697.

Kneip, A., Simar, L., and Wilson, P. W. (2011). A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. Computational Economics, 38(4), 483-515.

Korostelev, A., Simar, L., and Tsybakov, A. (1995). Efficient estimation of monotone boundaries. The Annals of Statistics, 476-489.

Krüger, J. J. (2012). A Monte Carlo study of old and new frontier methods for efficiency measurement. European Journal of Operational Research, 222(1), 137-148.

Kuosmanen, T. (2008). Representation theorem for convex nonparametric least squares. The Econometrics Journal, 11(2), 308-325.

Kuosmanen, T., and Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. Operations Research, 58(1), 149-160.

Levačić, R., and Vignoles, A. (2002). Researching the links between school resources and student outcomes in the UK: A review of issues and evidence. Education Economics, 10(3), 313-331.

Mayen, C. D., Balagtas, J. V., and Alexander, C. E. (2010). Technology adoption and technical efficiency: organic and conventional dairy farms in the United States. American Journal of Agricultural Economics, 92(1), 181-195.

Mayston, D. (2003). Measuring and managing educational performance. Journal of the Operational Research Society, 54(7), 679-691.

McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile. Economics of Education Review, 22(2), 131-141.

Morita, H., and Avkiran, N. K. (2009). Selecting inputs and outputs in data envelopment analysis by designing statistical experiments. Journal of the Operations Research Society of Japan, 52(2), 163.

Nataraja, N. R., and Johnson, A. L. (2011). Guidelines for using variable selection techniques in data envelopment analysis. European Journal of Operational Research, 215(3), 662-669.

Orme, C., and Smith, P. (1996). The potential for endogeneity bias in data envelopment analysis. Journal of the Operational Research Society, 73-83.

Pedraja-Chaparro, F., Salinas-Jimenez, J., and Smith, P. (1999). On the quality of the data envelopment analysis model. Journal of the Operational Research Society, 50(6), 636-644.

Perelman, S., and Santin, D. (2011). Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. Education Economics, 19(1), 29-49.

Peyrache, A., and Coelli, T. (2009). Testing procedures for detection of linear dependencies in efficiency models. European Journal of Operational Research, 198(2), 647-654.

Ruggiero, J. (2003). Comment on estimating school efficiency. Economics of Education Review, 22(6), 631-634.

Ruggiero, J. (2004). Performance evaluation when non-discretionary factors correlate with technical efficiency. European Journal of Operational Research, 159(1), 250-257.

Ruggiero, J. (2005). Impact assessment of input omission on DEA. International Journal of Information Technology and Decision Making, 4(03), 359-368.

Schlotter, M., Schwerdt, G., and Woessmann, L. (2011). Econometric methods for causal evaluation of education policies and practices: a non-technical guide. Education Economics, 19(2), 109-137.

Simar, L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise? Journal of Productivity Analysis, 28(3), 183-201.

Simar, L., and Wilson, P. W. (2001). Testing restrictions in nonparametric efficiency models. Communications in Statistics-Simulation and Computation, 30(1), 159-184.

Simar, L., and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: recent developments and perspectives. The measurement of productive efficiency and productivity growth, 421-521.

Simar, L., and Wilson, P. W. (2011). Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. Foundations and Trends® in Econometrics, 5, 183-337.

Smith, P. (1997). Model misspecification in data envelopment analysis. Annals of Operations Research, 73, 233-252.

Solis, D., Bravo-Ureta, B. E., and Quiroga, R. E. (2007). Soil conservation and technical efficiency among hillside farmers in Central America: a switching regression model*. Australian Journal of Agricultural and Resource Economics, 51(4), 491-510.

Webbink, D. (2005). Causal effects in education. Journal of Economic Surveys, 19(4), 535-560.

Wilson, P. W. (2003). Testing independence in models of productive efficiency. Journal of Productivity Analysis, 20(3), 361-390.

Wooldridge, J. (2012). Introductory econometrics: A modern approach: Cengage Learning.

## 1.7 Figures and Tables

Figure 1.1: True frontier and DEA-BCC estimates under the assumption of exogenously distributed true efficiency

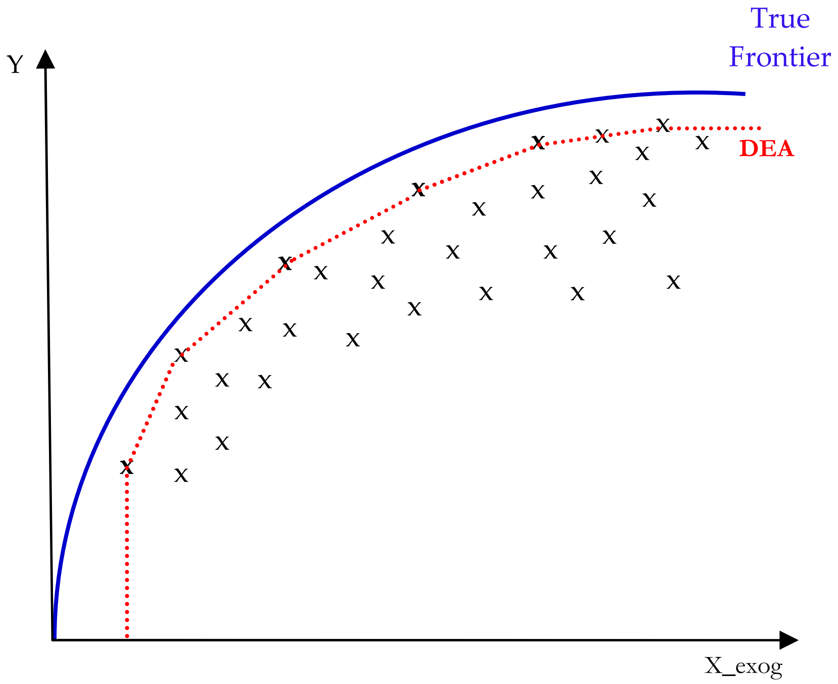


Figure 1.2: True frontier and DEA-BCC estimates under positive and high correlation between the true efficiency level and one input

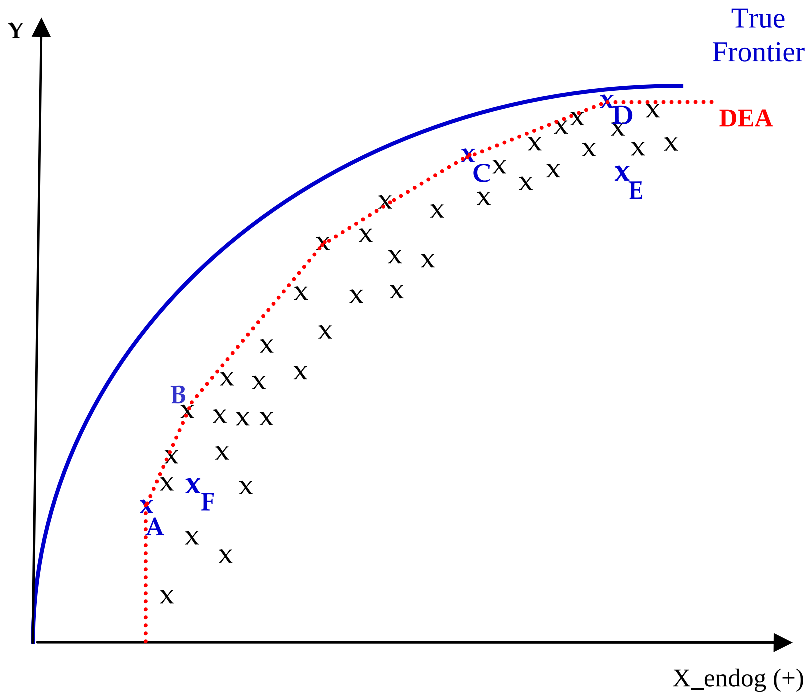



Figure 1.3: True frontier and DEA-BCC estimates under negative and high correlation between the true efficiency level and one input

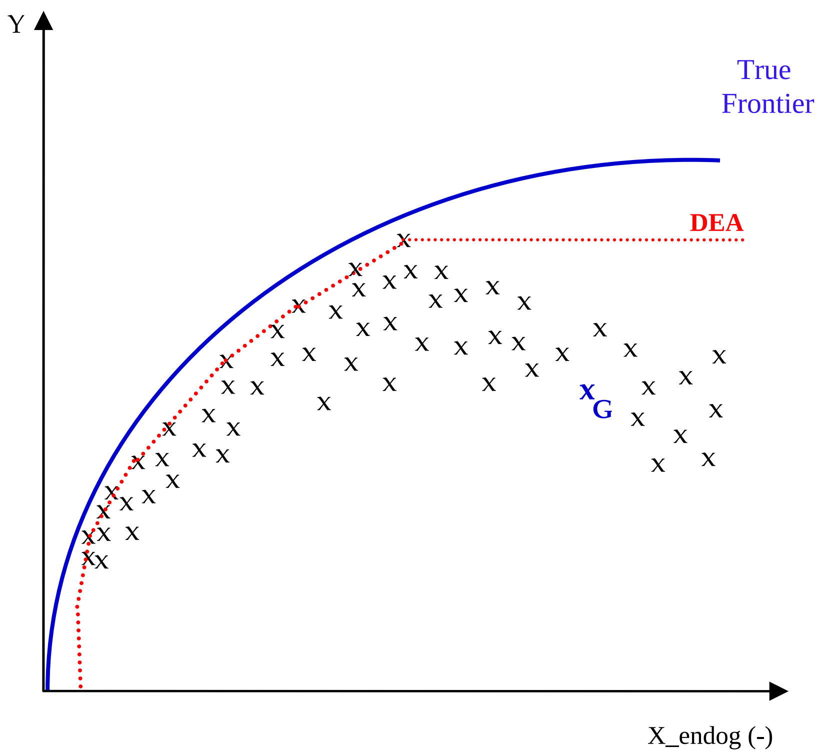


Table 1.1: Descriptive statistics of the correlation between true efficiency and the endogenous input in Monte Carlo scenarios

| | **Negative correlation** | | | **Positive correlation** | | |
|---|---|---|---|---|---|---|
| | HIGH (ρ ≅ – 0.8) | MEDIUM (ρ ≅- 0.4) | LOW (ρ ≅ - 0.2) | LOW (ρ ≅ 0.2) | MEDIUM (ρ ≅ 0.4) | HIGH (ρ ≅ 0.8) |
| | | | Cobb-Douglas Technology | | | |
| Mean | -0.809 | -0.422 | -0.235 | 0.231 | 0.426 | 0.811 |
| Std. Deviation | 0.033 | 0.084 | 0.096 | 0.095 | 0.083 | 0.036 |
| | | | *Translog* Technology | | | |
| Mean | -0.813 | -0.425 | -0.230 | 0.239 | 0.427 | 0.812 |
| Std. Deviation | 0.033 | 0.085 | 0.096 | 0.097 | 0.081 | 0.034 |

*Note*: Mean values after 1,000 replications. Sample size N=100.

Table 1.2: Accuracy measures of DEA estimates in Monte Carlo simulations (Cobb Douglas)

| | **Spearman's correlation** | | **Estimated mean efficiency** | | **MAE** | |
|---|---|---|---|---|---|---|
| | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** |
| **ρ ≅ - 0.8** | 0.695 | 0.574 | 0.898 | 0.937 | 0.058 | 0.085 |
| **ρ ≅ - 0.4** | 0.774 | 0.689 | 0.887 | 0.916 | 0.049 | 0.067 |
| **ρ ≅ - 0.2** | 0.778 | 0.686 | 0.885 | 0.913 | 0.048 | 0.064 |
| **ρ ≅ 0** | 0.778 | 0.671 | 0.884 | 0.912 | 0.049 | 0.065 |
| **ρ ≅ 0.2** | 0.754 | 0.622 | 0.886 | 0.915 | 0.051 | 0.068 |
| **ρ ≅ 0.4** | 0.715 | 0.560 | 0.890 | 0.919 | 0.055 | 0.073 |
| **ρ ≅ 0.8** | 0.520 | 0.300 | 0.911 | 0.942 | 0.073 | 0.094 |

*Note*: Mean values after 1,000 replications. Sample size N=100.

Table 1.3: Accuracy measures of DEA estimates in Monte Carlo simulations (*Translog*)

| | **Spearman's correlation** | | **Estimated mean efficiency** | | **MAE** | |
|---|---|---|---|---|---|---|
| | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** |
| **$\rho \cong$ - 0.8** | 0.713 | 0.708 | 0.800 | 0.957 | 0.089 | 0.097 |
| **$\rho \cong$ - 0.4** | 0.717 | 0.765 | 0.803 | 0.895 | 0.084 | 0.073 |
| **$\rho \cong$ - 0.2** | 0.700 | 0.757 | 0.808 | 0.893 | 0.083 | 0.071 |
| **$\rho \cong$ 0** | 0.669 | 0.729 | 0.815 | 0.893 | 0.083 | 0.072 |
| **$\rho \cong$ 0.2** | 0.619 | 0.675 | 0.827 | 0.898 | 0.084 | 0.078 |
| **$\rho \cong$ 0.4** | 0.564 | 0.612 | 0.841 | 0.905 | 0.086 | 0.085 |
| **$\rho \cong$ 0.8** | 0.305 | 0.342 | 0.892 | 0.936 | 0.105 | 0.116 |

*Note*: Mean values after 1,000 replications. Sample size N=100.

Table 1.4: Monte Carlo results of DEA-CRS and DEA-VRS ability to correctly assign observations into quintiles (Cobb Douglas)

| | % Assigned to the correct quintile | | % Assigned two or more quintiles from actual | | % Assigned to bottom quintile actually in bottom | | % Assigned to top quintile actually in top | | % Assigned to bottom quintile actually in the two first quintiles | | % Assigned to top quintile actually in the two last quintiles | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** |
| $\rho \cong -0.8$ | 43.5 | 35.2 | 16.9 | 24.5 | 68.3 | 57.8 | 47.8 | 37.6 | 1.2 | 4.9 | 12.2 | 17.3 |
| $\rho \cong -0.4$ | 48.5 | 42.2 | 12.1 | 17.0 | 74.3 | 69.4 | 52.6 | 43.0 | 0.4 | 0.8 | 7.3 | 12.1 |
| $\rho \cong -0.2$ | 48.8 | 43.1 | 11.8 | 16.9 | 74.9 | 70.8 | 52.3 | 42.8 | 0.4 | 0.7 | 7.0 | 13.1 |
| $\rho \cong 0$ | 48.5 | 42.6 | 11.9 | 17.5 | 74.9 | 70.1 | 52.1 | 42.3 | 0.3 | 0.7 | 7.0 | 14.4 |
| $\rho \cong 0.2$ | 46.6 | 40.2 | 13.3 | 20.0 | 72.4 | 66.3 | 49.9 | 40.2 | 0.5 | 1.0 | 8.5 | 17.6 |
| $\rho \cong 0.4$ | 44.1 | 37.3 | 15.7 | 23.4 | 69.0 | 60.7 | 48.0 | 38.6 | 0.8 | 1.8 | 10.6 | 20.8 |
| $\rho \cong 0.8$ | 34.4 | 27.0 | 26.2 | 36.5 | 50.1 | 36.6 | 41.4 | 32.6 | 4.5 | 10.8 | 21.6 | 31.8 |

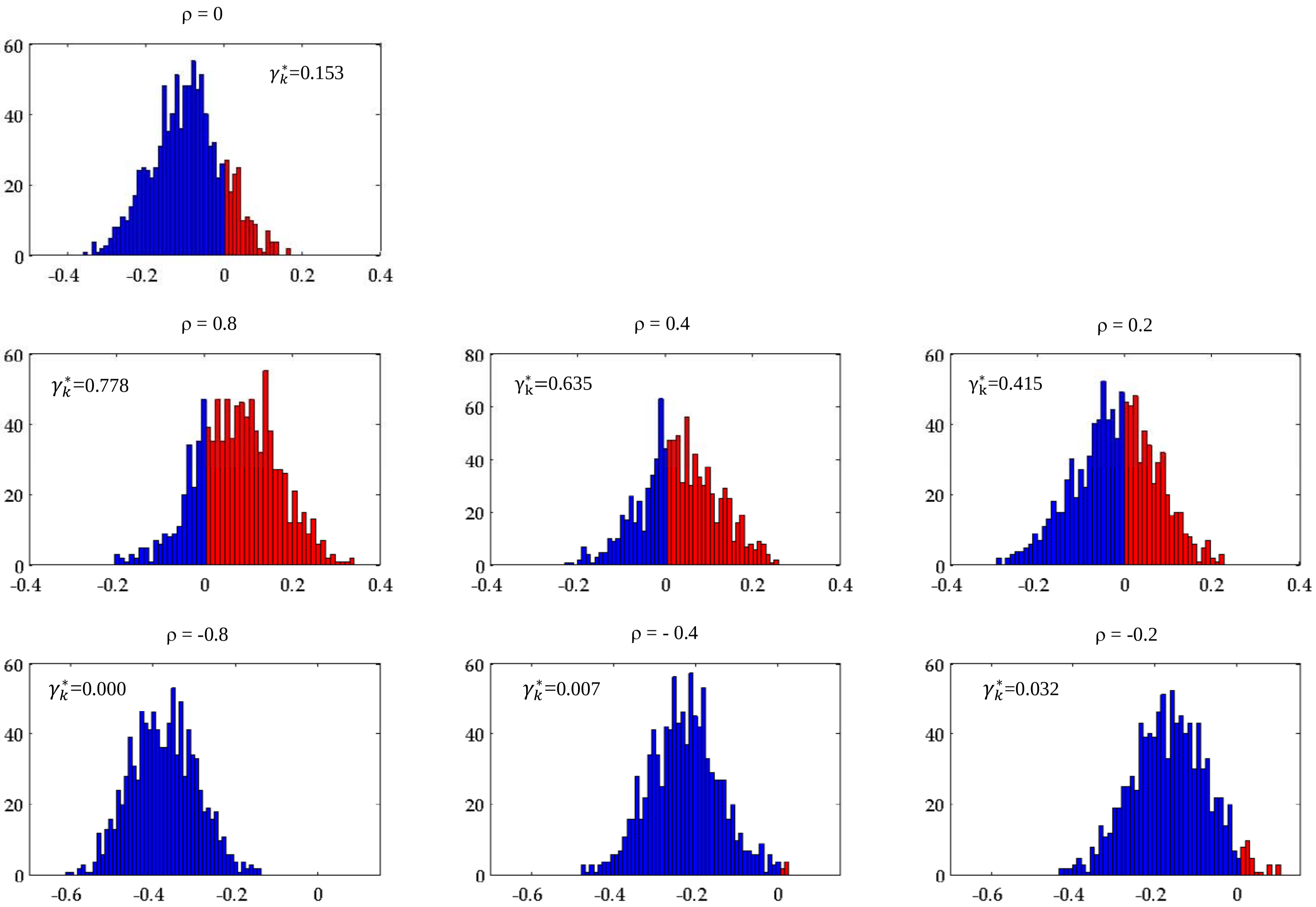
*Note*: Mean values after 1,000 replications. Sample size N=100.

Table 1.5: Monte Carlo results of DEA-CRS and DEA-VRS ability to correctly assign observations into quintiles (*Translog*)

| | **% Assigned to the correct quintile** | | **% Assigned two or more quintiles from actual** | | **% Assigned to bottom quintile actually in bottom** | | **% Assigned to top quintile actually in top** | | **% Assigned to bottom quintile actually in the two first quintiles** | | **% Assigned to top quintile actually in the two last quintiles** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** | **CRS** | **VRS** |
| **$\rho \cong -0.8$** | 42.8 | 42.7 | 16.6 | 16.0 | 64.1 | 70.9 | 53.8 | 43.3 | 1.0 | 1.4 | 9.3 | 10.4 |
| **$\rho \cong -0.4$** | 42.7 | 48.2 | 16.9 | 11.5 | 66.7 | 75.7 | 51.2 | 48.3 | 1.5 | 0.2 | 6.8 | 9.3 |
| **$\rho \cong -0.2$** | 41.4 | 48.8 | 18.2 | 11.8 | 66.0 | 76.1 | 49.1 | 48.6 | 2.0 | 0.2 | 6.7 | 10.3 |
| **$\rho \cong 0$** | 39.5 | 47.2 | 20.3 | 13.3 | 64.4 | 74.8 | 45.8 | 46.9 | 2.6 | 0.2 | 7.5 | 12.3 |
| **$\rho \cong 0.2$** | 37.0 | 44.0 | 23.3 | 16.3 | 61.1 | 70.4 | 41.7 | 44.4 | 3.9 | 0.3 | 9.8 | 15.5 |
| **$\rho \cong 0.4$** | 34.8 | 40.1 | 26.3 | 19.8 | 57.8 | 64.8 | 38.4 | 42.1 | 5.5 | 0.7 | 12.8 | 18.6 |
| **$\rho \cong 0.8$** | 26.9 | 28.0 | 37.5 | 34.8 | 39.8 | 40.8 | 28.3 | 32.8 | 17.1 | 8.2 | 27.8 | 30.3 |

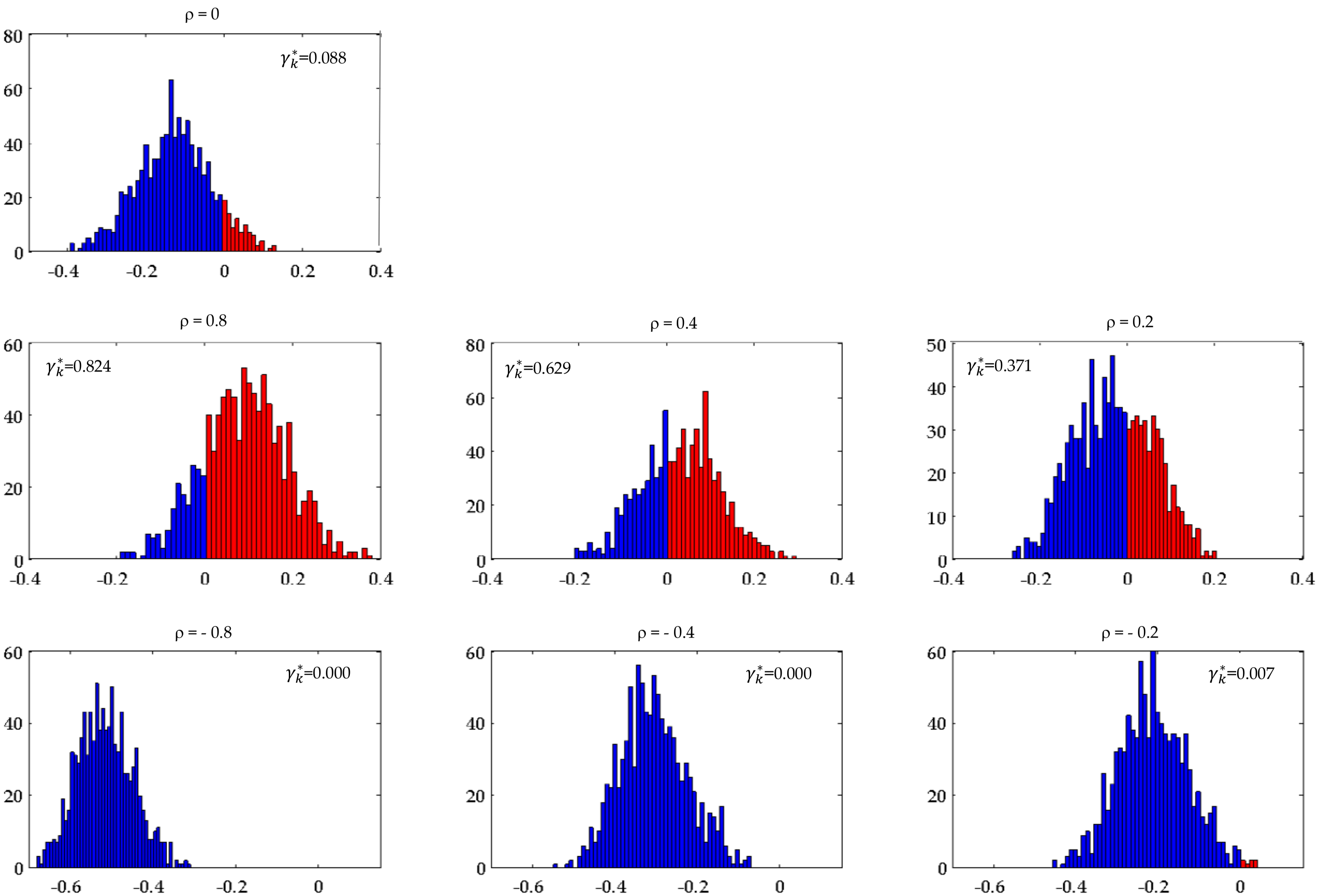
*Note*: Mean values after 1,000 replications. Sample size N=100.

Figure 1.4: Monte Carlo inputs classification using the proposed heuristic method to identify endogenous inputs (Cobb Douglas)




*Note*: Bootstrap size B = 1,000. Sample size N=100. DEA and II-DEA were estimated under CRS.

Figure 1.5: Monte Carlo inputs classification using the proposed heuristic method to identify endogenous inputs (*Translog*)




*Note*: Bootstrap size B = 1,000. Sample size N=100. DEA and II-DEA were estimated under CRS.

Table 1.6: Accuracy measures for conventional DEA and II-DEA estimates in Monte Carlo simulations (Cobb Douglas)

| | Spearman's correlation | | Estimated mean efficiency | | MAE | |
|---|---|---|---|---|---|---|
| | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** |
| **ρ ≅ - 0.8** | 0.695 | -0.270 | 0.898 | 0.849 | 0.058 | 0.136 |
| **ρ ≅ - 0.4** | 0.774 | 0.146 | 0.887 | 0.794 | 0.049 | 0.133 |
| **ρ ≅ - 0.2** | 0.778 | 0.289 | 0.885 | 0.777 | 0.048 | 0.134 |
| **ρ ≅ 0** | 0.778 | --- | 0.884 | --- | 0.049 | --- |
| **ρ ≅ 0.2** | 0.754 | 0.586 | 0.886 | 0.750 | 0.051 | 0.138 |
| **ρ ≅ 0.4** | 0.715 | 0.693 | 0.890 | 0.744 | 0.055 | 0.139 |
| **ρ ≅ 0.8** | 0.520 | 0.881 | 0.911 | 0.732 | 0.073 | 0.141 |

*Note*: Mean values after 1,000 replications. Sample size N=100.

Table 1.7: Accuracy measures for conventional DEA and II-DEA estimates in Monte Carlo simulations (*Translog*)

| | Spearman's correlation | | Estimated mean efficiency | | MAE | |
|---|---|---|---|---|---|---|
| | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** |
| **ρ ≅ - 0.8** | 0.708 | 0.128 | 0.957 | 0.893 | 0.097 | 0.127 |
| **ρ ≅ - 0.4** | 0.765 | 0.362 | 0.895 | 0.846 | 0.073 | 0.109 |
| **ρ ≅ - 0.2** | 0.757 | 0.439 | 0.893 | 0.831 | 0.071 | 0.105 |
| **ρ ≅ 0** | 0.729 | --- | 0.893 | --- | 0.072 | --- |
| **ρ ≅ 0.2** | 0.675 | 0.605 | 0.898 | 0.810 | 0.078 | 0.100 |
| **ρ ≅ 0.4** | 0.612 | 0.657 | 0.905 | 0.804 | 0.085 | 0.099 |
| **ρ ≅ 0.8** | 0.342 | 0.760 | 0.936 | 0.794 | 0.116 | 0.097 |

*Note*: Mean values after 1,000 replications. Sample size N=100.

Table 1.8: Monte Carlo results of conventional DEA and II-DEA ability to correctly assign observations into quintiles (Cobb-Douglas)

| | % Assigned to the correct quintile | | % Assigned two or more quintiles from actual | | % Assigned to bottom quintile actually in bottom | | % Assigned to top quintile actually in top | | % Assigned to bottom quintile actually in the two first quintiles | | % Assigned to top quintile actually in the two last quintiles | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** |
| **ρ ≅ - 0.8** | 43.5 | 14.7 | 16.9 | 56.3 | 68.3 | 12.3 | 47.8 | 7.1 | 1.2 | 72.2 | 12.2 | 42.2 |
| **ρ ≅ - 0.4** | 48.5 | 21.9 | 12.1 | 43.6 | 74.3 | 25.5 | 52.6 | 25.2 | 0.4 | 41.1 | 7.3 | 22.2 |
| **ρ ≅ - 0.2** | 48.8 | 25.2 | 11.8 | 38.1 | 74.9 | 32.2 | 52.3 | 32.0 | 0.4 | 31.0 | 7.0 | 16.7 |
| **ρ ≅ 0** | 48.5 | --- | 11.9 | --- | 74.9 | --- | 52.1 | --- | 0.3 | --- | 7.0 | --- |
| **ρ ≅ 0.2** | 46.6 | 35.6 | 13.3 | 24.6 | 72.4 | 50.4 | 49.9 | 48.1 | 0.5 | 11.5 | 8.5 | 7.8 |
| **ρ ≅ 0.4** | 44.1 | 41.1 | 15.7 | 18.4 | 69.0 | 58.2 | 48.0 | 54.7 | 0.8 | 6.2 | 10.6 | 5.0 |
| **ρ ≅ 0.8** | 34.4 | 57.0 | 26.2 | 5.1 | 50.1 | 77.8 | 41.4 | 67.6 | 4.5 | 0.1 | 21.6 | 1.2 |

*Note*: Mean values after 1,000 replications. Sample size N=100.

Table 1.9: Monte Carlo results of conventional DEA and II-DEA ability to correctly assign observations into quintiles (*Translog*)

| | % Assigned to the correct quintile | | % Assigned two or more quintiles from actual | | % Assigned to bottom quintile actually in bottom | | % Assigned to top quintile actually in top | | % Assigned to bottom quintile actually in the two first quintiles | | % Assigned to top quintile actually in the two last quintiles | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** | **DEA** | **II-DEA** |
| **$\rho \cong -0.8$** | 42.7 | 26.2 | 16.0 | 39.8 | 70.9 | 41.7 | 43.3 | 46.5 | 1.4 | 43.7 | 10.4 | 26.7 |
| **$\rho \cong -0.4$** | 48.2 | 30.9 | 11.5 | 31.5 | 75.7 | 45.3 | 48.3 | 47.2 | 0.2 | 25.8 | 9.3 | 20.8 |
| **$\rho \cong -0.2$** | 48.8 | 33.2 | 11.8 | 28.3 | 76.1 | 47.5 | 48.6 | 47.3 | 0.2 | 19.9 | 10.3 | 19.0 |
| **$\rho \cong 0$** | 47.2 | --- | 13.3 | --- | 74.8 | --- | 46.9 | --- | 0.2 | --- | 12.3 | --- |
| **$\rho \cong 0.2$** | 44.0 | 40.0 | 16.3 | 20.4 | 70.4 | 57.9 | 44.4 | 46.8 | 0.3 | 7.4 | 15.5 | 17.0 |
| **$\rho \cong 0.4$** | 40.1 | 43.1 | 19.8 | 17.1 | 64.8 | 62.6 | 42.1 | 46.5 | 0.7 | 4.0 | 18.6 | 16.8 |
| **$\rho \cong 0.8$** | 28.0 | 51.8 | 34.8 | 10.0 | 40.8 | 75.7 | 32.8 | 46.9 | 8.2 | 0.1 | 30.3 | 15.6 |

*Note*: Mean values after 1,000 replications. Sample size N=100.

# Chapter 2

# Measuring educational efficiency drivers under endogeneity: an application to public schools in Uruguay

## 2.1 Introduction

The interest in improving school performance and educational attainment through efficiency gains is growing basically in response to three main findings. First, improved academic outcomes have been proven to have a positive impact on development and economic growth (Barro and Lee, 2012;Hanushek and Woessmann, 2012). Second, public expenditure on education is one of the largest public budget items, and the public sector is the main provider of education in most countries. In fact, the public sector is the main provider of secondary education in Uruguay where, in 2012, 84% of the students were enrolled in public schools[1]. Thirdly, there is still no concluding empirical evidence to show that a higher level of resources leads *per se* to better results, which leads to the suspicion that there are great inefficiencies in several education systems (Hanushek, 2003).

The level of educational expenditure and its percentage share of GDP are indicators commonly used to measure a country's educational investment. In this sense, Public expenditure on education accounted for 3.53% of Uruguay's GDP in 2000, whereas ten years later it had risen to 4.5%[2]. But unfortunately this significant budgetary effort has not been accompanied by adequate reforms and public policies leading to better educational achievement in public schools. Conversely, the Uruguayan education system has entered into stagnation and recession in recent years, particularly at the public secondary education level, which has recorded high repetition and drop-out rates as well as a steady decline in academic performance. For example, the repetition rate from 1st to 4th grades in public schools has increased between 2003 and 2012 from 21.3% to 27% while the attainment rate was reduced from 72.7% to 67.4% in the same period[3] . In addition, as evidenced by the latest results published in the PISA 2012 (Programme for International Student Assessment) Report from the OECD (Organisation for Economic Co-operation and Development), results in public schools remain steady across the first three waves in which Uruguay has participated, showing a downward trend in the last cycle (416, 420, 419 and 399 average points in 2003, 2006, 2009 and 2012, respectively). Part of this educational decline could be explained by the increase in public secondary school enrolment that has occurred in the last decade (4.3% from 2006 to 2012)[4]. In general, students who joined the public system have poorer socio-economic status and educational outcomes. However, given the significant increase in resources, it would have been expected that they were invested in promoting educational practices that would enable a properly adaptation of these students, so they could achieve similar results to those already in the public system.

As a consequence of these poor results, the Uruguayan public educational system problems are a recurring concern, not only for educational policy-makers and the government but also for teachers and families involved in the education process. Still, in many cases, the discussion

[1]Education Observatory, National Administration of Public Education (ANEP).
[2]The GDP grew by 33% in real terms over this period (Uruguayan Central Bank (BCU)).
[3]Education Observatory, National Administration of Public Education (ANEP).
[4]Education Observatory, National Administration of Public Education (ANEP).

primarily focuses on increasing public resources expended on education, although as we noted above there is no concluding empirical evidence to show that a higher level of resources leads per se to better results.

These contexts highlight the need to evaluate the current public education system from a different perspective based not only on educational outcomes, but on exploring the existence of inefficient behaviours in the production process and exploring the sources of these inefficiencies. The presence of inefficient schools in the education system means that it is not being made the most of the educational resources and, therefore, that it would be feasible to increase academic results with the current levels of resources, which is one of the educational authorities main targets.

The most popular models applied for explaining the sources of inefficiency in public sector are the semi-parametric two-stage models popularized by Ray (1991) and McCarty and Yaisawarng (1993)[5]. Under this approach, in a first stage a Data Envelopment Analysis (DEA) model is used to estimate a production frontier, which defines both the efficient and inefficient units. In the second stage, a regression technique is applied to explain the inefficient behaviours taking into account student and school contextual variables[6]. There are several international educational efficiency studies applying semi-parametric two-stage models for explaining schools efficiency (Charnes et al. 1997; Grosskopf et al. 1997; Xue and Harker 1999; Mancebón and Mar-Molinero 2000; Afonso and St.Aubyn 2006; Hoff 2007; Simar and Wilson 2007; Cordero et al. 2008; 2010; Alexander et al. 2010; De Jorge and Santín 2010)[7]. However, in the Latin American context, there is little available research and in particular, to the best of our knowledge, for the Uruguayan case there are no studies using this efficiency approach[8].

Two-stage models differ primarily in the regression model specified in the second stage to explain efficiency scores. The most commonly applied methodology has been the censored regression model (Tobit), followed by ordinary least squares (OLS) and the truncated regression. Xue and Harker (1999) were the first to point out the main drawback of the two-stage approach. They underline that two-stage model results are bound to be biased due to the fact that the radial efficiency scores estimated in the first stage (the dependent variable in the second stage) depend on each other. Hence, conventional inference methods are invalid in this context because the error term is serially correlated, and this violates the basic econometric assumption of independence within the sample. To overcome this drawback, Simar and Wilson (2007, 2011) proposed a new estimation methodology based on the use of bootstrapping. However, as the discussion about which is (are) the best model(s) to be run in the second-stage regression is ongoing, we run four different second-stage specifications following the available literature in order to check the

[5]In a recent Study Liu et al. (2013) surveys the DEA literature by applying a citation-based approach and find that 'two-stage contextual factor evaluation framework' is the most active DEA sub-area in recent years.

[6]See Simar and Wilson (2007) for a detailed review of two-stage models.

[7]For a more detailed review see Simar and Wilson (2007).

[8]In Uruguay, interest has traditionally focused on education system coverage rates, the system's redistributive effect and its impact on poverty and growth rather than the quality of the services provided and the academic outputs (Llambí and Perera, 2008; Llambí, Perera and Messina, 2009; Fernández, 2009).

robustness of the results[9].

As we evidenced in Chapter 1, there is a major and recurrent issue in production processes, namely, the presence of endogeneity, which is frequently overlooked when practitioners apply DEA (that is, the first stage in semi-parametric models). In fact, the education sector is one of the most illustrative contexts where this problem can be frequently observed (Webbink 2005, Schlotter et al. 2011). For example, highly motivated parents invest more time and money in choosing the best schools for their children. If parent motivation is correlated with family socio-economic level (Hoxby, 2000; Sacerdote, 2001), such pupils (and thus the school they attend) will tend to obtain better academic results for two reasons. First, because socio-economic level is an essential input in the educational production function. Second, because parents motivation (which is unobserved) also positively affects pupil academic achievement. As a result, schools with pupils from high socio-economic backgrounds will be more prone to be fully efficient. In this case, the mechanism of self-selection results in a positive correlation between schools technical efficiency and their average pupils socio-economic background. The same reasoning can be applied for teachers motivation. Not only the most disadvantaged students are less motivated, but also teachers and principals in these schools are less motivated (due to a self-selection problem if teachers can choose the school to teach or because, even if they were more motivated at the beginning, in more disadvantaged contexts they loose their initial motivation). Again, this process derives in a positive correlation between the input level and the school efficiency. The intensity of this correlation depends not only on the importance of parents, students or/and teachers motivation, but also on the correlation between this unobservable variables and the observed input (socio-economic level). That is, the greater stratification in the education system, the higher level of endogeneity.

If we analyze the Uruguayan academic results more in detail, we observe that pupils' socio-economic contexts have a great impact on schools performance. In fact, in the last PISA 2012 Report Uruguay shows the higher score gap between students by socio-economic level in mathematics of all Latin American countries who participated in the study[10]. If we consider the level of educational resources in public schools, the differences between schools are notably less pronounced than in the case of educational outcomes. This situation leads us to suspect that public schools from more disadvantaged contexts not only have pupils from poorer socio-economic status and poorer educational resources, but they are also less efficient than public schools from more advantaged contexts. Moreover, in Uruguay the school type is strongly associated with pupils socio-economic status. Figure 2.1 shows the distribution of pupils into public and private schools, from which we can observe almost a perfectly split market. Students from high income families attend private schools, while pupils from more disadvantaged backgrounds are enrolled in public schools. This great segmentation reinforces the endogeneity problem in the public sector where schools do not have to compete to attract students (making it possible for inefficient

[9]See Hoff (2007), Banker and Natarajan (2008), Mc Donald (2009) and Ramalho, Ramalho and Henriques (2010).

[10]For a detailed analysis see http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=38667314

schools to survive over time). This great market segmentation by school type suggests that public and private sectors operate under different circumstances and they use different technologies, therefore to estimate the efficiency level in both sector they should be analyzed separately.

In the previous chapter, using synthetic data generated in a Monte Carlo experiment we found that although DEA is robust to negative endogeneity, the presence of a higher or medium positive endogeneity severely biases DEA performance. However, this problem is frequently ignored when practitioners apply non parametric techniques, included DEA, to estimate technical efficiency in the education sector. In this chapter, we take this problem into account. To do this, we apply the proposed method to detect the presence of endogenous inputs in the Uruguayan public secondary education system and we apply the proposed *Instrumented Input Data Envelopment Analysis* (II – DEA) method to deal with this problem in order to obtain more reliable DEA estimates.

The aim of this research is twofold. First, we explore the sources of inefficiency of the Uruguayan public secondary schools in order to provide new timely and complementary evidence for the current national debate about which public policies and educational practices could contribute to improve public schools academic results with the current resources. For this purpose, using data from PISA 2012 we apply a semi-parametric two-stage model that incorporates a new II-DEA in the first stage which allow us to take into account the presence of the endogeneity problem. Secondly, we investigate the impact of considering or omitting this issue in empirical terms and the implications for public policy recommendations. To explore that, we also provide the results using the conventional DEA in the first step and we compare them with those from the II-DEA method.

The chapter is organized as follows. In section 2.2 we present the main methodological concepts. Section 2.3 briefly describes the Uruguayan education system, the PISA program and presents the selected variables for the analysis. In section 2.4 we discuss the results. Finally, the last section 2.5 is devoted to the conclusions and their implications in terms of educational public policies.

## 2.2 Methodology

### 2.2.1 The Educational Production Function

The concept of educational production function refers to the relationship between inputs and outputs for a given production technology. The theoretical approach used in this paper for linking resources to educational outcomes at the school level is based on the well-known educational production function proposed by Levin (1974) and Hanushek (1979), and more recently by Hanushek et al. (2013)

$$A_i = f(B_i, S_i) \tag{2.1}$$

where sub index $i$ refers to school, and $A_i$ represents the educational output vector for school $i$, normally measured through the average student score on standardized tests. On the other hand, educational inputs are divided into $B_i$ , which denotes the average student family background, and $S_i$ which are school educational resources.

The educational production function can be estimated considering the possible existence of inefficient behaviours in schools. Differences in efficiency may be due to multiple factors, such as poor teacher motivation, teaching and class organization issues, teacher quality or school management. All these factors may affect student performance significantly. In this case, we estimate a production frontier where fully efficient schools would belong to the frontier. These relatively efficient units achieve the maximum observed result given their resources allocation. Inefficient units do not belong to the estimated frontier, and their inefficiency level is measured by the radial distance between each school and the constructed frontier. The production frontier to be estimated at school level would be

$$A_i = f(B_i, S_i).u_i \tag{2.2}$$

where $0 \leq u_i \leq 1$ denotes the school efficiency level. Values of $u_i = 1$ imply that the analyzed schools are fully efficient, meaning that given the initial input endowment and the existing technology, these schools are maximizing and correctly managing the resulting outputs. Lower than one values $u_i \leq 1$ would indicate that the school is inefficient.

In short, three types of variables are involved in the production process: educational outputs $(A_i)$, educational inputs $(B_i, S_i)$ , and the estimated efficiency level $u_i$ for each school. Ray (1991) and McCarty and Yaisawarng (1993) were the first to propose applying a semi-parametric two-stage model to estimate efficiency scores and identify the main drivers. This approach uses a DEA model in the first stage which measures the technical efficiency, whereas a regression analysis conducted in the second stage seeks out the main explanatory factors of efficiency. A more detailed description of both stages of these semi-parametric models follows.

### 2.2.2 First stage: DEA and II-DEA models

The estimation of efficiency is associated with Farrel's concept of technical efficiency (Farrel 1957); who defines the production frontier as the maximum level of output that a decision-making unit (DMU) can achieve given its inputs and the technology (output orientation). In practice, the true production frontier and the technology are not available and should be estimated from the relative best practices observed in the sample.

There are basically two main groups of techniques for estimating the production frontier: parametric or econometric approaches (see Battese and Coelli 1988, 1992, 1995 for a review) and non-parametric methods based on mathematical optimization models. Although the use of the parametric approaches has increased in the last decades[11], nonparametric methods have been the most extensively applied for measuring educational technical efficiency.

[11] See, for example, Perelman and Santín (2011) and Crespo-Cebada et al. (2013).

Since the pioneering work by Charnes, Cooper and Rhodes (1981) and Banker, Charnes and Cooper (1984)[12], the DEA[13] model has been widely used to measure efficiency in several areas of public sector production. The main reason for its widespread application is its flexibility: it accounts for multiple outputs, for the uncertainty about the true production technology and for the lack of price information, making it well suited to the peculiarities of the public sector. In addition, it is a technique that can be easily translated to stakeholders and politicians, who are often not familiar with econometrics and statistics and therefore are somewhat reluctant to these techniques. The DEA model applies a linear optimization program to obtain a production frontier that includes all the efficient units and their possible linear combinations. As a result, the estimated efficiency score for each DMU is a relative measure calculated using all the production units that are compared. The formulation of the output-oriented DEA program under variable returns to scale (DEA-BBC model) for each analysed unit is

$$\varphi_i = \max_{\lambda,\varphi}\{\varphi_i | \varphi y_i \leq Y\lambda; x_i \geq X\lambda; n1'\lambda = 1; \lambda \geq 0\} \ \ \forall_i = 1, ..., n \tag{2.3}$$

where, for the ith DMU, $\varphi_i \geq 1$ is the estimated efficiency score, $y_i$ is the output vector $(q \times 1)$ and $x_i$ is the input vector $(p \times 1)$, and thus X and Y are the respective input $(p \times n)$ and output $(q \times n)$ matrices. The $(n \times 1)$ vector $\lambda$ contains the virtual weights of each unit determined by the problem solution. When $\varphi_i = 1$, the analyzed unit belongs to the frontier (is fully efficient), whereas $\varphi_i > 1$ indicates that the ith unit is inefficient, $\varphi_i$ being the radial distance between the ith unit and the frontier. In other words, indicates the equiproportional expansion over outputs needed to reach the frontier. Therefore, the higher the score value $\varphi_i$, the greater the inefficiency level.

In the previous chapter, using synthetic data generated in a Monte Carlo experiment we found that although DEA is robust to negative endogeneity, the presence of a higher or medium positive endogeneity severely biases DEA performance. Also, in the introductory section we have discussed the potential presence of endogeneity in public secondary schools in Uruguay associated with students socio-economic level. That is, we suspect that exists a positive correlation between the true technical efficiency of schools and the average students socio-economic level. To take this potential problem into account, in this chapter we adapt the conventional two-stages models by introducing a correction method in the efficiency estimates at the first stage. We apply the *'Instrumental Input DEA'* strategy proposed in section 1.4 in Chapter 1.

The idea behind this proposal is to include as an input in the DEA specification only the exogenous part of the endogenous input. To do this, we replace the endogenous input by an exogenous variable, which only contains the exogenous information of the original one, that is, that part which is uncorrelated with the efficiency.

[12]The DEA-CCR model and DEA-BBC model, respectively.
[13]See Worthington (2001, p. 253f) for a detailed review of available research that measures efficiency in education through frontier techniques and mostly DEA models.

Consider the single-output multi-input productive dataset $\chi = \{(X_i, Y_i)\ i = 1, ..., n\}$, where one input is significantly positive correlated with the true efficiency term (hereafter the endogenous input E). As in the classic Instrumental Variables (IV) approach, the first step is to find a good instrumental input G, which must satisfy at the same time two basic conditions:

i. Relevance: the instrument G must be significantly correlated with the endogenous input E, *i.e.* $E(E|G) \neq 0$;

ii. Exogeneity: the instrument G must be uncorrelated with the true efficiency term, *i.e.* $E(\varphi|G) = 0$

The first condition can be contrasted in empirical applications by testing the significance of the parameter $\tau$ in the following estimated regression $E = \alpha + \tau G + \xi$. If we do not reject$H_0 : \tau = 0$, we can assume that the instrument is relevant. However, the second condition cannot be directly tested because true efficiency is not observed in empirical settings. In this case, the exogeneity condition can be interpreted as the absence of a causal relationship between the instrumental input G and the output variable Y. That is, G should have no partial effect on Y (beyond the effect through the endogenous input). The II-DEA procedure is implemented following two simple steps:

1. The aim of the first step is to isolate the exogenous component of the endogenous input that is uncorrelated with the true efficiency. To do this, regress the endogenous input (E) over the instrumental input (G) and the rest of the exogenous inputs

$$E = \alpha + \delta_1 x_1 + ... + \delta_{k-1} x_{k-1} + \phi G + \mu_i \tag{2.4}$$

   where $x_{k-1}$ are the k-1 exogenous inputs, G is the instrumented input and $\mu_i$ is a random white noise component.

2. Secondly, in order to obtain the corrected DEA efficiency scores for each DMU replace the endogenous input (E) by the estimated exogenous variable $\hat{E}_i$ in the conventional DEA linear program (2.3).

In this research we apply both methods, the conventional DEA and the II-DEA, to investigate the empirical implications for educational public policy recommendations of considering or not the endogeneity problem. After estimating efficiency scores, these are regressed on different contextual variables to explore the sources of the inefficient behaviours of public secondary schools.

In order to test if the students' socio-economic level is an endogenous input in our sample, we apply the simple heuristic proposed in section 1.4 in chapter 1. This method relies on the analysis of the correlation coefficients between the inputs and the estimated efficiency scores. From a microeconomic viewpoint and assuming that inputs are exogenous, the correlation coefficient between the inputs and the DEA estimated efficiency scores should be slightly negative and close to zero (or at least non-positive), as DEA assumes that for a given output, the more input level the higher inefficiency. Then, the proposed heuristic is based on these expected correlation coefficients in order to classify the nature of each input included in the DEA model. In practice, we proceed in six steps as follows:

i. From the empirical dataset $\chi = \{(X_i, Y_i)\ i = 1, ..., n\}$ randomly draw with replacement a bootstrap sample B=1,000 $\chi_b^* = \{(X_{ib}^*, Y_{ib}^*)\ i = 1, ..., n\}$

ii. Estimate $\hat{\theta_{ib}^*} = \dfrac{1}{\varphi_{ib}^*} \leq 1\ \ i = 1, ..., n$ using the LP (2.3)

iii. For each input $k = 1, ..., p$ compute the Pearson's correlation coefficient between the estimated efficiency score $\hat{\theta_{ib}^*}$ and the input $k$ $\rho_{kb}^* = corr(x_{ik}^*, \hat{\theta_i^*})\ \ i = 1, ..., n\ \ k = 1, ..., p$

iv. Repeat steps 1-3 B=1,000 times in order to obtain for $k = 1, ..., p$ a set of correlations: $\{\rho_{kb}^*,\ \ b = 1, ..., B\}$

v. For each input $k$ compute $\gamma_k^* = \dfrac{1}{B}\sum_{b=1}^{B}[I_{[0,1]}(\rho_k^*)]_b\ \ for\ \ k = 1, ..., p$ where $I_{[0,1]}(\rho_k^*)$ is the Indicator Function defined by:

$$I_{[0,1]}(\rho_k^*) = \begin{cases} 1, & \text{if } 0 \leq \rho_k^* \leq 1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

vi. Finally, classify each input using the following criterion:

- If $\gamma_k^* < 0.25 \rightarrow$ Exogenous/Negative endogenous input $k$
- If $0.25 \leq \gamma_k^* < 0.5 \rightarrow$ Positive LOW endogenous input $k$
- If $0.5 \leq \gamma_k^* < 0.75 \rightarrow$ Positive MIDDLE endogenous input $k$
- If $\gamma_k^* \geq 0.75 \rightarrow$ Positive HIGH endogenous input $k$

### 2.2.3 Second stage specifications

The estimated efficiency scores $\hat{\varphi}_i \geq 1$ are regressed on a vector $Z = (z_1, z_2, ..., z_r)$ of school and student contextual variables, which are not inputs but are related to the learning process

$$\hat{\varphi}_i = f(Z_i, \beta_i) \tag{2.6}$$

The most used estimation method in this second stage is the censored regression model (Tobit), followed by ordinary least squares (OLS)[14], from which the main explanatory factors

[14] Some authors actually estimate both models simultaneously to verify results robustness.

of the efficiency scores can be drawn[15]

$$\hat{\varphi}_i = f(Z_i, \hat{\beta}_i) + \epsilon_i \tag{2.7}$$

Xue and Harker (1999) argued that these conventional regression models applied in the second stage yield biased results because the efficiency scores estimated in the first stage ($\hat{\varphi}_i$) are serially correlated. Accordingly, there has been a lively debate in recent years about which would be the most accurate model to perform in this second stage in order to provide consistent estimates. According to Simar and Wilson (2007) (hereinafter referred to as SW2007), the efficiency rates estimated by the DEA model in the first stage are correlated by construction (as they are relative measures), and therefore estimates from conventional (2.7) would be biased. Additionally, the possible correlation of the contextual variables $Z_i$ with the error term $\epsilon_i$ in (2.7) is another source of bias.

SW2007 state the need for bootstrapping to overcome these drawbacks. In their paper, SW2007 propose two algorithms[16] which incorporate the bootstrap procedure in a truncated regression model. They run a Monte Carlo experiment to examine and compare the performance of these two algorithms, and they prove that both bootstrap algorithms outperform conventional regression methods (Tobit and truncated regressions without bootstrapping), yielding valid inference methods. For small samples (problems with fewer than 400 units and up to three outputs and three inputs), Algorithm 1 fits results better than Algorithm 2, which is more efficient as of samples that exceed 800 units[17]. Since the sample analysed in our research is lower than 400 units, we apply the simple Algorithm 1 proposed by SW2007 (p. 41), which is described below.

1. Estimate efficiency scores $\hat{\varphi}_i \forall_i = 1, 2, ..., n$ solving DEA (2.3)
2. Estimate $\hat{\beta}_i$ y $\hat{\sigma}_\varepsilon$ by maximum likelihood in the truncated regression of $\hat{\varphi}_i$ on $z_i$ (Equation (2.6)), using $m < n$ observations, where $\hat{\varphi}_i > 1$.
3. Loop over the next steps ([3.1]-[3-3]) L times to obtain a set of bootstrap estimates $A = \{(\hat{\beta}^*, \hat{\sigma}^*_\epsilon)\}^L_{b=1}$
   (a) For each $i = 1, ...., m$ extract $\varepsilon_i$ from a normal distribution $N(0, \hat{\sigma}^2_\varepsilon)$ left-truncated in $(1 - Z_i\hat{\beta})$
   (b) Again, for each $i = 1, ..., m$ compute $\hat{\varphi}^*_i = z_i\hat{\beta} + \varepsilon_i$
   (c) Using maximum likelihood, estimate the truncated regression of $\varphi^*_i$ one $z_i$, obtaining $\hat{\beta}^*$ and $\hat{\sigma}^*_\varepsilon$

[15]For a detailed review of estimation methods used in the second stage of semi-parametric models, see Simar and Wilson (2007).
[16]The authors propose a simple Algorithm 1 and a double Algorithm 2. The difference lies in the fact that Algorithm 2 incorporates an additional bootstrap in the first stage, which amends the estimates of the efficiency scores.
[17]For a more detailed analysis of the results, see Simar and Wilson (2007, p. 45f.).

4. Use the bootstrap values in A and the original $\beta$ and $\sigma_\epsilon$ to construct estimated confidence intervals for $\beta$ and $\sigma_\epsilon$

Later, Hoff (2007), McDonald (2009) and Ramalho, Ramalho and Henriques (2010) took up the discussion about the use of OLS, Tobit and fractional regression models in the second stage. Unlike Hoff (2007), who concluded that both (Tobit and OLS) models yield consistent estimations, McDonald (2009) shows that only the Tobit produces consistent results. Meanwhile, Banker and Natarajan (2008) (BN2008) provide a statistical model which yields consistent second-stage OLS estimations. Simar and Wilson (2011) again took part in the ongoing debate and compared the consistency between truncated regressions and the BN2008 OLS specification. They conclude that only the truncated regression model proposed by SW2007 and, under very particular and unusual assumptions, the OLS model presented by BN2008 provide consistent estimates. Furthermore, they prove that in both cases only bootstrap methods were capable of statistical inference.

Building upon this evidence we conclude that there is yet no agreement in the available literature about which is (are) the most consistent regression model(s). For this reason, two-stage model practitioners find the selection of the second-stage regression model baffling, as they are unsure about whether or not results will vary significantly with their choice of specification. To clarify this point, we have chosen to estimate four alternative regression models in the second stage and compare the results. First, we specify the conventional Tobit (censored regression model), as it is the most commonly used in the literature. Then, we estimate three regression models applying the bootstrap procedure: Algorithm 1 proposed by SW2007 based on a truncated regression; and a Tobit regression and an OLS model with bootstrapping. The aim here is to explore the real implications of this methodological discussion for policy recommendations derived from an empirical analysis of real educational data.

## 2.3 Data

### 2.3.1 Brief description of the Uruguayan education system

The Uruguayan national education system is composed of four levels: three years of infant education (three to five years old), six years of primary education (six to eleven years old), six years of secondary education (twelve to seventeen years old), and tertiary education from age eighteen. Secondary education is divided into three years of basic secondary education *(Ciclo Básico Común)* and three years of upper secondary education *(Bachillerato)*[18]. Compulsory education covers 14 years: the two last years of early education (four and five years old), primary and secondary education[19].

[18]In Uruguay there are two types of institutions to study secondary education: secondary schools and technical vocational schools.

[19]Art. 10 of the *General Education Law* N.18.437 of December 12, 2008.

In terms of public and private education coverage, the public sector takes absolute primacy over the private sector in all education levels, and particularly in secondary education. In 2012, 88% of high school students attended public schools (*Education Observatory*, National Administration of Public Education). This highlights how important the performance of public institutions is for national academic results, and therefore the need to assess both the management and the teaching practices implemented by these schools.

Uruguay has historically occupied a leading position in Latin America in terms of educational achievement, according to the main standard indicators and international studies. However, the Uruguayan public education system (particularly at the secondary level) is currently undergoing a phase of stagnation and recession. The major budgetary effort made by the government in the last decade has not been accompanied by effective reforms and policies that achieve improved educational outcomes.

The results of PISA 2012 corroborate that Uruguay is still in an advantageous position within the region[20], but also confirm that results have not improved compared to previous waves. In addition, test scores in the three analysed areas are more highly dispersed than in other countries, which mirror the high social segmentation of the education system. Comparing student's performance by the school's socio-economic context, it is noteworthy that while almost 89% of students who attend to schools in "very unfavourable circumstances" do not reach the minimum 'competence threshold' defined by the OECD in mathematics[21]., this figure drops to 13%[22] for students who attend to schools in "very favourable circumstances"[23]. By contrast, analysing the percentage of top-scoring students (performance levels four to six) defined by PISA analysts, we find that this proportion rises to almost 30% of students in "very favourable circumstances" whereas students from "very unfavourable circumstances" account for less than 1%. As we exposed in the introduction section, this great inequality in academic results leads us to suspect for the presence of endogeneity related to the school's socio-economic context. That is, schools from more disadvantaged contexts not only have pupils from poorer socio-economic status and poorer educational resources, but they are also less efficient than schools from more advantaged contexts.

[20]Uruguay is placed in the third position in the three evaluated areas between all Latin American countries that participated in PISA 2012.

[21]"At Level 2 students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational model. Students at this level can employ basic algorithms, formulae, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results".For more details, see OECD (2013a)

[22]National Administration of Public Education (ANEP), "*Informe Ejecutivo Preliminar Uruguay en PISA 2012*". Available at http://www.anep.edu.uy/anep/index.php/presentaciones-2012

[23]Schools are classified into five levels of socio-economic context based on the quintile distribution of the average socio-economic background of the students who attend to these schools (the average ESCS PISA index for each school). Levels are defined as 'Very unfavourable' (the bottom quintile), 'Unfavourable', 'Medium', 'Favourable' and 'Very favourable' (the top quintile).

### 2.3.2 PISA database

PISA is the only public source of data available for Uruguay that provides appropriate information about the academic results of students in secondary education schools -measured by objective test- and that provides contextual information about the students and schools. Moreover, the recent publication PISA allows us to obtain timely results for the current national educational debate.

PISA 2012 is the fifth edition of an initiative promoted by the OECD as of the late 1990s assessing 15-year-old students. The assessment focuses on measuring the extent to which students are able to apply their knowledge and skills to fulfil future real-life challenges rather than evaluating how they have mastered a specific school curriculum. It is a cross-curricular assessment which emphasis is on the mastery of processes, the understanding of concepts and the ability to function in various situations within each domain. The evaluation addresses three knowledge areas: reading, mathematical and scientific literacy, and each wave tests in depth a major domain. In 2000 and in 2009 the major domain was reading, in 2003 it was mathematics, in 2006 science and finally, in 2012, it is again mathematics. The measurement of student's abilities or skills is measured through the Rasch item response theory, from which a continuous scale score for each test is obtained with a mean score of 500 and standard deviation of 100 among OECD countries.

In addition to academic achievement data, the PISA database contains a vast amount of contextual information about students, their households and the schools they attend. Additionally, the database provides information through synthetic indexes, elaborated by OECD experts, by clustering responses to related questions provided by students and school's principals. The advantage of working with these indexes is that they have been constructed considering both theoretical and empirical studies, and have therefore been extensively tested at the international level (OECD, 2013b).

The 2012 PISA cycle is the fourth wave in which Uruguay has taken part, and it assessed 5,315 students from 180 public and private schools. For the purposes of this research, this database was refined. Firstly, as we are focused in the public sector, we eliminate private schools. Secondly, we eliminate schools which only offer basic secondary education (1st, 2nd and 3rd year of high school) or only offer upper secondary education (4th, 5th and 6th year of high school). The cut-off age between the two cycles in Uruguay is just 15 years old and, since PISA evaluates students of this age, those students attending schools where only basic secondary education is offered are inevitably repeaters and, on the contrary, students attending schools where only upper secondary education is offered are all non repeaters. As a result, in schools where only basic secondary education is imparted, 100% of the assessed students in PISA are repeaters in at least one previous course and, in those schools where only upper secondary education is imparted, 100% of the assessed students are on the right course. Therefore, these institutions are not comparable when estimating the production frontier.

This debugging based on the education levels offered in schools implies removing of the analysis of almost all public schools located in Montevideo, since these type of schools are almost all located in the capital. On the contrary, in the interior of the country secondary schools offer both cycles. In sum, this analysis is carried out for 71 mixed public schools (which provide both cycles of secondary education) located in the interior of the country. Therefore, as the context in the capital and outside it are different, the results from this research should be interpreted as a first approach to the problem of educational efficiency in public secondary schools and in the future it would be interesting to consider this problem in the PISA sample design in order to have comparable information for the whole country.

### 2.3.3 Relevant variables

It is difficult to empirically quantify the education received by an individual, especially when the focus is on analysing its quality beyond the years of education acquired. However, there is a consensus in the literature about considering that educational outputs can be approximated by the results obtained in standardized test, as they are difficult to forge and, above all, are taken into account by parents and politicians when making decisions on education (Hoxby, 1999). In fact, Hanushek (1986) found that two thirds of the educational studies use tests results as measures of educational outputs. In this research, as PISA 2012 is focused in mathematics, we selected the school average result in mathematics (Maths) as the output of the educational process[24].

Regarding educational inputs, three variables were selected taking into account the educational production function in Equation 1, which represent the classical inputs required to carry out the learning process (raw material, physical and human capital):

- ESCS (economic, social and cultural status): is an index developed by the PISA analysts to indicate the student socio-economic status. It therefore represents the "raw material" to be transformed through the learning process[25]. It is the result of running a categorical principal component analysis with three variables: the highest occupational status of either parent (HISEI), the highest educational level of either parent measured in years of education (PARED), and finally an index of home possessions (HOMEPOS)[26].
- SCMATEDU (school educational resources): is an index of the quality of educational resources in the school. It is therefore associated with the physical capital. It is elaborated from the responses by principals to seven questions related to the scarcity or lack of laboratory equipment, institutional materials, computers, Internet, educational software,

[24]The DEA technique can deal with multi-outputs problems but, since we apply the II-DEA technique described in the preceding chapter, in order to maintain consistency with the Monte Carlo experiment results we decided to only include one educational output. In future research, it would be interesting to extend the analysis to a multi-outputs context and compare the results with those of the present study.

[25]Both the ESCS index and the clustered variables are standardized with mean to zero and standard deviation equal to one across equally weighted OECD countries.

[26]For further details, see OECD (2013b).

library materials and, finally, audiovisual resources. The higher the index, the better the quality of the school's material resources.

- PROPCERT (proportion of fully certified teachers): this index reflects the quality of teachers, and therefore the school's human capital. The index is constructed by dividing the total number of certified teachers in the schools (with a teaching degree)[27] by the total number of teachers in the school. This variable is especially relevant in the case of Uruguay since not all teachers have received the teaching training required to qualify as teachers.

To ensure a correct DEA model specification, it is necessary to verify the monotonicity assumption, that is, all selected inputs must show a non-negative correlation with the output. Table 2.1 presents the bivariate correlations of the selected output and inputs where all correlations are positive.

As it was set in the introduction, the endogeneity problem is frequently observed in education and, particularly, in the case of the public secondary education sector in Uruguay, where we suspect that schools efficiency is positive and highly correlated with the school socio-economic level. To deal with this issue, we apply the II-DEA proposed in the first chapter, for which we need to find a good instrumental input for the school socio-economic background.

In order to know if it is necessary to apply the II-DEA we use the heuristic exposed in section II to identify potential endogenous inputs. From Figure 2.2, we can observe that SCMATEDU is classified as exogenous or negative correlated with the true efficiency term and PROPCERT seems to be low positive correlated with the true efficiency. However, the school socio-economic level (ESCS) appears to be a high positive endogenous input. As a result, we decide to instrument it and correct our estimations.

In order to apply the II-DEA the first step is to find a good instrument, which is not easy at all. As said in section 2.2, a good instrument should be correlated with the endogenous input (ESCS) but uncorrelated with the true efficiency. In empirical applications this means that there should not exist a clear causal relationship between the instrumental input variable and the output variable (*Maths*). Following this, we find an instrumental input that fulfils both conditions, the "*Percentage of students in the school that have had access to Internet before thirteen years old*" (ACCINT hereafter). The correlation between this variable and the school socio-economic levels is 0.2 (similar to the correlation assumed in the Monte Carlo experiment in the previous chapter). Furthermore, there is not clear evidence in the literature that just having access to Internet or TICs leads to better academic results per se; the effects will depend on how they are used and on parental monitoring and supervision (Angrist and Lavy 2002, Fuchs et al. 2004, Goolsbee and Guryan 2006). We apply the proposed heuristic to detect endogenous inputs with this new data set (Maths, ACCNT, SCMATEDU and PROPCERT)[28] and we find

[27] Certified teachers in Uruguay are required to complete a four-year degree at the *Instituto de Profesores Artigas* (IPA), a higher education institution which provides specialized secondary teacher training.

[28] We run the II-DEA in order to compute the Pearson's correlation coefficients between inputs and the estimated

that in this case, all inputs are classified as exogenous inputs or negative correlated with the efficiency term (Figure 2.3).

Finally, regarding the explanatory variables (Z vector in Equation (2.6)) of the estimated efficiency scores considered in the second stage, based on international evidence we select fourteen variables associated with students and schools that could be associated with technical efficiency. These variables reflect not only students and schools characteristics, but also some key aspects of management, school organization and the teaching-learning processes conducted in the classroom.

- *TECHVOC*: dummy variable that takes value one if the institution is a vocational technical school.
- *RURAL*: dummy variable that takes value one if the school is located in a town with less than 3,000 inhabitants.
- *SCHSIZE*: total number of students enrolled in the school.
- *PCTGIRL*: the percentage of female students in the school.
- *ICTSCH*: an index developed by PISA analysts that reflects the Information, Communication and Technology (ICT) availability at the school. It is elaborated from student responses to five questions regarding the availability at school of a desktop computer, a portable laptop, Internet connection, a printer, and a USB (memory) stick. The higher the index, the more ICT resources available at schools.
- *PCTCORRECT*: the percentage of students assessed in the school who are in the academic year that a 15-year student should really be in. This variable reflects the grade retention policy, and it is an important focus of attention in current educational discussions because there is no consensus about its net effect on educational results.
- *ANXMAT*: the index of mathematics anxiety is constructed by PISA analysts using student responses about the level of agreement to five statements when they are asked to think about studying mathematics[29]: 'I often worry that it will be difficult for me in mathematics classes'; 'I get very tense when I have to do mathematics homework'; "I get very nervous doing mathematics problems'; 'I feel helpless when doing a mathematics problem' and ; 'I worry that I will get poor grades in mathematics. The higher the index, the more the anxiety observed in the student.
- *PCTMATHEART*: percentage of students in the school that have answered yes to the statement 'When I study for a mathematics test, I learn as much as I can off by heart'. This variable reflects the learning skills acquired along the student's academic life.

efficiency scores.

[29] The levels of agreement are 'strongly agreed', 'agreed', 'disagreed' or 'strongly disagreed'.

- *TEACHGOAL*: percentage of students in the school that have reported that the teacher sets clear goals in 'every lesson' or 'most lessons'. This variable provides information about the teaching practices in the classroom.

- *TEACHCHECK*: percentage of students that have reported that the teacher makes questions to check students understanding 'every lesson' or 'most lessons'. Again, this variable inform about the teaching practices in the classroom.

- *HINDTEACH*: is a dummy variable that takes value one when the school's principal perceives that the learning of students is hindered 'a lot' or 'to some extent' by the presence of teachers not being well prepared for classes.

- *TEACHMORAL*:is a dummy variable that takes value one when the school's principal answers 'Strongly agree' or 'Agree' to the statement: 'The morale of teachers in this school is high'.

- *RESPCUR*: the index of the school responsibility for curriculum and assessment was constructed by PISA analysts from the principals answers about the responsibility that different stakeholders have related to four items: i) establishing student assessment policies; ii) choosing which textbooks are used; iii) determining course content; and iv) deciding which courses are offered. The ratio of the number of responsibilities that 'principals' and/or 'teachers' have for these four items to the number of responsibilities that 'regional or local education authority' and/or 'national education authority' have for these four items was computed. The higher index value, the relatively more responsibility for schools than for local, regional or national education authorities.

- *RESPRES*: the index of school responsibility for resource allocation was constructed by PISA analysts from the principals answers about the responsibility related to i) selecting teachers for hire; ii) firing teachers; iii) establishing teachers' starting salaries; iv) determining teachers' salary increases; v) formulating the school budget; and vi) deciding on budget allocations within the school. The ratio of the number of responsibilities that 'principals' and/or 'teachers' have for these six items to the number of responsibilities that 'regional or local education authority' and/or 'national education authority' have for these six items was computed. The higher index value, the relatively more responsibility for schools than for local, regional or national education authorities.

Table 2.2 presents the main descriptive statistics of all selected variables: output, inputs and contextual variables.

## 2.4 Results

### 2.4.1 First stage results

Table 2.3 and Figure 2.4 illustrate the distribution of efficiency scores, , estimated by the output-oriented II-DEA model under VRS. Results show that only 15.5% of the schools behave efficiently. On average, educational results could be increased by 17% given the available resources. We find that 25% of the evaluated schools could increase their academic achievements by up to 10% if they were fully efficient and 22.5% of the schools could raise their educational results between 10% and 20%. Moreover, a quarter of the evaluated schools could improve their outcomes by 20% to 30% with their current inputs; while one in ten schools could improve their results by over 30% to reach the frontier.

We also present the results from the conventional DEA model under the endogenous scenario (i.e. using the ESCS as an input) in order to compare them with those obtained using the II-DEA. We observe that in the first scenario not only the average efficiency is overestimated, but the distribution of all estimated efficiency scores is shifted to the left (Figure 2.4). These results are consistent with those arising from the Monte Carlo experiment presented in Chapter 1. As a result, potential improvements in public schools educational outcomes are considerably lower when we do not take into account the endogeneity problem, compared to those resulting from the II-DEA estimation. Table 4 provides the mean estimated efficiency scores with both models (DEA and II-DEA) and the mean ESCS by quintiles according to the endogenous input level (the school socio-economic context), the estimated efficiency score using the conventional DEA model (dhat-end) and the estimated efficiency score using the II-DEA approach (dhat-inst). We also compute the estimated bias for each school as the absolute difference between the estimated efficiency score from the II-DEA model and the estimated efficiency score from the conventional DEA

$$bias_i = \hat{\varphi}_{i,II-DEA} - \hat{\varphi}_{i,DEA} \tag{2.8}$$

This complementary analysis allows us to evaluate the differences between both specifications at different points of the distribution and hence, it is a helpful tool to locate the main damage of the endogeneity in this empirical application. From Table 2.4 we verify that the main effect of the endogeneity is on schools with a more disadvantaged socio-economic context. Schools at the bottom quintile according to the socio-economic context show the greatest bias (0.206) while for those schools located at the top quintile the bias is not significant at all. In fact, if we take into account the endogeneity issue in our estimations, schools from the bottom quintile could improve their results on average in 28.6% while, if we do not take into account this issue, the potential improvement reduces to only 8%. For schools at the top quintile (from most advantaged contexts) the potential improvements are similar in both scenarios, taking and not taking into account endogeneity (7.6% and 7.9%). We also observe that the better the average socio-economic schools context, the greater the estimated average school efficiency. On the contrary, if we analyse schools performance by quintiles according to the estimated efficiency score using DEA (under endogeneity) we cannot find an association between the mean school socio-economic level and the estimated efficiency scores. Again, these results corroborate what we found in the Monte Carlo experiment in Chapter 1.

Finally, in Table 2.5 we present three individual examples to illustrate the effect of taking or not into account the presence of endogeneity in the estimation of schools efficiency scores using DEA[30] . Schools A, B and C show similar average results in mathematics (Maths) and socio-economic context (ESCS) but schools A and C have considerably more resources (inputs) in terms of school's educational material (SCMATEDU) and the proportion of certified teacher at the school (PROPCERT) than school B. In other words, school A and C are really inefficient schools compared with school B. However, as we discussed in Chapter 1, under the presence of positive and high endogeneity conventional DEA misidentifies inefficient units with low levels of the endogenous input (ESCS in our case). From Table 2.5 we confirm this result. The DEA estimated efficiency scores for schools A and C are 1.036 and 1.045 respectively, which implies that they only could improve their mathematics results in around 4%. Conversely, when we take into account the endogeneity problem and estimate efficiency score using the II-DEA method, they are correctly identified as highly inefficient schools and in this case they could increase their results in almost 25% (21 percentage points more than with the conventional DEA estimation).

These findings should be taken into account specially if DEA is conducted with the purpose of benchmarking or informing performance-based reforms, since there would be some schools identified as benchmarks when they actually are not (e.g. schools A and C). Moreover, this misidentification is also detrimental if we are focused in exploring the explanatory variables of efficiency, because we will be trying to explain a dependent variable (the estimated efficiency score) that significantly differs from the true efficiency level. Therefore, the identified associations in the second-stage will not reflect the true ones[31].

### 2.4.2 Second stage results

We regress the II-DEA estimated efficiency scores over the contextual variables using four model specifications: the truncated regression with bootstrap proposed by Simar and Wilson (2007), the conventional Tobit, the Tobit regression with bootstrap and, finally, the OLS model with bootstrap. Results are shown in Table 2.6.

The first conclusion from the comparative analysis of the four specified models is that there are only minor discrepancies between the results. The sign and significance of almost all variables are the same in all models, implying that the educational policy recommendations derived from them would be basically the same, adding robustness to the findings discussed above. Taking into account this general conclusion we will consider the specification proposed by Simar and

[30] In Appendix A we present the results for all analysed schools.

[31] In Appendix B we present the results of the second stage using estimated efficiency scores from the DEA and the II-DEA where we corroborate that the results are totally different which also implies radically different educational public policy recommendations.

Wilson (2007) as a reference to discuss the results.

Firstly, there is a set of variables that do not affect efficiency scores. The first variable showing no significant effect is the dummy variable that indicates whether the school is a secondary school or a technical school (TECHVOC). Uruguayan public schools have on average better average academic results than technical schools. The results of this research show that these schools perform better due to higher initial input endowments and not due to a better management of them. In the same vein, school location does not seem to affect the efficiency (RURAL). Again, on average, schools in rural areas or small villages have worse educational outcomes than those located in bigger cities. The fact that the town size does not affect significantly the efficiency implies that the higher results are due to a greater allocation of educational resources and not to a better use of them.

Thirdly, the scale of production represented by the school size seems to slightly affect the schools efficiency (SCHSIZE). Larger schools which have on average better academic results, also have on average higher levels of educational inputs and better efficiency results. On the other hand, the percentage of female students at the school is not significant (PCTGIRLS), which indicates that the gender composition of the school does not affect efficiency. The availability of ICT in school (ICTSCH), or the fact that the teacher sets clear goals in lessons (TEACHGOAL) do not affect efficiency either.

Finally, none of the three variables associated with schools autonomy are significant. Decentralization of budget allocation decision, curriculum design or evaluation policies does not affect the schools efficiency. This is an interesting finding, since the decentralization issue is part of most current education discussions. International evidence shows that decentralization is successful in countries where there is also a school accountability practice properly regulated and with standardized criteria (Hanushek et al. 2013; OECD 2013b). This is not the case of Uruguay, where there is great heterogeneity in accountabilities and where, in many cases, there is not even a systematic way of presenting them. Therefore, the results of this research could be associated with this international evidence, which points out that decentralization would only have positive effects on improving academic results if it is carried out accompanied by an appropriate accountability system. Another possible interpretation of this result lies in the fact that the autonomy indexes were computed from the principals' responses and their perceived autonomy and therefore might not be reflecting the true degree of autonomy they actually have. In Uruguay, public high schools generally have low levels of autonomy; however, the variables included in this analysis show a certain degree of variance (Table 2.2), which could suggest some distortion between reality and principals' perceptions regarding their responsibility and autonomy.

On the contrary, there is a group of variables associated with student's characteristics and teaching practices that are significant and show the expected sign. First, the percentage of students that are in the right year (PCTCORRECT) appears to be a positive and significant driver of efficiency. This result calls into question the adequacy of current Uruguayan grade retention policies at all levels of the education system. Therefore, it would perhaps be better to attempt to identify younger (primary) students who are at risk of repeating and provide them with additional support in order to prevent their retention.

Another variable associated with students that has a significant negative effect on the school educational efficiency is the degree of anxiety of students to mathematics (ANXMAT). There is international evidence that supports this behaviour can be induced by the attitudes and expectations of teachers and parents (Zavaslvsky 1994). Thus it would be desirable to work on reducing student anxiety when solving math problems (and other disciplines) both in the classroom and at home. This means, not only to work at school but also to foster greater families commitment to support students work at home. Although this research is focused in secondary education, such practices should be encouraged from the beginning of the student's academic life in previous cycles, when it is most effective to impact on their non-cognitive skills (Heckman and Kautz, 2013). Thirdly, studying for mathematics tests by heart (PCTMATHEART) also has a negative impact on efficiency. This variable reflects the students study skills acquired along their academic life and, as in the previous case, this ability could be associated with classroom teaching techniques adopted by teachers. Thus, this factor should be considered by school managers and educational authorities, especially in the early stages of the learning process when students are assimilating the learning techniques to be used throughout their academic life.

In addition, the fact that the teacher checks student understanding in lessons (TEACHCHECK) positively affects efficiency and thus this practice should be promoted in order to obtain better results. Fifthly, efficiency is positive and significantly affected by the fact that the school principal perceives that the learning of students is hindered by the presence of teachers not being well prepared for classes (HINDTEACH). Therefore, this result suggests that principals who perceive greater shortage of qualified teachers manage school resources in a better way, obtaining most of their actual educational resources. This finding suggest that, it is not only necessary to have prepared teachers to produce education but also crucial to make a better use of them. Finally, teacher's morale (TEACHMORAL) has a positive impact on efficiency. Therefore, this study provides evidence that if one wants to improve the academic performance in Uruguayan public secondary schools through efficiency it is necessary to foster an incentive based teaching career or other similar tools in order to increase the morale of teachers in schools.

## 2.5 Concluding remarks

Modern countries agree about the need and importance of having a more and better educated population in order to ensure economic growth based on the high productivity of a skilled labour force. The high percentage of public expenditure invested in education reflects this conviction. During the last decade, the Uruguayan government has made a huge effort to increase educational resources; however, academic results have not improved. On the contrary, public education system (especially public secondary education) is in a deep crisis and the current educational national debate mainly focuses on the need to put more resources into the system instead of exploring how to make better use of available inputs, i.e., how to achieve a more efficient education system. This is the main focus of this chapter: to explore the sources of the inefficiency in Uruguayan public secondary schools and to provide new evidence to the national educational debate.

To do that, we use one of the most popular methods applied in the literature: the semi-parametric two-stage model. Under this approach, we estimate efficiency scores using a DEA model and, in a second stage, we regress these efficiency scores over different contextual variables related to students and schools. Two-stage models differ mainly in the regression model specified in the second stage to explain efficiency scores. However, as we discussed in the previous chapter, these models are affected by a major and recurrent issue in educational production processes which is frequently overlooked when practitioners apply DEA (that is, the first stage of the semi-parametric models), namely, the presence of endogeneity.

The Uruguayan public secondary education sector is a very illustrative case of this issue, for we observe that schools' socio-economic context is highly positive correlated with schools technical efficiency. As shown in Chapter 1, the presence of this high correlation severely biases DEA performance. To overcome this problem, we apply the proposed *Instrumental Input Data Envelopment Analysis* strategy in order to obtain more reliable estimates. In this respect, we also aim to investigate whether taking or not into account this problem really matters in empirical terms and which are its implications for public policy recommendations. To explore that, we provide the results obtained using the conventional DEA in the first step, and we compare them with those from the proposed II-DEA method.

Our first results evidence that the evaluated public secondary schools could increase their results in mathematics on average by 17% if they were fully efficient. If we do not consider the problem of endogeneity in our estimations, this potential improvement reduces to only 10%. We also observe that not only the average efficiency is overestimated, but also the distribution of all estimated efficiency scores is shifted to the left. In addition, we corroborate that the greatest damage of the endogeneity problem in DEA efficiency estimates is in those schools with lower levels of the endogenous input, i.e. schools from more disadvantaged socio-economic contexts. Schools at the bottom quintile according to their socio-economic context show the greatest bias (0.203) while schools located at the top quintile do not show almost any bias. In fact, if we take into account the endogeneity issue in our estimations, schools from the bottom quintile could improve their results on average in 28% while if we do not take into account this issue this potential improvement reduces to only 8%. These findings evidence the importance of taking into account the endogeneity issue in the efficiency estimation basically for two reasons. Firstly, because under endogeneity DEA misidentifies the most inefficient schools which are in fact the first schools which should work to improve their results. Secondly, if the efficiency scores are biased, we cannot find the real sources or explanatory factors of the true inefficiencies. The endogeneity problem associated with the socio-economic context of schools in terms of efficiency implies greater inequality as the most inefficient public schools are those with students with fewer opportunities and more unfavourable contexts. Therefore, failing in considering this problem in the estimation of educational efficiency has deep implications in terms of public educational policy.

Additionally, promoting teaching and learning techniques to reduce student's anxiety and improve self-confidence in solving mathematics problems and to discourage students from studying mathematics by heart would produce significant improvements in academic outcomes. Although these practices should be promoted mainly in the classroom, commitment and parents support at home is also needed to ensure the effectiveness of these practices. Therefore, educational policies should also try to increase the involvement of families in the learning process of their children. Although this research is focused in secondary education, most of the policies and practices suggested above should be encouraged from the beginning of the students' academic life in previous cycles, when it is most effective to influence their non-cognitive skills.

Another relevant finding from our estimations is that school efficiency is positive and significantly affected by the fact that the school principal perceives that the learning of students is hindered by the presence of teachers not being well prepared for classes. Thus, it seems that principals who perceive a shortage of prepared teachers make a better use of them. Finally, teacher's morale is a key factor to improve efficiency in public secondary schools. In this sense, it would be appropriate to promote teacher compensation systems that establish teacher incentives (professional career, monetary incentives, etc.) linked to their performance (measured by multiple tools).

It is noteworthy that these results should be interpreted with some caution. Although a priori all these recommended practices do not generate additional direct costs, some of them could lead to associated indirect costs for some individuals involved in the educational process. Therefore, it is essential to achieve a general commitment by all stakeholders in the educational process (authorities, teachers, students, families and society) to ensure effective improvements of educational efficiency in public schools.

Finally, the results of this research should be interpreted as a first milestone to the efficiency issue in Uruguayan public secondary schools and, therefore, further research is necessary in this direction. Having access to national databases adapted to the Uruguayan reality seems to be the logical first step to expand the scope of this research (both in geographical terms and to other levels of the education system).

## 2.6 References

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

Alexander, W. R. J., Haug, A. A., and Jaforullah, M. (2010). A two-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools. Journal of Productivity Analysis, 34(2), 99-110. Angrist, J., and Lavy, V. (2002). New Evidence on Classroom Computers and Pupil Learning*. The Economic Journal, 112(482), 735-765.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-1092.

Banker, R. D., and Natarajan, R. (2008). Evaluating contextual variables affecting productivity using data envelopment analysis. Operations Research, 56(1), 48-58.

Barro, R. J., and Lee, J.-W. (2012). A new data set of educational attainment in the world, 1950–2010. Journal of Development Economics, 104, 184–198.

Battese, G. E., and Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of econometrics, 38(3), 387-399.

Battese, G. E., and Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis, 3(1-2), 153-169.

Battese, G. E., and Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics, 20(2), 325-332.

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

Crespo-Cebada, E., Pedraja-Chaparro, F., and Santín, D. (2013). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. Journal of Productivity Analysis. doi: 10.1007/s11123-013-0338-y

Charnes, A., Cooper, W., Lewin, A. Y., and Seiford, L. M. (1997). Data envelopment analysis theory, methodology and applications. Journal of the Operational Research Society, 48(3), 332-333.

Charnes, A., Cooper, W. W., and Rhodes, E. (1981). Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. Management Science, 27(6), 668-697.

De Jorge, J., and Santín González, D. (2010). Determinantes de la eficiencia educativa en la Unión Europea. Hacienda pública española(193), 131-155.

Fernández, T. (2009). La desafiliación en la educación media en Uruguay. Una aproximación con base en el panel de estudiantes evaluados por PISA 2003. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación (REICE), 7(nº4).

Goolsbee, A., and Guryan, J. (2006). The impact of Internet subsidies in public schools. The Review of Economics and Statistics, 88(2), 336-347.

Grosskopf, S., Hayes, K. J., Taylor, L. L., and Weber, W. L. (1997). Budget-constrained frontier measures of fiscal equality and efficiency in schooling. Review of Economics and Statistics, 79(1), 116-124.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. Journal of human Resources, 351-388.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. Journal of economic literature, 24(3), 1141-1177.

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

Hanushek, E. A., and Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Journal of Economic Growth, 17(4), 267-321.

Heckman, J. J., and Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition. National Bureau of Economic Research, No. w19656.

Hoff, A. (2007). Second stage DEA: Comparison of approaches for modelling the DEA score. European Journal of Operational Research, 181(1), 425-435.

Hoxby, C. M. (1999). The productivity of schools and other local public goods producers. Journal of Public Economics, 74(1), 1-30.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. National Bureau of Economic Research, No. w7867.

Levin, H. M. (1974). Measuring efficiency in educational production. Public Finance Quarterly, 2(1), 3-24.

Liu, J. S., Lu, L. Y., Lu, W.-M., and Lin, B. J. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. Omega, 41(1), 3-15.

Llambí, C., Perera, M., and Messina, P. (2009). Desigualdad de oportunidades y el rol del sistema educativo en los logros de los jóvenes uruguayos. Investigación financiada por el Fondo "Carlos Filgueira", del Programa Infancia, Adolescencia y Familia-Ministerio de Desarrollo Social, Uruguay.

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

McCarty, T. A., and Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In H. O. Fried, C. A. Knox Lovell and S.

Smith (Eds.), The measurement of productive efficiency: Techniques and applications (pp. 271-287). New York: Oxford University Press.

OECD (2013a), PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing.

OECD (2013b), PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I), PISA, OECD Publishing.

OECD (2012). PISA 2009 Technical Report. PISA, OECD Publishing.

Perelman, S., and Santin, D. (2011). Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. Education Economics, 19(1), 29-49.

Perera, M., and Llambí, C. (2009). La Función de Producción Educativa: el posible sesgo en la estimación de efectos "institucionales" con los datos PISA. El caso de las escuelas de Tiempo Completo. Documento de Trabajo DT, 3, 2009.

Ramalho, E. A., Ramalho, J. J., and Henriques, P. D. (2010). Fractional regression models for second stage DEA efficiency analyses. Journal of Productivity Analysis, 34(3), 239-255.

Ray, S. C. (1991). Resource-use efficiency in public schools: A study of Connecticut data. Management Science, 37(12), 1620-1628.

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

Simar, L., and Wilson, P. W. (2011). Two-stage DEA: caveat emptor. Journal of Productivity Analysis, 36(2), 205-218.

Webbink, D. (2005). Causal effects in education. Journal of Economic Surveys, 19(4), 535-560.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. Education Economics, 9(3), 245-268.

Xue, M., and Harker, P. T. (1999). Overcoming the inherent dependency of DEA efficiency scores: a bootstrap approach. Working Paper, Wharton Financial Institutions Center, University of Pennsylvania.

Zaslavsky, C. (1994). Fear of math: How to get over it and get on with your life: Rutgers University Press.

## 2.7 Figures and Tables

Figure 2.1: Pupils distribution into public and private schools by socio-economic level in PISA 2012

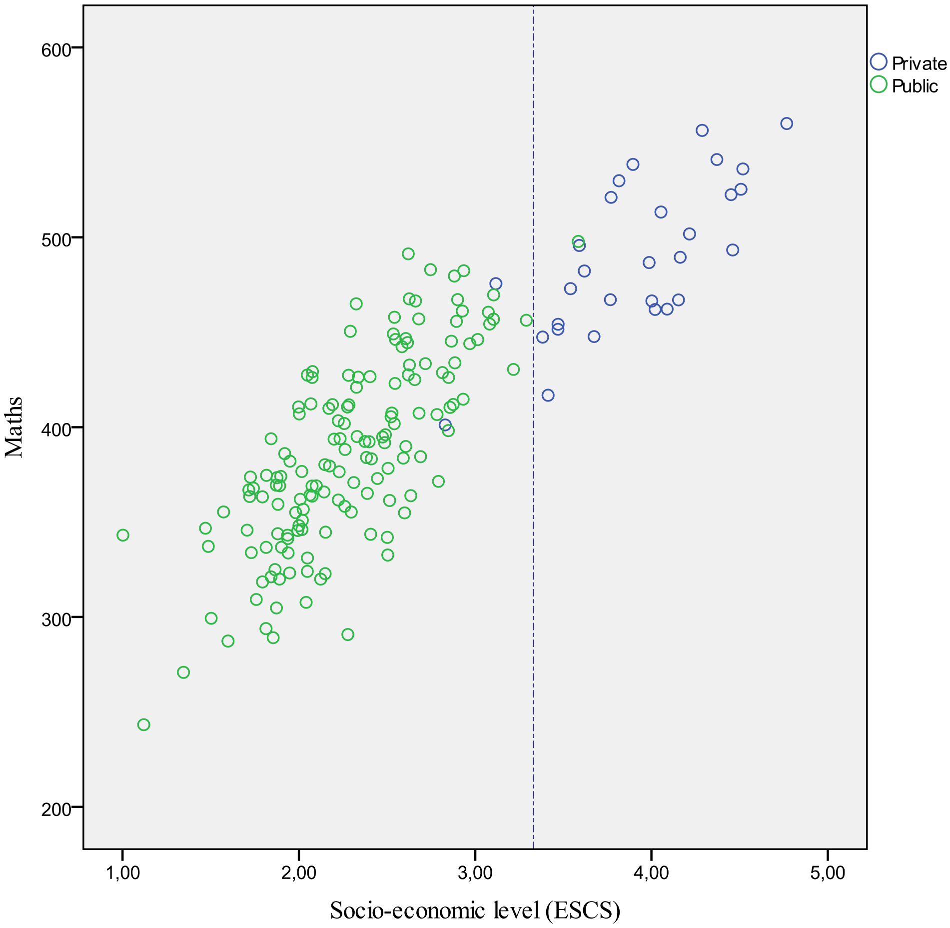
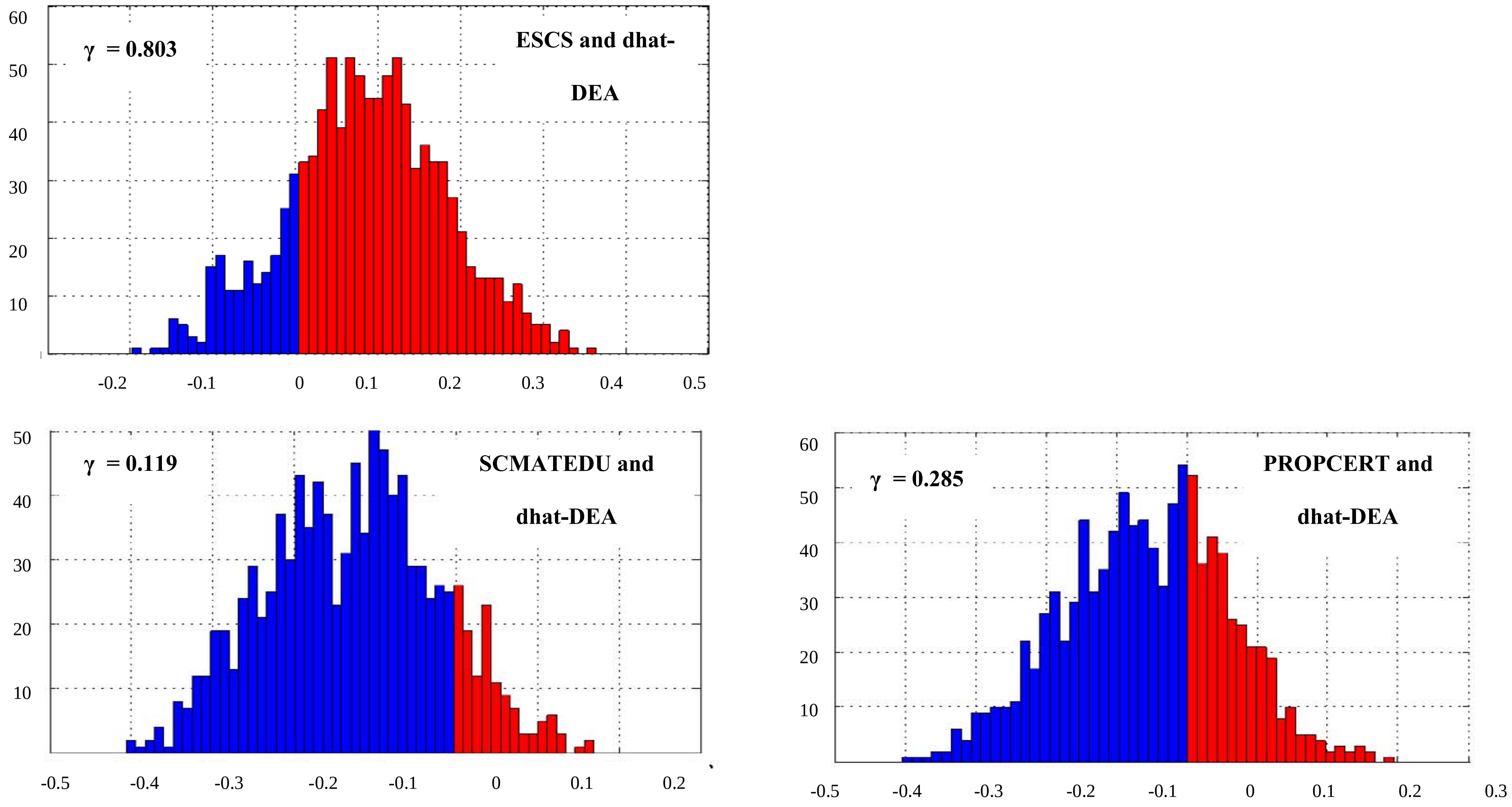


Figure 2.2: Input classification using the proposed heuristic to identify endogenous inputs using conventional DEA



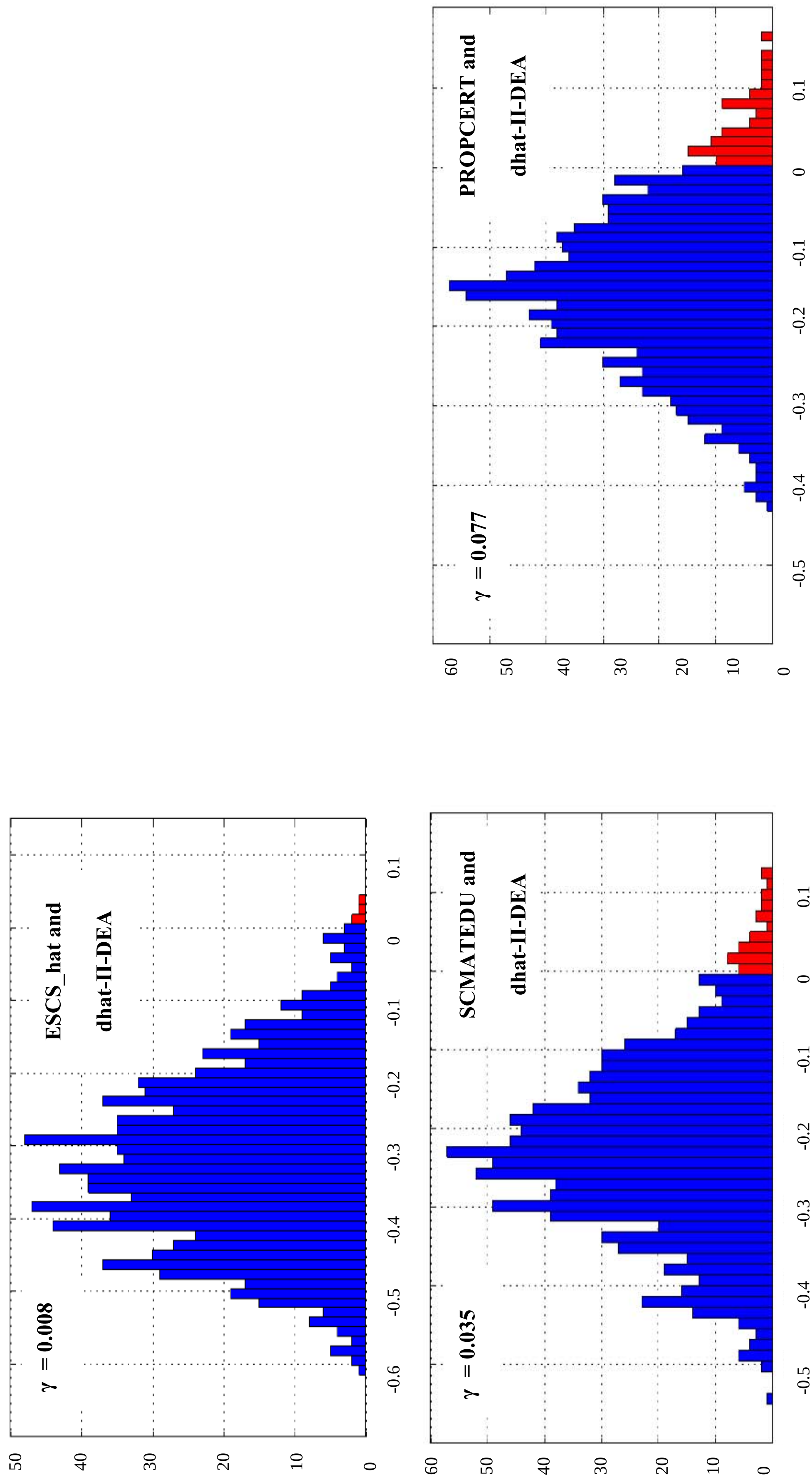


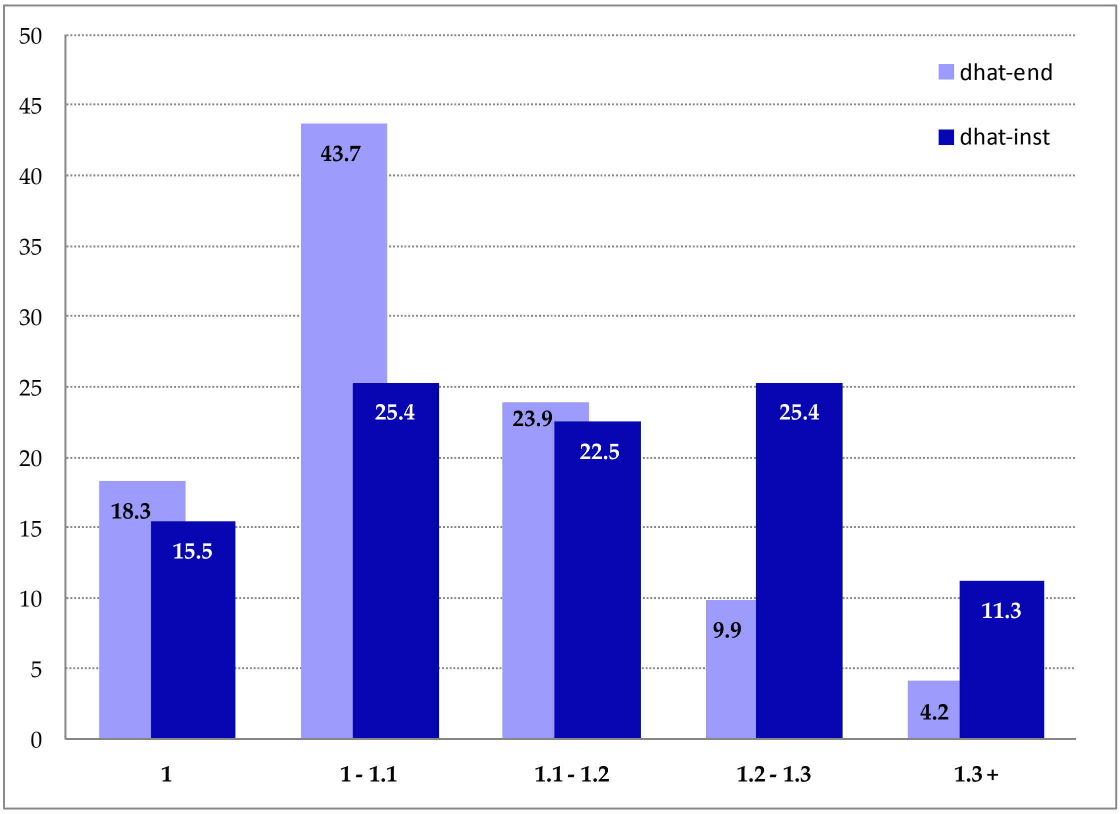

Figure 2.3: Input classification using the proposed heuristic to identify endogenous inputs using II-DEA

Table 2.1: Bivariate correlations between inputs and *Maths*

| | ESCS | SCMATEDU | PROPCERT |
|---|---|---|---|
| Maths | 0.693 | 0.083 | 0.101 |

*Note*: Pearson's correlation coefficients // Sample size = 71 // All correlations are significant at 0.01%.

Figure 2.4: Estimated efficiency scores distribution (DEA-BBC)




*Note*: Values equal to one represent full efficient units. The higher the score value, the greater inefficiency.

Table 2.2: Descriptive statistics of output, inputs and explanatory variables of efficiency

| Variable | Mean | Std. Dev. | Min | Max | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|---|
| Maths | 382.7 | 44.2 | 270.9 | 466.5 | 354.9 | 388.2 | 411.7 |
| ESCS | 2.20 | 0.42 | 1.35 | 3.29 | 1.88 | 2.08 | 2.53 |
| SCMATEDU | 4.50 | 1.11 | 2.30 | 6.57 | 3.72 | 4.42 | 5.05 |
| PROPCERT | 0.52 | 0.20 | 0.15 | 0.94 | 0.35 | 0.52 | 0.67 |
| ACCINT | 0.86 | 0.08 | 0.60 | 1.00 | 0.81 | 0.86 | 0.90 |
| *TECHVOC*[a] | 0.32 | 0.47 | 0 | 1 | 0 | 0 | 1 |
| *RURAL*[a] | 0.13 | 0.34 | 0 | 1 | 0 | 0 | 0 |
| SCHSIZE | 910 | 645 | 74 | 3,292 | 442 | 797 | 1,281 |
| *PCTGIRL* | 0.52 | 0.15 | 0.11 | 0.87 | 0.41 | 0.54 | 0.61 |
| *ICTSCH* | 3.49 | 0.35 | 2.58 | 4.20 | 3.25 | 3.48 | 3.79 |
| *PCTCORRECT* | 0.55 | 0.25 | 0 | 1 | 0.39 | 0.57 | 0.75 |
| ANXMAT | 3.81 | 0.29 | 3.10 | 4.71 | 3.59 | 3.81 | 3.99 |
| *PCTMATHEART* | 0.17 | 0.08 | 0.00 | 0.39 | 0.11 | 0.17 | 0.23 |
| *TEACHGOAL* | 0.48 | 0.10 | 0.22 | 0.67 | 0.42 | 0.49 | 0.55 |
| *TEACHCHECK* | 0.46 | 0.10 | 0.21 | 0.67 | 0.38 | 0.47 | 0.54 |
| *HINDTEACH*[a] | 0.46 | 0.50 | 0 | 1 | 0 | 0 | 1 |
| *TEACHMORAL*[a] | 0.25 | 0.44 | 0 | 1 | 0 | 0 | 1 |
| RESPCUR | 1.28 | 0.29 | 1.00 | 2.57 | 1.00 | 1.23 | 1.45 |
| RESPRES | 1.05 | 0.09 | 1.00 | 1.44 | 1.00 | 1.00 | 1.10 |

*Note*: (d) Dummy variables where the mean represents the proportion of schools in the reference category // References categories are: vocational technical school (TECHVOC); school located in rural are (RURAL);school's principal perceives teachers not being well prepared (HINDTEACH) and school's principal perceives high teachers morale (TEACHMORAL).

Table 2.3: Descriptives of the estimated efficiency scores (DEA and II-DEA)

| Efficiency | Mean | Std- Dev. | Min. | Max. | Q1 | Q2 | Q3 |
|---|---|---|---|---|---|---|---|
| dhat-end | 1.101 | 0.102 | 1.000 | 1.468 | 1.015 | 1.074 | 1.158 |
| dhat-inst | 1.167 | 0.149 | 1.000 | 1.640 | 1.023 | 1.137 | 1.258 |

Table 2.4: Mean ESCS, efficiency scores and bias by quintiles (DEA and II-DEA results)

| | Mean ESCS | Mean dhat-inst | Mean dhat-end | Mean \|Bias\| |
|---|---|---|---|---|
| **Quintiles by ESCS** | | | | |
| Bottom quintile | **1.68** | 1.286 | 1.079 | 0.206 |
| 4th quintile | **1.92** | 1.229 | 1.132 | 0.097 |
| 3rd quintile | **2.13** | 1.146 | 1.107 | 0.050 |
| 2nd quintile | **2.40** | 1.106 | 1.108 | 0.011 |
| Top quintile | **2.82** | 1.076 | 1.079 | 0.003 |
| **Quintiles by dhat-inst** | | | | |
| Bottom quintile | 1.88 | **1.386** | 1.213 | 0.174 |
| 4th quintile | 1.92 | **1.233** | 1.126 | 0.111 |
| 3rd quintile | 2.37 | **1.139** | 1.100 | 0.043 |
| 2nd quintile | 2.50 | **1.059** | 1.049 | 0.021 |
| Top quintile | 2.33 | **1.003** | 1.008 | 0.008 |
| **Quintiles by dhat-end** | | | | |
| Bottom quintile | 2.07 | 1.325 | **1.257** | 0.071 |
| 4th quintile | 2.29 | 1.202 | **1.139** | 0.075 |
| 3rd quintile | 2.24 | 1.115 | **1.070** | 0.059 |
| 2nd quintile | 2.27 | 1.107 | **1.026** | 0.082 |
| Top quintile | 2.13 | 1.076 | **1.000** | 0.076 |

Table 2.5: Outputs, inputs and estimated efficiency scores DMU's A,B and C

| SCH | MATH | ESCS | SCMATEDU | PROPCERT | ACCINT | dhat_end | dhat-inst | Bias |
|---|---|---|---|---|---|---|---|---|
| A | 367 | 1.72 | 4.4218 | 0.5170 | 0.86 | 1.0362 | 1.2478 | 0.2117 |
| B | 363 | 1.72 | 2.3016 | 0.2220 | 0.60 | 1.0000 | 1.0000 | 0.0000 |
| C | 368 | 1.74 | 4.6100 | 0.7310 | 0.86 | 1.0446 | 1.2467 | 0.2021 |

Table 2.6: Efficiency explanatory variables: alternative second-stage estimations

| Dependent variable: dhat | Truncated + bootstrap | | | Conventional Tobit | | | Tobit + bootstrap | | | OLS + bootstrap | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Coef** | **Std. Err.** | **z** | **Coef** | **Std. Err.** | **z** | **Coef** | **Std. Err.** | **z** | **Coef** | **Std. Err.** | **z** |
| *TECHVOC*[a] | 0.0097 | 0.057 | 0.17 | 0.0180 | 0.040 | 0.45 | 0.0180 | 0.050 | 0.36 | 0.0211 | 0.043 | 0.49 |
| *RURAL*[a] | -0.0062 | 0.074 | -0.08 | -0.0019 | 0.049 | -0.04 | -0.0019 | 0.062 | -0.03 | -0.0023 | 0.053 | -0.04 |
| SCHSIZE | -0.0001 | 0.000 | -1.81 * | 0.0000 | 0.000 | -2.17 ** | 0.0000 | 0.000 | -1.64 * | 0.0000 | 0.000 | -2.03 ** |
| *PCTGIRL* | 0.0249 | 0.165 | 0.15 | -0.0269 | 0.100 | -0.27 | -0.0269 | 0.137 | -0.20 | -0.0335 | 0.128 | -0.26 |
| *ICTSCH* | -0.0395 | 0.067 | -0.59 | -0.0637 | 0.044 | -1.43 | -0.0637 | 0.055 | -1.16 | -0.0651 | 0.050 | -1.31 |
| *PCTCORRECT* | -0.2898 | 0.117 | -2.47 ** | -0.3122 | 0.080 | -3.91 *** | -0.3122 | 0.099 | -3.15 *** | -0.2683 | 0.080 | -3.37 *** |
| ANXMAT | 0.2410 | 0.077 | 3.14 *** | 0.1500 | 0.054 | 2.80 *** | 0.1500 | 0.065 | 2.31 ** | 0.1557 | 0.057 | 2.76 *** |
| *PCTMATHEART* | 0.5081 | 0.268 | 1.89 * | 0.1967 | 0.176 | 1.12 | 0.1967 | 0.227 | 0.87 | 0.1824 | 0.194 | 0.94 |
| *TEACHGOAL* | 0.3965 | 0.253 | 1.57 | 0.2760 | 0.192 | 1.43 | 0.2760 | 0.222 | 1.24 | 0.2429 | 0.188 | 1.29 |
| *TEACHCHECK* | -0.5443 | 0.228 | -2.39 ** | -0.3084 | 0.162 | -1.91 * | -0.3084 | 0.193 | -1.59 | -0.2978 | 0.163 | -1.82 * |
| *HINDTEACH*[a] | -0.0873 | 0.039 | -2.24 ** | -0.1022 | 0.028 | -3.63 *** | -0.1022 | 0.034 | -3.01 *** | -0.0850 | 0.027 | -3.12 *** |
| *TEACHMORAL*[a] | -0.1056 | 0.049 | -2.13 ** | -0.0574 | 0.032 | -1.78 * | -0.0574 | 0.038 | -1.52 | -0.0609 | 0.031 | -1.97 ** |
| RESPCUR | -0.0962 | 0.064 | -1.50 | -0.0363 | 0.043 | -0.85 | -0.0363 | 0.057 | -0.63 | -0.0448 | 0.048 | -0.93 |
| RESPRES | 0.1902 | 0.199 | 0.95 | 0.3807 | 0.145 | 2.63 ** | 0.3807 | 0.196 | 1.94 * | 0.3127 | 0.174 | 1.80 * |
| _cons | 0.5361 | 0.423 | 1.27 | 0.7095 | 0.280 | 2.53 | 0.7095 | 0.341 | 2.08 | 0.7716 | 0.318 | 2.42 |
| /sigma | 0.0926 | 0.01 | 8.65 | 0.0999 | 0.01 | | 0.0999 | 0.01 | | 0.1001 | | |

*Note*: Left-truncated regression at value one // 2,000 bootstrap replications //'Coef' are the estimated coefficient. Negative values increase efficiency // Std.Err. are the robust standard error of the estimated coefficients // Sample size = 71 // $***p-value < 0.01$; $**p-value < 0.05$ ; $*p-value < 0.10$ // (d) Dummy variables where the mean represents the proportion of schools in the reference category // References categories are: vocational technical school (TECHVOC); school located in rural are (RURAL);school's principal perceives teachers not being well prepared (HINDTEACH) and school's principal perceives high teachers morale (TEACHMORAL).

## 2.8 Appendix A: Data and estimated efficiency scores for each school

Table 2.7: Database Uruguayan public schools and efficiency estimated scores

| ID_sch (A) | MATH (B) | ESCS (C) | SCMATEDU (D) | PROPCERT (E) | ACCINT (F) | dhat-inst (G) | dhat-end (H) | Mean Bias \|(G) - (H)\| |
|---|---|---|---|---|---|---|---|---|
| 7 | 363.46 | 1.72 | 2.3016 | 0.2220 | 0.60 | 1.0000 | 1.0000 | 0.0000 |
| 25 | 382.07 | 1.95 | 2.3016 | 0.1530 | 0.78 | 1.0000 | 1.0000 | 0.0000 |
| 27 | 410.63 | 2.00 | 4.4218 | 0.1530 | 0.88 | 1.0000 | 1.0000 | 0.0000 |
| 42 | 410.66 | 2.28 | 3.8979 | 0.2910 | 0.75 | 1.0000 | 1.0546 | 0.0546 |
| 44 | 411.74 | 2.28 | 2.8995 | 0.2340 | 0.83 | 1.0000 | 1.0053 | 0.0053 |
| 45 | 450.51 | 2.29 | 4.6100 | 0.4960 | 0.83 | 1.0000 | 1.0000 | 0.0000 |
| 47 | 426.32 | 2.34 | 3.3447 | 0.1530 | 0.94 | 1.0000 | 1.0000 | 0.0000 |
| 54 | 407.43 | 2.53 | 2.3016 | 0.4500 | 0.86 | 1.0000 | 1.0000 | 0.0000 |
| 55 | 457.89 | 2.54 | 4.4218 | 0.4280 | 0.89 | 1.0000 | 1.0000 | 0.0000 |
| 61 | 466.49 | 2.66 | 6.5680 | 0.5970 | 0.88 | 1.0000 | 1.0000 | 0.0000 |
| 63 | 433.47 | 2.72 | 2.8995 | 0.4280 | 0.85 | 1.0000 | 1.0000 | 0.0000 |
| 71 | 456.32 | 3.29 | 5.0526 | 0.9210 | 0.83 | 1.0090 | 1.0090 | 0.0000 |
| 23 | 386.07 | 1.92 | 4.2438 | 0.7100 | 0.67 | 1.0146 | 1.0492 | 0.0346 |
| 33 | 427.45 | 2.05 | 5.7539 | 0.6730 | 0.79 | 1.0230 | 1.0000 | 0.0230 |
| 35 | 412.26 | 2.07 | 2.6300 | 0.5080 | 0.93 | 1.0275 | 1.0275 | 0.0000 |
| 67 | 410.43 | 2.86 | 2.6300 | 0.7370 | 0.85 | 1.0277 | 1.0277 | 0.0000 |
| 56 | 446.22 | 2.55 | 4.8160 | 0.6850 | 0.87 | 1.0297 | 1.0286 | 0.0011 |
| 65 | 428.74 | 2.81 | 3.3447 | 0.7500 | 0.86 | 1.0298 | 1.0310 | 0.0013 |
| 64 | 406.67 | 2.78 | 5.3444 | 0.1560 | 0.91 | 1.0480 | 1.0480 | 0.0000 |
| 69 | 446.15 | 3.01 | 5.3444 | 0.4690 | 0.89 | 1.0526 | 1.0616 | 0.0089 |
| 60 | 432.76 | 2.63 | 4.8160 | 0.3900 | 0.91 | 1.0578 | 1.0000 | 0.0578 |
| 59 | 427.56 | 2.62 | 5.3444 | 0.3930 | 0.90 | 1.0689 | 1.0645 | 0.0044 |
| 38 | 393.71 | 2.20 | 2.8995 | 0.4120 | 0.79 | 1.0740 | 1.0142 | 0.0598 |
| 51 | 426.61 | 2.40 | 4.6100 | 0.5718 | 0.88 | 1.0750 | 1.0642 | 0.0109 |
| 29 | 407.00 | 2.00 | 4.2438 | 0.3480 | 0.84 | 1.0762 | 1.0762 | 0.0000 |
| 70 | 430.45 | 3.22 | 5.7539 | 0.6350 | 0.89 | 1.0859 | 1.1443 | 0.0584 |
| 11 | 374.63 | 1.82 | 4.8160 | 0.8300 | 0.71 | 1.0868 | 1.0023 | 0.0845 |
| 36 | 429.24 | 2.08 | 6.5680 | 0.7140 | 0.89 | 1.0905 | 1.0905 | 0.0000 |
| 68 | 414.74 | 2.93 | 4.0706 | 0.5718 | 0.95 | 1.0931 | 1.0896 | 0.0036 |
| 52 | 383.34 | 2.41 | 4.4218 | 0.2640 | 0.81 | 1.1112 | 1.0848 | 0.0264 |
| 40 | 401.94 | 2.26 | 3.7220 | 0.5710 | 0.83 | 1.1120 | 1.0457 | 0.0663 |
| 39 | 403.38 | 2.22 | 4.6100 | 0.3930 | 0.84 | 1.1233 | 1.0692 | 0.0542 |
| 37 | 409.87 | 2.17 | 5.0526 | 0.9440 | 0.82 | 1.1259 | 1.1330 | 0.0071 |
| 62 | 407.37 | 2.68 | 5.3444 | 0.9250 | 0.97 | 1.1331 | 1.1331 | 0.0000 |
| 41 | 388.24 | 2.26 | 4.4218 | 0.3480 | 0.87 | 1.1357 | 1.1563 | 0.0206 |
| 53 | 405.48 | 2.52 | 6.5680 | 0.5160 | 0.89 | 1.1373 | 1.1283 | 0.0090 |

Table 2.8: Database Uruguayan public schools and efficiency estimated scores (cont)

| ID_sch (A) | MATH (B) | ESCS (C) | SCMATEDU (D) | PROPCERT (E) | ACCINT (F) | dhat-inst (G) | dhat-end (H) | Mean Bias \|(G) - (H)\| |
|---|---|---|---|---|---|---|---|---|
| 22 | 374.07 | 1.90 | 4.8160 | 0.5980 | 0.77 | 1.1527 | 1.0739 | 0.0789 |
| 50 | 392.34 | 2.40 | 5.7539 | 0.4680 | 0.85 | 1.1563 | 1.1563 | 0.0000 |
| 66 | 398.18 | 2.85 | 5.0526 | 0.7770 | 0.94 | 1.1594 | 1.1427 | 0.0167 |
| 48 | 392.48 | 2.37 | 4.2438 | 0.5300 | 0.91 | 1.1609 | 1.0162 | 0.1447 |
| 12 | 393.92 | 1.84 | 5.0526 | 0.6640 | 0.97 | 1.1688 | 1.0121 | 0.1566 |
| 49 | 384.05 | 2.38 | 4.0706 | 0.5860 | 0.86 | 1.1776 | 1.1620 | 0.0156 |
| 5 | 355.35 | 1.57 | 4.6100 | 0.3210 | 0.76 | 1.1784 | 1.1784 | 0.0000 |
| 58 | 389.89 | 2.61 | 4.8160 | 0.7500 | 0.89 | 1.1857 | 1.0805 | 0.1052 |
| 34 | 364.27 | 2.06 | 4.2438 | 0.6670 | 0.77 | 1.2025 | 1.0527 | 0.1498 |
| 10 | 363.30 | 1.79 | 6.5680 | 0.4000 | 0.80 | 1.2064 | 1.2044 | 0.0020 |
| 46 | 355.33 | 2.30 | 3.3447 | 0.2970 | 0.86 | 1.2252 | 1.0817 | 0.1435 |
| 21 | 369.12 | 1.89 | 4.0706 | 0.5630 | 0.97 | 1.2305 | 1.0754 | 0.1551 |
| 17 | 373.42 | 1.88 | 4.8160 | 0.5490 | 0.94 | 1.2390 | 1.1558 | 0.0832 |
| 6 | 366.94 | 1.72 | 4.4218 | 0.5170 | 0.86 | 1.2447 | 1.0903 | 0.1544 |
| 8 | 367.87 | 1.74 | 4.6100 | 0.7310 | 0.86 | 1.2467 | 1.0446 | 0.2021 |
| 15 | 369.47 | 1.87 | 6.5680 | 0.5560 | 0.88 | 1.2478 | 1.0362 | 0.2117 |
| 24 | 323.20 | 1.95 | 3.7220 | 0.3220 | 0.72 | 1.2585 | 1.0000 | 0.2585 |
| 2 | 346.76 | 1.47 | 3.7220 | 0.2410 | 1.00 | 1.2585 | 1.2463 | 0.0122 |
| 30 | 350.96 | 2.02 | 4.0706 | 0.5080 | 0.83 | 1.2628 | 1.2873 | 0.0245 |
| 32 | 324.05 | 2.05 | 4.0706 | 0.3640 | 0.72 | 1.2762 | 1.2246 | 0.0516 |
| 16 | 304.82 | 1.87 | 4.2438 | 0.5718 | 0.65 | 1.2780 | 1.1818 | 0.0962 |
| 26 | 345.54 | 1.99 | 4.2438 | 0.5700 | 0.80 | 1.2798 | 1.2101 | 0.0697 |
| 13 | 321.18 | 1.84 | 3.3447 | 0.3480 | 0.74 | 1.2902 | 1.1604 | 0.1298 |
| 18 | 343.87 | 1.88 | 4.6100 | 0.3040 | 1.00 | 1.2967 | 1.1886 | 0.1080 |
| 28 | 348.14 | 2.00 | 4.3779 | 0.3960 | 0.88 | 1.2975 | 1.2975 | 0.0000 |
| 57 | 354.87 | 2.60 | 5.0526 | 0.8570 | 1.00 | 1.2980 | 1.1246 | 0.1734 |
| 19 | 359.38 | 1.88 | 6.5680 | 0.9150 | 0.96 | 1.3012 | 1.1942 | 0.1071 |
| 3 | 337.21 | 1.49 | 3.5628 | 0.5950 | 0.86 | 1.3170 | 1.0164 | 0.3006 |
| 14 | 324.96 | 1.86 | 6.5680 | 0.3470 | 0.81 | 1.3493 | 1.3058 | 0.0435 |
| 20 | 319.98 | 1.89 | 4.0706 | 0.4440 | 0.97 | 1.4134 | 1.2477 | 0.1658 |
| 43 | 290.79 | 2.28 | 3.3447 | 0.2940 | 0.89 | 1.4881 | 1.4684 | 0.0197 |
| 9 | 309.26 | 1.76 | 5.3444 | 0.5050 | 0.91 | 1.4925 | 1.2514 | 0.2411 |
| 31 | 307.72 | 2.04 | 4.8160 | 0.7060 | 1.00 | 1.4932 | 1.3697 | 0.1235 |
| 4 | 299.35 | 1.50 | 6.5680 | 0.7460 | 0.89 | 1.5584 | 1.1736 | 0.3848 |
| 1 | 270.88 | 1.35 | 3.5628 | 0.5718 | 0.92 | 1.6396 | 1.0000 | 0.6396 |

## 2.9 Appendix B: Second stage estimations for DEA and II-DEA efficiency scores

Table 2.9: Efficiency explanatory variables under the endogeneity and exogeneity assumptions

| Dependent variable: dhat | Truncated + bootstrap (II-DEA) | | | | Truncated + bootstrap (DEA) | | | |
|---|---|---|---|---|---|---|---|---|
| | **Coef** | **Std. Err.** | **z** | | **Coef** | **Std. Err.** | **z** | |
| TECHVOC(d) | 0.0097 | 0.057 | 0.17 | | 0.0536 | 0.990 | 0.32 | |
| RURAL(d) | -0.0062 | 0.074 | -0.08 | | -0.0255 | 0.087 | -0.29 | |
| SCHSIZE | -0.0001 | 0.000 | -1.81 | * | -0.0001 | 0.000 | -1.53 | |
| PCTGIRL | 0.0249 | 0.165 | 0.15 | | -0.1433 | 0.166 | -0.87 | |
| ICTSCH | -0.0395 | 0.067 | -0.59 | | -0.0395 | 0.049 | -0.80 | |
| PCTCORRECT | -0.2898 | 0.117 | -2.47 | ** | -0.1300 | 0.089 | -1.46 | |
| ANXMAT | 0.2410 | 0.077 | 3.14 | *** | 0.1255 | 0.064 | 1.96 | ** |
| PCTMATHEART | 0.5081 | 0.268 | 1.89 | * | -0.0087 | 0.243 | -0.04 | |
| TEACHGOAL | 0.3965 | 0.253 | 1.57 | | -0.3214 | 0.227 | -1.41 | |
| TEACHCHECK | -0.5443 | 0.228 | -2.39 | ** | -0.0017 | 0.189 | -0.01 | |
| HINDTEACH(d) | -0.0873 | 0.039 | -2.24 | ** | -0.0497 | 0.037 | -1.35 | |
| TEACHMORAL(d) | -0.1056 | 0.049 | -2.13 | ** | -0.0253 | 0.036 | -0.71 | |
| RESPCUR | -0.0962 | 0.064 | -1.50 | | -0.0661 | 0.072 | -0.92 | |
| RESPRES | 0.1902 | 0.199 | 0.95 | | 0.1696 | 0.221 | 0.77 | |
| _cons | 0.5361 | 0.423 | 1.27 | | 1.0170 | 0.401 | 2.53 | |
| /sigma | 0.0926 | 0.01 | 8.65 | | 0.0751 | --- | --- | |

*Note*: Left-truncated regression at value one // 2,000 bootstrap replications //'Coef' are the estimated coefficient. Negative values increase efficiency // Std.Err. are the robust standard error of the estimated coefficients // Sample size = 71 // $***p-value < 0.01; **p-value < 0.05$ ; $*p-value < 0.10$ // (d) Dummy variables where the mean represents the proportion of schools in the reference category // References categories are: vocational technical school (TECHVOC); school located in rural are (RURAL);school's principal perceives teachers not being well prepared (HINDTEACH) and school's principal perceives high teachers morale (TEACHMORAL).

# Chapter 3

# The teacher effect: an efficiency analysis from a randomized natural experiment in Spanish schools

## 3.1 Introduction

The number and quality of the years of education received by an individual throughout his live will determine his future welfare. More and better educated people have on average not only better access and adaptability to the labour market but also higher salaries as a consequence of their greater productivity. For this reason, the investment that a country makes in quality education is essential to ensure its sustainable development and economic growth (Barro and Lee, 1996, 2012; Hanushek and Kimko, 2000; De la Fuente, 2011; Hanushek and Woessmann, 2012a, 2012b). Thus, it is not surprising that public expenditure on education is one of the largest public budget items, and that the public sector is the main provider of education in most countries. Indeed, in most OECD countries the trend has been to increase the public expenditure devoted to education, although this process stopped and even reversed in some countries as a consequence of the global economic crisis. However, can we conclude that more educational resources guarantee better educational quality? The answer to this question is not straightforward and, at least in developed countries with an actual high educational expenditure, the evidence is negative. In this context, governments should not be concerned only with improving academic results through educational public expenditure and more attention should be put to making a better use of this expenditure, that is, to be more efficient in the use of educational resources.

In this regard, teachers play a key role because it is inside the classrooms where the educational production process takes place and the innovation in teaching methods and educational practices can actually improve educational quality. Drawing on the Coleman's Report (Coleman *et al.*, 1966) many studies have argued in the last decades that differences in school resources have a limited influence in academic outcomes, concluding that family background and the peer group effect are the most important variables to explain education results. Furthermore, in the last decade some works demonstrate that teachers' observed characteristics (experience, academic training, etc.) do not show a consistent relationship with students' test scores (Rivkin *et al.*, 2005; Hanushek and Rivkin, 2006; Clotfelter *et al.*, 2007; Kane *et al.*, 2008). This lack of relation could be interpreted as teachers' quality not mattering in their effectiveness or –alternatively- it could reveal that quantifiable teacher variables are not good proxies for their quality. (Hanushek y Rivkin, 2012).

Based on this evidence, researchers and policy makers are turning their attention to the impact measures of teachers' performance, using the value added in the academic outcome of students as the main indicator of the quality and effectiveness of teachers (Hanushek and Rivkin 2010; Rothstein, 2010). In fact, several recent studies have shown that there exist substantial differences in teachers' quality and that these differences have large impacts not only over students' test scores (Rockoff, 2004; Rivkin *et al.*, 2005; Gordon *et al.*, 2006; Hanushek and Rivkin, 2010; Aaronson *et al.*, 2007; Kane and Steiger, 2008; Kane *et al.*, 2008) but also on their long term earnings (Chetty et al. 2011). However, most of these works do not consider the fact that students are generally not randomly allocated neither in schools nor in classrooms within schools. Therefore, estimations about teachers' true impact on students' academic outcomes could be biased if this endogeneity problem is not taken into account (Rothstein, 2010; Hanushek and Rivkin, 2012).

The aim of this research is to shed new light about the differences in teachers' quality and their impact on students' academic results controlling by the presence of endogeneity. For this purpose, we employ a different methodological approach through the measurement of the technical efficiency inside the classrooms. Once we control for differences in student' background and school resources between classrooms within schools, the classroom technical efficiency reflects the teacher's global impact on students test scores. Our strategy lies on exploiting the exogenous variation between classrooms that is produced within each school when students are randomly assigned to classrooms, thus creating a natural experiment. (Schlotter *et al.*, 2011).

To do this we use the 'General Diagnostic Assessment' Database (*Evaluación General de Diagnóstico* in Spanish) that captures information about principals, teachers, parents and students in their fourth grade of primary education in Spain during 2009. For half of schools assessed, this database contains information about two classrooms inside the same school and it allows us to know whether or not students were randomly assigned into classrooms in each school. Randomization guarantees that on average students' observable and non observable characteristics are similar in both classrooms inside the same school. Parents can self-select in schools but they cannot decide the classroom in which their children will be allocated within the school. Under this framework the only difference between two groups at the starting school date is the teacher that has been randomly assigned to the classroom. Differences on average students' academic results between two classrooms will be directly related with teachers' differences in performance. In this research we will measure teachers' quality using the technical efficiency of the teacher in producing education. We consider that each teacher seeks to maximize average students' results taking into account its inputs (school resources and students' background) available.

Thus, in a first stage we estimate technical efficiency at classroom level using a Data Envelopment Analysis (DEA) model for those schools that randomly assign students to classrooms. In the second stage we analyse classrooms' efficiency differences within these schools, *i.e.* teachers' efficiency differences inside each school. As it was shown in Chapter 1, endogeneity due to self-selection can bias the DEA efficiency scores obtained in the first stage However, this bias can be mitigated if we analyze differences in efficiency between classrooms within each school.

Our results agree with previous works cited above and corroborate that there exist considerable differences in teachers' quality in primary schools in Spain. Within school difference between the most and the least efficient teacher is on average 4.4 efficiency points, which represents 0.82 standard deviations with respect the mean efficiency. Once this differences are computed, it is possible to evaluate the average impact of having the most efficient teacher in the classroom in terms of academic results. Randomization produces a natural experiment allocating efficient and not so efficient teachers to similar classrooms. Therefore, we can conclude that the average difference between the classroom with the most efficient teacher (treated group) and the less efficient one (control group) is a measure of the impact of having the better teacher. Our estimation finds that for Spain this impact is even larger than the results found for the United States in previous studies. In our case, students randomly assigned to the most efficient teacher obtain on average 0.43 (0.44) standard deviations more in maths (reading) test scores. According to Hanushek and Rivkin (2010, p. 269) these impacts vary between 0.11 and 0.36 (0.08 and 0.26) standard deviations in maths (reading).

Last, but not least, we explore whether some observable variables can explain efficiency differences between teachers, aiming to contribute to the debate about which educational policies could be implemented to select and retain the best teachers. To answer this question we regress the efficiency ratio between the most and the least efficient teacher in each school over some observable teacher and classroom variables together with other environmental school variables and students' characteristics. Our results corroborate that neither teacher experience nor academic training explain teachers' quality or efficiency (Rivkin *et al.*, 2005; Hanushek y Rivkin, 2006; Kane *et al.*, 2008). On the other hand, we find several factors significantly correlated with teachers' efficiency. For example, to be a female teacher, having worked more than five years in the evaluated school and to repeat a second year with the same group of students positively affects efficiency. Likewise, having fewer students per classroom positively impacts the results.

In sum, this research presents an original approach to evaluate teacher's quality and its effect on students, and has various contributions to the existing economics of education literature. Firstly, this is the first work that analyses teachers' quality through the measurement of technical efficiency using a natural experiment, thus allowing to deal with the presence of endogeneity. Secondly, most of previous works focused on the measurement of educational efficiency and its explanatory factors cited in Chapter 2 used school or student level analysis. The only exceptions that employ classrooms as production units in efficiency analysis are Cooper and Cohn (1997) analysing 541 classrooms in South Carolina to explain the effect of teacher incentives on results; De Witte and Rogge (2011) using Belgian data to measure teachers' quality based on student' evaluations and finally Klaveren y De Witte (2014) who carry out an efficiency study using German data (second grade of secondary education) from TIMSS 2003 to examine what teaching activities maximize students' results. However, these works are far away in their aims from ours and most of them do not consider the endogeneity problem. Finally, although there is a wide agreement about the importance of investing educational resources at early ages (Heckman and Kautz, 2013), most efficiency studies have focused their efforts in analyzing secondary and tertiary education. There exist some works studying primary education for different countries (Mancebón y Mar-Molinero, 2000; Grosskopf *et al.*, 2001; Mizala *et al.*, 2002; Thanassoulis, 2002; Banker *et al.*, 2004; Blackburn *et al.*, 2013; Casalprim *et al.*, 2013), however, to the best of our knowledge this is the first research about measuring the technical efficiency in primary education for Spain.

The rest of Chapter 3 is organized as follows. Section 3.2 presents the main methodological concepts and our identification strategy to measure teacher efficiency free of endogeneity problems. Section 3.3 briefly describes the database and the variables included in the analysis. Section 3.4 reports the estimation results. Finally, Section 3.5 discusses the conclusions of this research and their implications for educational policy makers.

## 3.2 Methodology

The theoretical approach used in this paper for linking resources to educational outcomes at school level is based on the well-known educational production function proposed by Levin (1974), Hanushek (1979) and Hanushek et al (2013):

$$A_i = f(B_i, S_i) \tag{3.1}$$

where subindex $i$ refers to school, and $A_i$ represents the educational output vector for school $i$. This output is normally measured through the students' average scores in standardized tests. On the other hand, educational inputs are divided into $B_i$, which denotes average student family and socio-economic background, and $S_i$, which are the school educational resources. The educational production function is frequently estimated considering the possible existence of inefficient behaviours in schools following Equation 3.2,

$$A_i = f(B_i, S_i).u_i \tag{3.2}$$

where $0 \leq u_i \leq 1$ denotes the efficiency level of school $i$. Values of $u_i = 1$ imply that the analysed schools are fully efficient, meaning that given the initial input endowment and the existing technology, these schools are maximizing their outputs and managing correctly the school. Values $u_i \leq 1$ would indicate that the school is inefficient. The estimation of Equation 3.2 assumes that inputs are exogenous or, in other words, that the efficiency term and the educational inputs are uncorrelated $E(u_i|B_i, S_i) > 0$. However, this assumption frequently does not hold in the production of education because students are not randomly assigned to schools (Schlotter *et al.*, 2011). Most motivated parents and those who give more value to education put more effort and resources in selecting the best schools for their children. These parents gather more formal and informal information about schools results and peer group in order to choose the best available option for their children (Hoxby 2000; Sacerdote, 2001). As a consequence, children from most motivated parents attending together the same schools will obtain better results for two reasons. On one hand, because it is expected that these schools will have a higher average socio-economic level $B_i$ . On the other hand, because these schools also have the students with most motivated parents, a variable that also influences students' results. Since parents' motivation is an unobservable variable, its effect over academic results would be captured in Equation 3.2 by the efficiency term. To disentangle this effect we can rewrite Equation 3.2 as follows:

$$A_i = f(B_i, S_i).u_i = f(B_i, S_i).\theta_i.\gamma_i \tag{3.3}$$

where $0 \leq \theta_i \leq 1$ captures the managerial efficiency in school $i$ and $0 \leq \gamma_i \leq 1$ denotes the non-observable students' characteristics, particularly the average parental motivation of students attending at school $i$. When $\gamma_i = 1$ the school shows the maximum motivation of students' parents, a value that progressively declines when $\gamma_i$ moves away from one. As it was discussed previously, parents' motivation is positively correlated with the average school socio-economic level $E(\gamma_i|B_i) > 0$ and for this reason the global average efficiency will be also correlated with $B_i$, $E(u_i|B_i) > 0$. This positive correlation implies the presence of positive endogeneity in the estimation. From Chapter 1 we know that this endogeneity can turn into flawed efficiency estimations for the term $u_i$.

In short, the lack of randomization in allocating students to schools and the existence of non observable variables implies that when we try to estimate $u_i$ we really estimate the confounding term $\theta_i.\gamma_i$ that in practice cannot be decomposed and is a biased estimate of the true managerial efficiency of schools. Although $\theta_i$ is exogenous, the presence of $\gamma_i$ biases its estimation because this term is positively correlated with $B_i$. Under these circumstances a direct estimation of $u_i$ is biased, in order to deal with this problem we have to look for an identification strategy to measure $\theta_i$ independently of $\gamma_i$. To do this we propose to employ impact evaluation insights (Schlotter *et al.*, 2011) as a way to improve performance measurements when the presence of positive endogeneity affects our data. In this research we take advantage of a database of schools from Spain in which we can identify those schools where students were randomly assigned to classrooms. This randomization produces a natural experiment where by chance one classroom has been assigned to the most efficient teacher and the other one to the least efficient teacher.

From Equation 3.4 we know that the average result of $N$ students $n = 1, 2, \ldots, N$ distributed in $K$ classrooms $k = 1, 2, \ldots, K$ in school $i$ $i = 1, 2, \ldots., M$ are determined through the following production function:

$$A_{ik} = f(B_{ik}, S_{ik}).u_{ik} = f(B_{ik}, S_{ik}).\tau_{ik}.\omega_{ik}.\gamma_{ik} \tag{3.4}$$

where $A_{ik}$ denotes the educational output vector for classroom $k$ at school $i$. This output depends on a set of observable variables $(B_{ik}, S_{ik})$ and non observable variables captured by the efficiency term $0 \leq u_{ik} \leq 1$. Technical efficiency $u_{ik}$ can be decomposed in three terms at classroom level: average non observed characteristics of students $0 \leq \gamma_{ik} \leq 1$; the school managerial efficiency $\tau_{ik}$, $0 \leq \tau_{ik} \leq 1$ which is the same for all classrooms in the same school $\tau_{i1} = \tau_{i2} = ... = \tau_{iK} = \tau_i$ ; and the teacher efficiency $0 \leq \omega_{ik} \leq 1$ that captures the teacher's ability to deal with the educational process at his classroom. For $\tau_{ik}$ and $\omega_{ik}$ a value equal to one corresponds to the maximum level of performance, which declines when this value decreases. If there exists randomization in the assignment of students into classrooms this three components are expected to be independent of each other.

The direct efficiency estimation at classroom level $u_{ik}$ suffers from the same endogeneity problems discussed before and it would be biased. However, the fact that we have information about two classrooms inside the same school allows us to work with the difference between both groups to correct the bias produced in the direct estimation. Let assume that in every school that randomly assigns students to classes we have two groups $k = 1, 2$[1]. If we estimate the technical efficiency at each classroom in those schools using Equation 3.4 and we compute the efficiency ratio between the two classrooms within each school, we have

$$E\left(\frac{u_{i1}}{u_{i2}}\right) = E\left(\frac{\tau_{i1}.\omega_{i1}.\gamma_{i1}}{\tau_{i2}.\omega_{i2}.\gamma_{i2}}\right) = E\left(\frac{\tau_{i1}}{\tau_{i2}}\right) E\left(\frac{\omega_{i1}}{\omega_{i2}}\right) E\left(\frac{\gamma_{i1}}{\gamma_{i2}}\right) \tag{3.5}$$

Given that the school technical efficiency is the same for both classrooms, it is straightforward to conclude that $E\left(\frac{\tau_{i1}}{\tau_{i2}}\right) = 1$. Likewise, if students are randomly assigned to classrooms inside every school (for example by alphabetical order), then it is expected that on average students from both groups are similar not only in observed characteristics (*e.g.* socio-economic level) but also in the non-observed ones (*e.g.* parents' motivation). More motivated parents can self-select into best schools but we assume that, because of the random allocation process within schools, they cannot choose the best classroom inside the school. Randomization therefore guarantees that the expected value of the ratio of average non observable characteristics (motivation) between both classrooms be equal to one, $E\left(\frac{\gamma_{i1}}{\gamma_{i2}}\right) = 1$ . Thus, we can conclude that observed differences in the estimated efficiency between classrooms will be due to differences between teachers. In this case we can derive that $E\left(\frac{u_{i1}}{u_{i2}}\right) = E\left(\frac{\omega_{i1}}{\omega_{i2}}\right)$[2].

Therefore, although the estimated efficiency scores for each classroom are biased due to self-selection, taking efficiency ratios between classrooms within random class assignment schools allows us to correctly identify the true differences in teachers' performance. And if these differences are significant the next question that comes up is: what is the effect of these differences on students' test scores? As a consequence of the natural experiment, this impact can be computed as the difference of the average results of treated classrooms (assigned to the most efficient teacher) with the average results in the control classroom (assigned to the least efficient teacher).

Finally, we analyse which factors can explain the efficiency gap between teachers. In other words, is it possible to relate some teachers' observable characteristics to their efficiency? To answer this question we regress the ratio of classrooms' efficiency against a set of control variables associated with students and schools characteristics and with observed teachers' variables such as gender, experience, academic training, etc.

Schematically, our methodological strategy in this research can be summarized in the following steps:

1. From a school sample that uses randomization to allocate students to classrooms within schools we estimate the technical efficiency at the classroom level using the Data Envel-

[1] For the sake of simplicity the model is described for two groups. In the case of more groups the model extension is trivial taking a group $k$ as reference and calculating $k-1$ differences.

[2] The use of ratios instead of differences is necessary to isolate the difference is teachers' performance. Calculating $u_{i1} - u_{i2} = (\tau_{i1}.\omega_{i1}.\gamma_{i1}) - (\tau_{i2}.\omega_{i2}.\gamma_{i2}) = (\gamma_{i1} - \gamma_{i2}).\omega_i.\tau_i$ where now the $\omega_i.\tau_i$ term is not the same for every school and confounds again the difference in teachers' performance.

opment Analysis (DEA) method introduced by Charnes, Cooper and Rhodes (1978, 1981) and Banker, Charnes and Cooper (1984). The technique implements a linear optimization program to obtain a production frontier comprising all the efficient units and their possible linear combinations. Thus, the estimated efficiency score for each Decision Making Unit (DMU) is a relative measure calculated using all the production units that are compared. The formulation of the output-oriented DEA program under variable returns to scale (DEA-BBC model) for each analyzed unit is:

$$\varphi_i = \max_{\lambda,\varphi}\{\varphi_i|\varphi y_i \leq Y\lambda; x_i \geq X\lambda; n1'\lambda = 1; \lambda \geq 0\} \ \forall_i = 1, ..., n \tag{3.6}$$

where, for the $kth$ DMU in the $ith$ school, ($\varphi_{ik} = \frac{1}{\hat{u}_{ik}} \geq 1$ is the efficiency score, $y_{ik}$ is the output vector $(q \times 1)$ and $x_{ik}$ is the input vector $(p \times 1)$, and thus X and Y are the respective input $(p \times nk)$ and output $(q \times nk)$ matrices. The $(nk \times 1)$ vector $\lambda$ contains the virtual weights of each unit determined by the problem solution. When $\varphi_{ik} = 1$, the analysed unit belongs to the frontier (is fully efficient), whereas $\varphi_{ik} > 1$ indicates that the $ith$ unit is inefficient, $\varphi_{ik}$ being the radial distance between the $ith$ unit and the frontier. In other words, $\varphi_{ik}$ indicates the equiproportional expansion over outputs needed to reach the frontier. Therefore, the higher the score value $\varphi_{ik}$, the greater the inefficiency level.For example, $\varphi_{ik} = 1.2$ suggest that this DMU is inefficient because it could obtain 20% more output with its available inputs.

2. Once the efficiency $u_{ik}$ is estimated for each group $k = 1, 2$ inside each school $i$, we identify the most efficient teacher, that corresponds with the treated classroom $(T)$, and the least efficient one, that corresponds with the control classroom $(C)$, $0 \leq \hat{u}_{iT} \leq \hat{u}_{iC} \leq 1$. To measure the impact of having been assigned to the most efficient teacher we compare academic outcomes of both the treated and the control groups.

$$\Delta Y = \frac{1}{M}\sum_{i=1}^{M}\Delta\bar{Y}_i = \frac{1}{M}\sum_{i=1}^{M}(\bar{Y}_{iT} - \bar{Y}_{ic}) \tag{3.7}$$

where $\bar{Y}_{iT}$ $(\bar{Y}_{iC})$ is the impact, in terms of test scores, of having the most efficient teacher, $\bar{Y}_{iT}$ is the average results of students assigned to the treated group and $\bar{Y}_{iC}$ is the average result of students assigned to the control group.

3. Lastly, to isolate the teacher effect for the 'treated' group at each school we compute the efficiency ratio between both teachers as follows:

$$\Delta\hat{u}_i = \frac{\hat{u}_{iT}}{\hat{u}_{iC}} \geq 1 \ \forall i \tag{3.8}$$

Efficiency ratios $\Delta\hat{u}_i$ are regressed over a set of observed teachers' characteristics. The regression model also includes other control variables related with the school and the classroom that allow controlling for exogenous variables that may also be explaining the efficiency gaps. The model to be estimated is:

$$\Delta\hat{u}_i = (\alpha_{iT} - \alpha_{iC}) + \beta_P(P_{iT} - P_{iC}) + \beta_Z(Z_{iT} - Z_{iC}) + \beta_W W_i + \varepsilon_i \tag{3.9}$$

where $(P_{iT} - P_{iC})$ denotes differences in observed teachers' characteristics; $(Z_{iT} - Z_{iC})$ represents the vector of students' differences between both classrooms and finally $W_i$ is a set of school variables to capture school and principal characteristics. Taking into account that by construction $\Delta\hat{u}_i \geq 1$, Equation 3.9 is estimated through a left-censored regression model censored at value one.

Finally, in order to empirically quantify the impact that not controlling by the existence of endogeneity would have in the estimates, we carry out an efficiency analysis considering all the schools in the sample, as it is usually done in standard efficiency estimations. In other words, we also include in the analysis those schools where students are not randomly assigned and those with information only available for one group. In that case, the model to estimate is a semi-parametric two-stage model proposed by Ray (1991) and McCarty and Yaisawarng (1993). The first stage of this approach is to apply a DEA model that measures technical efficiency at classroom level, whereas a regression analysis conducted in the second stage seeks out the main explanatory factors of efficiency. Following Simar and Wilson (2007) in the second stage we estimate a truncated regression model with bootstrap.

$$\hat{u}_{ik} = \alpha_{ik} + \beta_P P_{ik} + \beta_Z Z_{ik} + \beta_W W_i + \varepsilon_{ik} \tag{3.10}$$

where in this case $\hat{u}_{ik}$ is the technical efficiency estimated in the first stage by DEA (Equation 3.6); $P_{ik}$ are the observed teacher's characteristics; $Z_{ik}$ is a vector of observed students' variables that could affect efficiency and finally $W_i$ are school characteristics common for all classrooms belonging to the same school.

## 3.3 Data

### 3.3.1 The EGD Database

To carry out our estimations we use data from the 'General Diagnostic Assessment' conducted by the Spanish Ministry of Education, Culture and Sport applied during 2009 to a sample of fourth-year primary students all over Spain (we will refer to this database as EGD from now on)[3]. EGD focuses on measuring the knowledge, skills and attitudes acquired by students in four core competencies: language, mathematics, social and civic education (social studies) and knowledge and interaction with the physical world (science). Like other international studies (PISA, TIMSS, PIRLS), a complementary questionnaire is also administered to

[3]A detailed description of this database including simple design and included variables can be found in INEE (2010).

students and their families, school principals and teachers to gather additional information on contextual factors, resources and organizational processes that allows further analysis of the students' performance.

Primary education in Spain is organized so that students have the same teacher during almost all of the school day and who teaches all four core subjects (reading, mathematics, science and social studies). The rest of the school day students are taught by one of the discipline specialists (for example, foreign language or physical education). In this work we evaluate the principal teacher in charge of the classroom.

The EGD respondents totalled 27,125 students distributed in 1,295 classrooms from 882 schools. In 442 schools two complete fourth grade classrooms were evaluated. This is a novel feature of the EGD that distinguishes it from other national and international education databases. Also, most interestingly, we can know how students were assigned to classrooms inside each school. When students are randomly assigned in classrooms, differences in test scores are mainly due to differences in teachers. Moreover, randomization of student allocation produces a natural experiment: by chance, students from one classroom will receive the best teacher and students from the other classroom the worst one.

In order to identify which schools randomly assign their students, EGD asks the school principal how students were grouped in classrooms. Table 3.1 shows these criteria and classifies them into random and not random criteria. Therefore, from 442 schools in which two classrooms were originally evaluated, we have excluded schools that employed a non random criterion: 'linguistic reasons', 'academic performance', 'looking for homogeneity of students' characteristics' and 'other criteria'. Finally, we analyse 213 schools[4] (426 classrooms) that use randomization to group students in classrooms. In this sample 66% are public schools while the remaining 34% are government dependent private schools (private schools publicly funded).

### 3.3.2 Variables

In this section we define and provide a brief statistical description of the selected outputs and inputs used to estimate classrooms' technical efficiency (Equation 3.6) and of the contextual variables used to explain differences between teachers (Equation 3.9).

The different cognitive and non-cognitive dimensions of the education received by an individual make it difficult to measure educational output. Still, there is a general consensus in the literature in favour of considering the results of standardized tests (for example the EGD) as educational outputs. Hanushek (1997) reports that around two thirds of educational research studies use test scores as output and Hoxby (1999) highlights that these tests are difficult to forge and, above all, that they are taken into account by parents and politicians when making decisions on education. Thus, for our study we have selected as output variables the average

[4]Initially, we identify 219 schools but we classify 6 schools as outliers, with extreme values in some of the relevant variables included in the efficiency analysis.

classroom result in reading (READ) and in maths (MATHS), which measure two vehicular and complementary cognitive dimensions[5].

The educational inputs were selected considering the classical educational production function (Equation 3.2), and they represent the inputs required to carry out the learning process: students' characteristics (raw material), teachers (human capital) and infrastructure (physical capital)[6]. Thus, the following four variables were included in the first stage of the DEA:

- ISECS: Average index of social, economic and cultural status of students in the classroom. This index was calculated by EGD analysts to measure student's background, and reflects the 'raw material' to be transformed in the learning process. The variable was calculated through a factor analysis considering four components: highest educational attainment of parents; highest professional status of parents; number of books in the household and level of domestic resources.
- PNAT: Percentage of native students in the classroom. This status also reflects the 'raw material' to be transformed in the learning process. Previous research in Spain shows that to be an immigrant significantly affects test scores (Calero *et al.*, 2009; Salinas and Santín, 2012; Zinovyeva *et al.*, 2013). Therefore, the percentage of native students should be considered to fairly compare classrooms.
- PCORR: Percentage of students in the correct grade within the classroom. Grade retention is a predictor variable of education outcomes (Jimerson *et al.*, 2002). As the socio-economic background and the native status, the non-repeater status reflects the 'raw material' to be transformed in the learning process. For this reason we also must include this input in the analysis to carry out a fair comparison.
- IQER: This index captures the quality of the educational resources in the school. It is elaborated through a factor analysis of the teachers' responses to four questions related to the scarcity or lack of: educational materials, computers for teaching, instructional support staff and other support staff. It is therefore associated with the human and physical capital available resources to produce education. The higher the index, the better the quality of the school's resources.

To be considered as an input in an efficiency analysis, a variable has to be significantly and positively correlated with all outputs. This monotonicity assumption implies that additional units of an input[7] will never decrease output. Table 3.2 presents the bivariate correlations of the selected outputs and inputs, and all correlations are found positive and statistically significant.

[5] The other competencies 'science' and 'social science' were not considered because they provide little additional information and are highly correlated with average results in Reading and mathematics as well.

[6] We focus on the quality rather than just the quantity of these inputs.

[7] We do not include the teacher-students ratio because it is negatively and significantly correlated with the output, thus breaking the monotonicity assumption. This negative correlation can be associated with the self-selection problem (Webbink, 2005. p.538). Best schools are more demanded and this raises classrooms sizes up to the legal limits, distorting the true effect of this variable on test scores. To deal with this problem we will include this variable in the second stage after controlling for the potential endogeneity.

This fulfilment of the monotonicity assumption is valid for classrooms belonging to schools that randomize and also for the whole pool of classrooms included in the EGD.

Regarding explanatory variables of teachers' efficiency, we selected and included in the analysis (Equation 3.9) the following variables:

- TEACHgen. Teacher gender. A dummy variable which takes the value one when the teacher is a female and zero when he is a male.
- TEACHcertified. A dummy variable which takes the value one when the teacher holds a teaching diploma and zero otherwise.
- TEACHgraduated. A dummy variable which takes the value one when the teacher holds a bachelor's diploma and zero otherwise.
- TEACHexp5. A dummy variable which takes the value one when the teacher has less than five years of teaching experience and zero otherwise.
- TEACHexp10. A dummy variable which takes the value one when the teacher has less than ten years of teaching experience and zero otherwise.
- TEACHexp30. A dummy variable which takes the value one when the teacher has more than thirty years of teaching experience and zero otherwise.
- TEACHschold. A dummy variable which takes the value one when the teacher has been working in the school less than five years and zero otherwise.
- TEACHtutor. A dummy variable which takes the value one when the teacher has been the teacher of the evaluated classroom in the last two academic years, *i.e.* third and fourth grades, and zero otherwise (just the current fourth year).

From these dummy variables at classroom level we define, as Equation 3.9 shows, variables taking differences between the most and least efficient teacher ($P_{iT} - P_{iC}$) within a given school. These variables can take values equal to -1, 0 and 1. For example, recalling the $TEACHgen$ variable, when the most efficient teacher in the i-th school is a male ($TEACHgen_{iT} = 0$) and the least efficient is a female ($TEACHgen_{iC} = 1$), then the difference between both groups is $\Delta TEACHgen = (TEACHgen_{iT} - TEACHgen_{ic}) = -1$. If both teachers have the same gender then $\Delta TEACHgen = 0$. And finally if the most efficient teacher in the i-th school is a female ($TEACHgen_{iT} = 1$) and the least efficient is a male ($TEACHgen_{iC} = 0$) , then the difference between both groups is equal to 1. By contrast, note that the estimation of Equation 3.10 is performed at school level so the defined dummy variables are directly included with no differences.

Furthermore, according to Equation 3.9 we include other variables related with classroom composition ($Z_{iT} - ZiC$) to control for the possibility that, by chance, most efficient classrooms had better conditions that explained some of the efficiency differences between teachers:

- PCGIRLS. Percentage of girls students in the classroom.
- EARLYSCH. Number of years that on average students in the classroom attended pre-primary education.

- PMONOPARENTAL. Percentage of students in the classroom that lives in single-parent families. The variable was built from students answers declaring not to live simultaneously with both biological parents.
- PQUARTER4. Percentage of students in the classroom that were born in the fourth quarter of the year.
- CLASSIZE. Number of student in the classroom.

Again, in Equation 3.9 we compute differences in these variables between the class of the most efficient teacher (T) and the class of the least efficient one (C), while to estimate Equation 3.10 these variables are directly introduced.

Finally, to figure out if school variables can explain part of the efficiency gap between teachers we also include the following $W_i$ control variables in Equations 3.9 and 3.10:

- SCHpublic. Dummy variable which takes the value one when the school is public and zero otherwise (private school publicly funded).
- SCHrural. A dummy variable which takes the value one when the school is located in a less than 10,000 inhabitants area and zero otherwise.
- SCHcity. A dummy variable which takes the value one when the school is located in a city with 500,000 or more inhabitants and zero otherwise.
- PPALfemale: Dummy variable which takes the value one when the school's principal is a woman and zero if he is a man.
- PPALexp5. A dummy variable which takes the value one when the school's principal has less than five years of experience as a principal and zero otherwise.

Tables 3.3 and 3.4 show, for schools with random assignment and for all classrooms that participated in the EGD survey respectively, the main descriptive statistics of outputs, inputs and explanatory variables of teachers' efficiency at classroom level. We can observe a slight advantage in results in favour of schools that use randomization but we do not appreciate large differences in the composition of both samples.

## 3.4 Results

Prior to estimating classrooms' efficiency, we check if students were actually assigned randomly into classrooms within schools and thus, if both teachers received similar students. To do that, we conduct a mean differences t-test between groups over the students' observable variables. We also test for differences in teachers' observable variables to verify that they have been randomly assigned to each classroom. Results in Table 3.5 confirm the randomization hypothesis. Differences in all variables are not significant and hence we cannot reject that both groups have similar means. Since no significant differences have been found in observable variables, it is also expected to find no differences in non observable variables between classrooms within schools.

Figure 3.1 shows the distribution of the estimated efficiency scores for each group in schools with random assignment. The estimated average mean efficiency is 91.6 with a standard deviation of 5.34. As we have exposed above, the estimated efficiency scores for each unit are biased by the presence of self-selection and their direct analysis in a second stage regression would be misleading. However, from these estimated efficiency scores we can compute the ratios between the most and the least efficient teachers in each school $\frac{\hat{u}_{iT}}{\hat{u}_{iC}} \geq 1$. Figure 3.2 shows the distribution of these efficiency ratios. One remarkable and straightforward finding is the presence of considerably differences between the teachers' efficiency within primary schools in Spain. The average ratio between the most and the least efficient teachers is 1.05 with a standard deviation of 0.04. In other words, the best teachers are on average 5% more efficient, which means 0.86 additional standard deviations from the estimated mean efficiency. Also, we find that more than one third of the schools (36%) show differences in teachers' efficiency greater than one standard deviation from the estimated average efficiency. Moreover, differences in efficiency are greater than 10% (almost two standard deviations) in 14% of schools.

Table 3.6 presents the mean differences in efficiency, outputs and inputs between classrooms with the most efficient and the least efficient teacher. These results restate that, in terms of initial educational inputs and students' characteristics, both classrooms are on average indeed similar. We only find differences in efficiency and therefore in the students' academic results (because for equal level of inputs, the greater the efficiency, the greater results). On average, the difference between the most and the least efficient teachers is 4.4 efficiency points. While the most efficient teachers have on average an estimated efficiency score of 93.8, this score decreases to 89.4 for the least inefficient teachers.

How do these differences in efficiency translate into students' academic outcomes? From Table 3.6 we can observe that classrooms assigned to the most efficient teacher obtained on average around 18 additional points in the reading and maths scores compared to students in classrooms assigned to the least efficient teachers[8]. In other words, moving one standard deviation up the distribution of teacher efficiency is expected to raise reading and maths test scores by 0.53 and 0.54 standards deviations respectively. This impact is notably greater than those found in previous studies for the United States, where the teacher quality impact ranges from 0.08 and 0.36 standard deviations (Hanushek and Rivkin, 2010). These differences have several potential explanations. First, a country effect, *i.e.* that in fact differences in teachers' quality in Spain are translated in a greater impact on students academic results. Since this is the first study of this type for the Spanish educational context we cannot check this intuition. A second possibility is that part of this difference is due to the analysis level of the study, we use the classroom level when most of previous studies are based on the student level. As the outcomes variance in the second case is naturally higher than in the first case, the impact measured in

[8]As is shown in Table 3.3, the average results in reading and maths test in the 426 analysed classrooms are 507 and 508 respectively, with a standard deviation of 41.9 and 42.4 respectively.

terms of standard deviations will be grater at the classroom level[9]. Thirdly, differences can be explained by the methodological methods conducted to estimate the teachers' quality as all previous works are based on the estimation of Value Added Models. Finally, but not less important, these differences could be explained by the fact that most of the analysis conducted before do not take into account the possible non-random assignment of students into classrooms (Rothstein, 2010), whereby the true impact of teacher's quality in terms of students outcomes in these studies could have been underestimated.

Regarding the explanatory factors of the differences in teacher's efficiency, Table 3.7 summarizes the main descriptive of the variables included in model 3.9 and Table 3.8 reports the main results. The first important finding is that, once we control for self-selection, differences in students' characteristics between classrooms (that could have occurred by chance) do not explain teachers' efficiency. This evidence implies that when we control for the presence of endogeneity, the estimated classroom efficiency only reflects the teacher efficiency, *i.e.* the way they manage their classrooms. Similarly, contextual variables that characterize schools seem to not affect the efficiency gap observed between teachers within schools.

Classroom size matters. The greater the class size, the lower the estimated relative efficiency of the best teacher. Increasing the class size in one standard deviation (almost three students)[10] is expected to reduce the teachers' efficiency gap in 0.6 standard deviations. This impact in terms of students' outcomes means a reduction in 0.07 standard deviations in both reading and maths scores. These findings disagree with some previous works who have found the relation between the number of students in class and their performance not statistically significant (Hanushek, 2003). However, most of these findings have recently been questioned by the fact that they do not account for the presence of endogeneity (Webbink, 2005). In fact, several studies that have analysed the impact of reducing class size on student's results through natural experiments or quasi-experiments, *i.e.* controlling for endogeneity as we do here, have found significant positive effects in reducing the number of students in classrooms[11]. In this sense, Angrist and Lavy (1999) found that reducing in eight students the class size in Israeli primary fourth grade schools increased students' outcomes between 0.13 and 0.29 standard deviations. In our case, the same reduction in class size would lead to an increase in the average result of 0.20 standard deviations in both reading and maths scores. This evidence is highly relevant, as the modification of class size has been (and still is) the focus of several educational policies implemented in different countries. Particularly, in the Spanish context, where several budget cuts conducted in the last years have led to an increment in the ratio of students *per* teacher.

Concerning the significance of teachers' characteristics we find, in line with earlier research, that neither teacher's experience nor their academic training impacts teachers' efficiency or quality (Kane *et al.*, 2008). However, we find that other observable factors do affect teachers'

[9]In fact, the outcome variance at the student level in our sample doubles the observed at the classroom level.

[10]On average, classrooms in our sample have 24 students with a standard deviation of 2.87 (Table 3.3).

[11]Krueger (1999), Angrist and Lavy (1999), Akerhielm (1995), Boozer and Rouse (2001), Case and Deaton (1998) and Lindahla (2001).

efficiency. Firstly, female teachers are more efficient than their male colleagues. The fact of being a female instead of a male increases the teacher's efficiency in 0.11 standard deviations, which implies that classrooms assigned to a female teacher obtain on average 0.08 additional standards deviations in both reading and maths compared to classrooms with male teachers. These results are consistent with previous works that have found a positive effect of female teachers in students' results in primary education (Krieg, *et al.*, 2005; Chudgar and Sankar, 2008) and also in Spanish secondary education (Escardíbul y Mora, 2013). Although literature about this issue is not conclusive, several countries (*e.g.* United States, United Kingdom or Finland) have promoted the male participation in teaching profession in primary education in the last decade. Our empirical results suggest that these kinds of policies seem to be inappropriate for the Spanish primary education. In this case, it would be recommendable to further investigate what attitudes characterize female teachers to try to compensate and reinforce male teachers' attitudes and their relation with the students.

Teachers' seniority in the school positively affects their efficiency. Having less than five years in the school reduces the teacher's efficiency in 0.13 standard deviations, which implies a decrease of 0.09 standard deviations in students' academic results. This effect can be explained by different factors. Firstly, due to the entrance costs that entails to enter in a new school. The first years teachers have to acknowledge the school work dynamics, the colleagues, etc. and that can affect their performance until they can be fully adapted to the school. Another potential explanation could be related to the current mechanisms for selecting, hiring and retaining teachers, which would operate differently in the private and the public sector. In the private sector hiring and firing teachers is relatively flexible, so our results evidence that efficient teachers are correctly identified and retained into the system while, conversely, inefficient teachers may be fired and have to start in a new school.

In the Spanish public sector things are not so flexible, and although the principal could detect the most inefficient teachers, he/she would be practically unable to let them go. In this sector, seniority and efficiency can be related through the teacher's selection criteria based on an entrance examination. International evidence shows that, in systems where exists this type of entrance examinations, the score that teachers obtain is positively related to their effectiveness in terms of student outcomes (Clotfelter *et al.* 2007). In Spain, the score obtained in the entrance examination determines not only the access to a permanent position in the public school system, but also the preference to choose the precise school to work in. The best teachers tend to obtain better scores and are able to choose the school they actually prefer to work in, being more likely to remain in this schools longer throughout their teaching career. On the contrary, teachers who do not achieve a minimum score to access to a permanent position will have accept a temporary position or try teach in the private sector. Even among those who obtain the minimum score to access the public system, the ones with lower scores will have to work in schools that were not there first choice and will be more likely to leave them when other preferable position opens up. So, the greater the score obtained in the entrance examination, the higher likelihood of remaining more years in the same school.

Finally, we find that having been with the same students two consecutive years increases teachers' efficiency. Teachers who have been the classroom teacher the last two academic years (third and fourth grade) are on average 0.14 standard deviations more efficient than those wo only have been the current teacher one year. This effect implies in terms of students educational results an increase of 0.10 standard deviations in reading and maths test scores. Therefore, it seems that the current Spanish organizational system of primary education in two-years academic cycles (first and second grades, third and fourth grades, five and sixth grades) is an effective policy. By working two consecutive years with the same group, teachers can know more about their pupils and have a more flexible medium-term planning, which according to this evidence seem to have positive effects on the students' results at the end of the academic cycle.

It must be noted that these findings must be cautiously interpreted. They are a first attempt to estimate teacher quality in Spanish primary education by the estimation of technical efficiency. Clearly further research is needed in this direction to deeply explore the channels through which these findings operate. Indeed, the fact that the constant in Equation 3.9 results highly significant in the estimates reinforces the idea that there are other (observable or not) factors behind the teachers' efficiency.

Finally, to empirically assess the impact of not controlling for the self-selection problem in Table 3.9 we present the results of the estimation of Equation 3.10 including all the classrooms evaluated in the EGD (*i.e.*, including also those schools where students were not randomly assigned) to explain teachers' efficiency. Results significantly differ from those in Table 3.8. When we do not take into account the endogeneity problem, some teachers characteristics are no longer significant and vice-versa. But, even more relevantly, some of the students' and schools' characteristics are now significant (due to the presence of non-observable characteristics). These results provide strong evidence that not taking into account the self-selection in the estimations, *i.e.* the endogeneity, can bias the results and lead to inaccurate conclusions about which factors explain the teachers' efficiency and to inappropriate educational public policy recommendations.

## 3.5 Conclusions

In this chapter we study the effect of teachers' efficiency on students' academic achievement in Spanish primary schools controlling by the presence of endogeneity due to educational self-selection. From the 'General Diagnostic Assessment' Database administered to students in their fourth grade of primary education in Spain during 2009, we can identify those schools with two classrooms where students and teachers were randomly assigned.

The results evidence the presence of significant differences between teachers' efficiency, *i.e.* teachers' quality, in primary schools, which also have a large impact on students outcomes. The best teachers are on average 5% more efficient than the worst teachers within schools. In terms of students' results, this difference in efficiency implies that students assigned to the most efficient teacher obtain on average 0.43 and 0.44 additional standard deviations on reading and maths test scores compared to students assigned to the least efficient teacher. This impact is notably greater than that found in previous studies for the U.S., where the teacher quality impact ranges from 0.08 to 0.36 standard deviations (Hanushek and Rivkin, 2010). These differences can have several potential explanations, but perhaps the most relevant is the fact that most of those previous analysis do not take into account the possible non-random assignment of students into classrooms (Rothstein, 2010), whereby, the true impact of teacher's quality in terms of student outcomes could result underestimated.

We also explored potential explanatory factors of teachers' efficiency. The first relevant finding is that, once we have controlled for the presence of self-selection, there is no statistically significant relationship between the estimated teacher's efficiency differences and the variables that characterized students and schools. In other words, the estimated efficiency scores actually reflect only issues related to teachers and how they manage their classroom. A second significant result with important implications for education policy is that class size matters. The smaller the class size, the grater the teacher's efficiency. This evidence is consistent with earlier studies that have analyzed the impact on students' outcomes of reducing the class size through natural experiments or quasi-experiments (e.g. Angrist and Lavy, 1999). Finally, neither teacher's experience nor their academic training impacts teachers' efficiency. Conversely, we find three teacher's characteristics to be positively associated with their performance: being female teacher, having more than five years in the school and having been the teacher of the group for two consecutive years improves teacher's efficiency. These findings suggest that the current methods for selecting and retaining teachers in primary schools are effective in both public and private sector. Also, the existing organizational academic system based on two-year cycles seems to be a useful educational policy to increase teachers' performance.

This chapter provides robust empirical evidence about the importance of teacher's efficiency on students' outcomes and about the existence of significant differences between teachers inside Spanish primary schools. The measurement of the channels through which these effects operate is not so straightforward. While we found some features that explain part of these differences in teachers' efficiency, a great part still remains to be explained. This part is probably associated with hardly measurable characteristics such as the attitudes of teachers toward students or the teaching practices conducted in class by each teacher. Unfortunately we do not have objective information about these variables in this study to further analyse how to improve classroom management by teachers, and this is a future line of research that we would like to address in the near future. Also, another limitation of this research is that we measured teacher's efficiency based solely on the effect on the cognitive skills of students. It would therefore be an interesting contribution to incorporate non-cognitive skills for a broader measure of the teachers' performance and to compare the results with those obtained in this chapter.

## 3.6 References


Akerhielm, K. (1995). Does class size matter? Economics of Education Review, 14(3), 229-241.

Angrist, J. D., and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly Journal of Economics, 114(2), 533-575.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-1092.

Banker, R. D., Janakiraman, S., and Natarajan, R. (2004). Analysis of trends in technical and allocative efficiency: An application to Texas public school districts. European Journal of Operational Research, 154(2), 477-491.

Barro, R. J., and Lee, J.-W. (2012). A new data set of educational attainment in the world, 1950–2010. Journal of Development Economics, 104, 184–198.

Barro, R. J., and Lee, J. W. (1996). International measures of schooling years and schooling quality. American Economic Review, 86(2), 218-223.

Blackburn, V., Brennan, S., and Ruggiero, J. (2013). Measuring efficiency in Australian Schools: A preliminary analysis. Socio-Economic Planning Sciences.

Boozer, M., and Rouse, C. (2001). Intraschool variation in class size: Patterns and implications. Journal of Urban Economics, 50(1), 163-189.

Calero Martínez, J., Choi de Mendizábal, A., and Waisgrais, S. (2009). Determinantes del rendimiento educativo del alumnado de origen nacional e inmigrante en PISA-2006. Cuadernos económicos de ICE(78), 281-310.

Casalprim, M., Rialp, J., Prior, D., and Sabrià, B. (2013). An Assessment of Schools' Efficiency of Different Educational Systems. Mediterranean Journal of Social Sciences, 4(11), 631.

Case, A., and Deaton, A. (1998). School quality and educational outcomes in South Africa.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. Economics of Education Review, 26(6), 673-682.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. (1966). Equality of educational opportunity. Washington, dc, 1066-5684.

Cooper, S. T., and Cohn, E. (1997). Estimation of a frontier production function for the South Carolina educational process. Economics of Education Review, 16(3), 313-327.

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.

Charnes, A., Cooper, W. W., and Rhodes, E. (1981). Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. Management Science, 27(6), 668-697.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood: National Bureau of Economic Research.

Chudgar, A., and Sankar, V. (2008). The relationship between teacher gender and student achievement: evidence from five Indian states. Compare, 38(5), 627-642.

De La Fuente, A. (2011). Human capital and productivity. Nordic Economic Policy Review, 2, 103-131.

De Witte, K., and Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. Economics of Education Review, 30(4), 641-653.

De Witte, K., and Van Klaveren, C. (2014). How are teachers teaching? A nonparametric approach. Education Economics, 22(1), 3-23.

Dobbelsteen, S., Levin, J., and Oosterbeek, H. (2002). The causal effect of class size on scholastic achievement: distinguishing the pure class size effect from the effect of changes in class composition. Oxford Bulletin of Economics and Statistics, 64(1), 17-38.

Escardíbul, J.-O., and Mora, T. (2013). Teacher Gender and Student Performance in Mathematics. Evidence from Catalonia (Spain). Journal of Education and Training Studies, 1(1), p39-46.

Fuentes, A. (2009). Raising Education Outcomes in Spain: OECD Publishing.

Gordon, R. J., Kane, T. J., and Staiger, D. (2006). Identifying effective teachers using performance on the job: Brookings Institution Washington, DC.

Grosskopf, S., Hayes, K. J., Taylor, L. L., and Weber, W. L. (2001). On the determinants of school district efficiency: Competition and monitoring. Journal of Urban Economics, 49(3), 453-478.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. Journal of Human Resources, 351-388.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. Educational Evaluation and Policy Analysis, 19(2), 141-164.

Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies. The Economic Journal, 113(485), F64-F98.

Hanushek, E. A., and Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. American Economic Review, 90(5), 1184-1208.

Hanushek, E. A., and Rivkin, S. G. (2006). Teacher quality. Handbook of the Economics of Education, 2, 1051-1078.

Hanushek, E. A., and Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. The American Economic Review, 267-271.

Hanushek, E. A., and Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. Annu. Rev. Econ., 4(1), 131-157.

Hanushek, E. A., and Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Journal of Economic Growth, 17(4), 267-321.

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

Heckman, J. J., and Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition: National Bureau of Economic Research.

Hoxby, C. M. (1999). The productivity of schools and other local public goods producers. Journal of Public Economics, 74(1), 1-30.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. The Quarterly Journal of Economics, 115(4), 1239-1285.

INEE (2010). Evaluación General de Diagnóstico 2009. Eduación Primaria. Cuarto Curso. Informe de Resultados. Ministerio de Educación, Madrid.

Jimerson, S., Egeland, B., and Teo, A. (1999). Achievement across time: A longitudinal study of deflections, considering early school and family factors. Journal of Educational Psychology, 91, 116-126.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. Economics of Education Review, 27(6), 615-631.

Kane, T. J., and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation: National Bureau of Economic Research.

Krieg, J. M. (2005). Student gender and teacher gender: What is the impact on high stakes test scores. Current Issues in Education, 8(9), n9.

Krueger, A. B. (1999). Experimental estimates of education production functions. Quarterly Journal of Economics, 114(2).

Levin, H. M. (1974). Measuring efficiency in educational production. Public Finance Quarterly, 2(1), 3-24.

Lindahl, M. (2001). Summer learning and the effect of schooling: Evidence from Sweden: IZA Discussion paper series.

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

McCarty, T. A., and Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In H. O. Fried, C. A. Knox Lovell and S. Smith (Eds.), The measurement of productive efficiency: Techniques and applications (pp. 271-287). New York: Oxford University Press.

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

Ray, S. C. (1991). Resource-use efficiency in public schools: A study of Connecticut data. Management Science, 37(12), 1620-1628.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. Econometrica, 73(2), 417-458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. American Economic Review, 247-252.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. The Quarterly Journal of Economics, 125(1), 175.

Sacerdote, B. (2001). Peer effects with random assignment: results for Dartmouth roommates. Quarterly Journal of Economics, 116(2), 681-704.

Salinas, J., and Santín, D. (2012). Selección escolar y efectos de la inmigración sobre los resultados académicos españoles en PISA 2006. Revista de educación, 358, 382-405.

Schlotter, M., Schwerdt, G., and Woessmann, L. (2011). Econometric methods for causal evaluation of education policies and practices: a non-technical guide. Education Economics, 19(2), 109-137.

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

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

Webbink, D. (2005). Causal effects in education. Journal of Economic Surveys, 19(4), 535-560.

Zinovyeva, N., Felgueroso, F., and Vazquez, P. (2013). Immigration and student achievement in Spain: evidence from PISA. Journal of the Spanish Economic Association.

## 3.7 Figures and Tables

Table 3.1: Criteria for grouping students in primary schools (Principal's questionnaire)

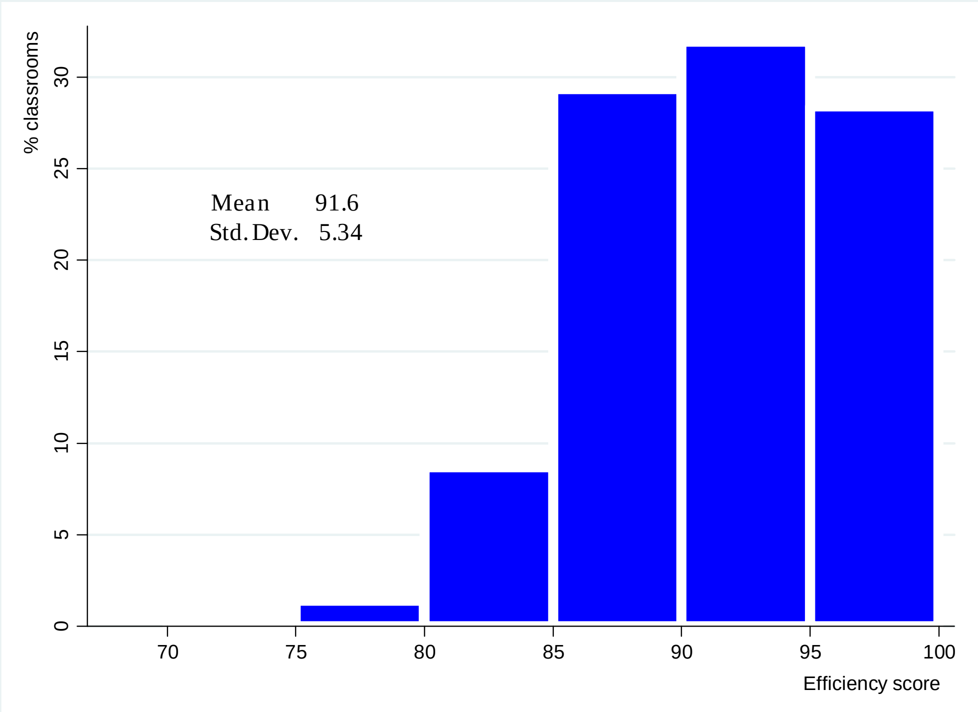
| **Question** | **Grouping criteria adopted in the school** | **Random** |
|---|---|---|
| PD26A | Surnames alphabetical order or other random criteria | YES |
| PD26B | Balance between girls and boys | YES |
| PD26C | Linguistic reasons | NO |
| PD26D | According to academic performance | NO |
| PD26E | Looking for homogeneity of student's characteristics | NO |
| PD26F | Pursuing heterogeneity among students | YES |
| PD26G | Other criteria | NO |

Table 3.2: Bivariate correlations between inputs and outputs

| | Classrooms with random assignment | | | | All classrooms EGD | | | |
|---|---|---|---|---|---|---|---|---|
| | ISECS | PNAT | PCORR | IQER | ISECS | PNAT | PCORR | IQER |
| READ | 0.646** | 0.200** | 0.483** | 0.150** | 0.665** | 0.205** | 0.490** | 0.079** |
| MATHS | 0.627** | 0.204** | 0.454** | 0.142** | 0.634** | 0.171** | 0.467** | 0.093** |

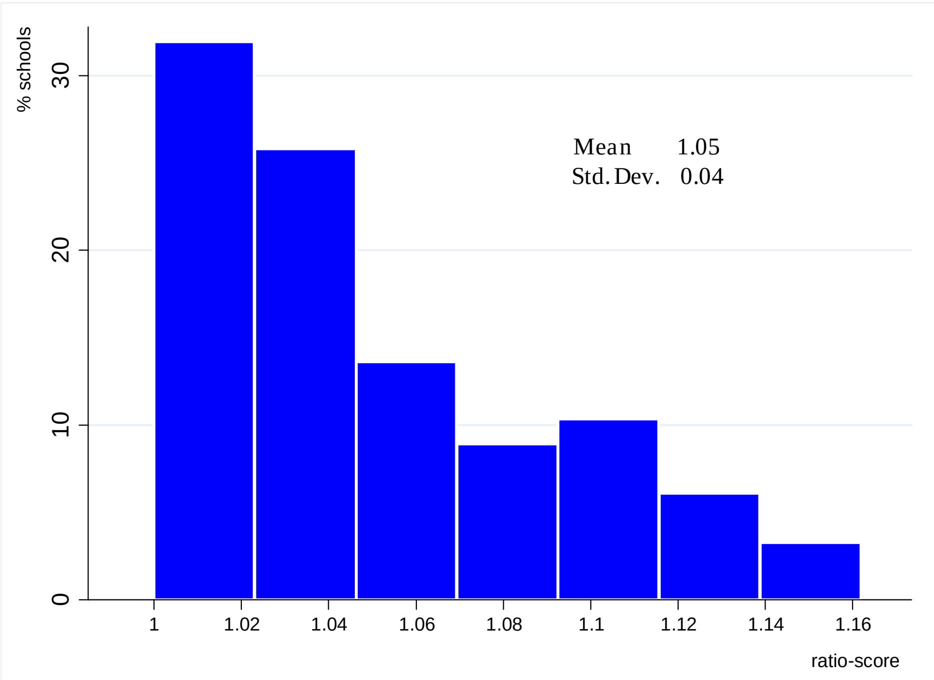
*Notes*: ** Correlation is significant at 1% // Sample size: Groups in schools with random assignment = 426// All groups evaluated in EGD = 1,295

Figure 3.1: Classrooms estimated efficiency distribution in schools with random assignment




*Note*: Sample size = 426 classrooms. Efficiency scores were estimated at classroom level in schools where students were randomly assigned (213 schools)

Figure 3.2: Distribution of teacher's efficiency ratios within schools with random assignment




*Note*: Sample size = 213 schools. Ratios were computed between the most efficient and the least efficient teacher.

Table 3.3: Descriptive statistics of outputs, inputs and teacher's efficiency explanatory variables for classrooms in schools with random assignment

| Variables | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| READ | 508.0 | 42.4 | 354.6 | 599.9 |
| MATHS | 506.6 | 41.9 | 375.4 | 635.7 |
| ISECS | 2.85 | 0.54 | 1.21 | 3.99 |
| PNAT | 0.90 | 0.12 | 0.30 | 1.00 |
| PCORR | 0.90 | 0.08 | 0.45 | 1.00 |
| IQER | 3.29 | 0.91 | 1.00 | 4.73 |
| TEACHgen* | 0.75 | 0.43 | 0.00 | 1.00 |
| TEACHcertified* | 0.77 | 0.42 | 0.00 | 1.00 |
| TEACHgraduated* | 0.19 | 0.39 | 0.00 | 1.00 |
| TEACHexp5* | 0.08 | 0.27 | 0.00 | 1.00 |
| TEACHexp10* | 0.18 | 0.39 | 0.00 | 1.00 |
| TEACHexp30* | 0.38 | 0.49 | 0.00 | 1.00 |
| TEACHschold* | 0.27 | 0.45 | 0.00 | 1.00 |
| TEACHtutor* | 0.76 | 0.43 | 0.00 | 1.00 |
| CLASSIZE | 24.13 | 2.87 | 12.00 | 34.00 |
| PGIRLS | 0.49 | 0.11 | 0.00 | 0.82 |
| PEARLYSCH | 3.78 | 0.39 | 2.32 | 4.73 |
| PMONOPARENTAL | 0.14 | 0.09 | 0.00 | 0.57 |
| PQUARTER4 | 0.23 | 0.10 | 0.00 | 0.50 |
| SCHpublic** | 0.66 | 0.48 | 0.00 | 1.00 |
| SCHrural** | 0.14 | 0.35 | 0.00 | 1.00 |
| SCHcity** | 0.08 | 0.26 | 0.00 | 1.00 |
| PPALfemale** | 0.46 | 0.50 | 0.00 | 1.00 |
| PPALexp5** | 0.24 | 0.43 | 0.00 | 1.00 |

*Notes*: Sample size = 426 classrooms in schools with random assignment // *Reference categories for teacher's dummies variables are: teacher gender male, non certified, non graduated, more than 5 year experience, more than 10 years experience, less than 30 years experience, more than 5 years in the school, non tutor respectively. // **Reference categories for school dummies variables are: private school, non rural, non big city, principal gender male and principal experience more than five years respectively.

Table 3.4: Descriptive statistics of outputs, inputs and classroom's efficiency explanatory variables in all evaluated classrooms in EGD

| **Variables** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|
| READ | 500.8 | 45.9 | 278.0 | 625.8 |
| MATHS | 500.9 | 44.8 | 323.8 | 641.8 |
| ISECS | 2.78 | 0.55 | 1.00 | 4.03 |
| PNAT | 0.89 | 0.13 | 0.17 | 1.00 |
| PCORR | 0.90 | 0.10 | 0.26 | 1.00 |
| IQER | 3.42 | 0.91 | 1.00 | 4.73 |
| TEACHgen* | 0.74 | 0.44 | 0.00 | 1.00 |
| TEACHcertified* | 0.80 | 0.40 | 0.00 | 1.00 |
| TEACHgraduated* | 0.17 | 0.37 | 0.00 | 1.00 |
| TEACHexp5* | 0.10 | 0.30 | 0.00 | 1.00 |
| TEACHexp10* | 0.22 | 0.42 | 0.00 | 1.00 |
| TEACHexp30* | 0.35 | 0.48 | 0.00 | 1.00 |
| TEACHschold* | 0.33 | 0.47 | 0.00 | 1.00 |
| TEACHtutor* | 0.69 | 0.46 | 0.00 | 1.00 |
| CLASSIZE | 22.89 | 4.39 | 5.00 | 36.00 |
| PGIRLS | 0.49 | 0.12 | 0.00 | 1.00 |
| PEARLYSCH | 3.74 | 0.43 | 1.60 | 4.93 |
| PMONOPARENTAL | 0.14 | 0.09 | 0.00 | 0.57 |
| PQUARTER4 | 0.23 | 0.10 | 0.00 | 0.83 |
| SCHpublic** | 0.69 | 0.46 | 0.00 | 1.00 |
| SCHrural** | 0.24 | 0.43 | 0.00 | 1.00 |
| SCHcity** | 0.09 | 0.29 | 0.00 | 1.00 |
| PPALfemale** | 0.48 | 0.50 | 0.00 | 1.00 |
| PPALexp5** | 0.34 | 0.47 | 0.00 | 1.00 |

*Notes*: Sample size = 1,295 (all classrooms evaluated in EGD) // *Reference categories for teacher's dummies variables are: teacher gender male, non certified, non graduated, more than 5 year experience, more than 10 years experience, less than 30 years experience, more than 5 years in the school, non tutor respectively.**Reference categories for school dummies variables are: private school, non rural, non big city, principal gender male and principal experience more than five years respectively.

Table 3.5: Differences of means in students and teachers characteristics between classrooms in schools with random assignment

| | GROUP A | | GROUP B | | Diff_mean | t | p-value |
|---|---|---|---|---|---|---|---|
| | Mean | Std. | Mean | Std. | | | |
| READ | 509.2 | 42.7 | 506.7 | 42.2 | 2.58 | 0.628 | 0.530 |
| MATHS | 506.4 | 41.4 | 506.8 | 42.5 | -0.41 | -0.100 | 0.920 |
| ISECS | 2.868 | 0.547 | 2.841 | 0.526 | 0.03 | 0.536 | 0.592 |
| PNAT | 0.907 | 0.120 | 0.902 | 0.121 | 0.00 | 0.424 | 0.672 |
| PREPEAT | 0.094 | 0.087 | 0.100 | 0.083 | -0.01 | -0.787 | 0.432 |
| PGIRLS | 0.482 | 0.112 | 0.492 | 0.115 | -0.01 | -0.922 | 0.357 |
| PEARLYSCH | 3.771 | 0.391 | 3.780 | 0.382 | -0.01 | -0.242 | 0.809 |
| PMONOPARENTAL | 0.133 | 0.092 | 0.147 | 0.097 | -0.01 | -1.561 | 0.119 |
| PQUARTER4 | 0.226 | 0.090 | 0.234 | 0.100 | -0.01 | -0.863 | 0.388 |
| CLASSIZE | 24.221 | 2.797 | 24.033 | 2.945 | 0.19 | 0.675 | 0.500 |
| TEACHgen* | 0.765 | 0.425 | 0.732 | 0.444 | 0.03 | 0.781 | 0.435 |
| TEACHcertified* | 0.779 | 0.416 | 0.761 | 0.428 | 0.02 | 0.460 | 0.646 |
| TEACHgraduated* | 0.183 | 0.388 | 0.197 | 0.399 | -0.01 | -0.370 | 0.712 |
| TEACHexp5* | 0.080 | 0.272 | 0.075 | 0.264 | 0.00 | 0.181 | 0.857 |
| TEACHexp10* | 0.164 | 0.371 | 0.197 | 0.399 | -0.03 | -0.880 | 0.379 |
| TEACHexp30* | 0.399 | 0.491 | 0.366 | 0.483 | 0.03 | 0.697 | 0.486 |
| TEACHschold* | 0.263 | 0.441 | 0.286 | 0.453 | -0.02 | -0.542 | 0.588 |
| TEACHtutor* | 0.770 | 0.422 | 0.742 | 0.439 | 0.03 | 0.675 | 0.500 |

*Notes*: Sample size 426 (213 in each group). // Group A and Group B are randomly defined.// *Reference categories for teacher's dummies variables are: teacher men, non certified, non graduated,more than 5 year experience, more than 10 years experience, less than 30 years experience, more than 5 years in the school, non tutor respectively.

Table 3.6: Differences in estimated efficiency scores, academic results and students' characteristics between classrooms with the most and the least efficient teacher in schools with random assignment

| | GROUP T | | GROUP C | | Diff_mean | t | p-value |
|---|---|---|---|---|---|---|---|
| | Mean | Std. | Mean | Std. | | | |
| SCORE | 93.78 | 5.10 | 89.40 | 4.72 | 4.38 | 9.283 | 0.000 |
| READ | 517.3 | 41.4 | 498.6 | 44.9 | 18.72 | 4.664 | 0.000 |
| MATHS | 515.6 | 39.6 | 497.6 | 45.1 | 18.06 | 4.548 | 0.000 |
| ISECS | 2.842 | 0.533 | 2.867 | 0.557 | -0.03 | -0.494 | 0.622 |
| PNAT | 0.901 | 0.118 | 0.908 | 0.122 | -0.01 | -0.651 | 0.516 |
| PREPEAT | 0.102 | 0.079 | 0.092 | 0.098 | 0.01 | 1.264 | 0.207 |
| PGIRLS | 0.490 | 0.113 | 0.484 | 0.113 | 0.01 | 0.474 | 0.635 |
| PEARLYSCH | 3.794 | 0.405 | 3.757 | 0.372 | 0.04 | 0.986 | 0.325 |
| PMONOPARENTAL | 0.146 | 0.094 | 0.134 | 0.095 | 0.01 | 1.299 | 0.195 |
| PQUARTER4 | 0.225 | 0.098 | 0.235 | 0.092 | -0.01 | -1.179 | 0.239 |
| CLASSIZE | 24.122 | 2.906 | 24.131 | 2.845 | -0.01 | -0.034 | 0.973 |

*Notes*: Sample size 426 (213 in each group). // Group T and Group C are defined as the group with the most and the least efficient teacher respectively.

Table 3.7: Descriptive statistics of efficiency ratios and explanatory variables in schools with random assignment

| Variables | Mean | Std. Dev. | Min | Max | t | p-value |
|---|---|---|---|---|---|---|
| Efficiency Ratio | 1.05 | 0.04 | 1.00 | 1.162 | 17.78 | 0.000 |
| d_TEACHgen* | 0.02 | 0.58 | -1.00 | 1.00 | 0.59 | 0.554 |
| d_TEACHcertified* | 0.05 | 0.60 | -1.00 | 1.00 | 1.15 | 0.252 |
| d_TEACHgraduated* | -0.02 | 0.57 | -1.00 | 1.00 | -0.60 | 0.548 |
| d_TEACHexp5* | -0.05 | 0.38 | -1.00 | 1.00 | -1.99 | 0.048 |
| d_TEACHexp10* | -0.09 | 0.49 | -1.00 | 1.00 | -2.65 | 0.009 |
| d_TEACHexp30* | 0.08 | 0.64 | -1.00 | 1.00 | 1.83 | 0.068 |
| d_TEACHschold* | -0.07 | 0.57 | -1.00 | 1.00 | -1.82 | 0.071 |
| d_TEACHtutor* | 0.10 | 0.55 | -1.00 | 1.00 | 2.75 | 0.006 |
| d_CLASSIZE* | -0.01 | 1.95 | -13.00 | 13.00 | -0.07 | 0.944 |
| d_PGIRLS* | 0.01 | 0.11 | -0.35 | 0.37 | 0.66 | 0.508 |
| d_PEARLYSCH* | 0.04 | 0.30 | -1.02 | 1.13 | 1.80 | 0.073 |
| d_PMONOPARENTAL* | 0.01 | 0.11 | -0.34 | 0.30 | 1.56 | 0.121 |
| d_PQUARTER4* | -0.01 | 0.13 | -0.41 | 0.29 | -1.21 | 0.227 |

*Notes*: Sample size = 213 schools with random assignment // *Variables in differences were computed as the difference between the group with the most efficient teacher and the group with the least efficient teacher.

Table 3.8: Explanatory factors of teacher's efficiency ratios in schools with random assignment

| Ratio score | Coef. | Robust Std. Err. | P>z | p-value | Marginal effects dy/dx | Std. Err. |
|---|---|---|---|---|---|---|
| Constant | 1.048 | 0.005 | 0.000 | 0.000 | --- | --- |
| d_TEACHgen* | 0.009 | 0.004 | 0.043 | 0.043 | 0.006 | 0.003 |
| d_TEACHcertified* | 0.013 | 0.012 | 0.279 | 0.279 | 0.009 | 0.008 |
| d_TEACHgraduated* | 0.004 | 0.013 | 0.770 | 0.770 | 0.003 | 0.009 |
| d_TEACHexp5* | -0.004 | 0.010 | 0.713 | 0.713 | -0.003 | 0.007 |
| d_TEACHexp10* | 0.010 | 0.008 | 0.227 | 0.227 | 0.007 | 0.006 |
| d_TEACHexp30* | -0.004 | 0.005 | 0.345 | 0.345 | -0.003 | 0.003 |
| d_TEACHschold* | -0.011 | 0.006 | 0.071 | 0.071 | -0.008 | 0.004 |
| d_TEACHtutor* | 0.012 | 0.005 | 0.026 | 0.026 | 0.008 | 0.004 |
| d_CLASSIZE | -0.003 | 0.001 | 0.024 | 0.024 | -0.002 | 0.001 |
| d_PGIRLS | -0.007 | 0.026 | 0.779 | 0.779 | -0.005 | 0.018 |
| d_EARLYSCH | 0.008 | 0.011 | 0.436 | 0.436 | 0.006 | 0.007 |
| d_PMONOPARENTAL | -0.016 | 0.025 | 0.515 | 0.515 | -0.011 | 0.018 |
| d_PQUARTER4 | -0.029 | 0.024 | 0.212 | 0.212 | -0.020 | 0.016 |
| SCHpublic** | 0.005 | 0.006 | 0.403 | 0.403 | 0.003 | 0.004 |
| SCHrural** | 0.008 | 0.009 | 0.403 | 0.403 | 0.006 | 0.007 |
| SCHcity** | 0.007 | 0.008 | 0.355 | 0.355 | 0.005 | 0.006 |
| PPALfemale** | -0.008 | 0.006 | 0.180 | 0.180 | -0.005 | 0.004 |
| PPALexp5** | -0.010 | 0.007 | 0.162 | 0.162 | -0.007 | 0.004 |
| /sigma | 0.039 | 0.002 | | | | |

*Notes*: Sample size = 213 schools // Dependent variable: -teacher efficiency ratio $\Delta\hat{u}_{iT}/\hat{u}_{iC} \geq 1$ // Tobit regression model with 9 left-truncated observations at value 1. // dy/dx = Marginal effects computed at covariates value equal to zero// *Variables in differences were computed as the difference between the group with the most efficient teacher and the group with the least efficient teacher // **Reference categories for school dummies variables are: private school, non rural, non big city, principal gender male and principal experience more than five years respectively.

Table 3.9: Explanatory factors of classrooms efficiency for all classrooms evaluated in EGD

| | Coef. | Bootstrap Std. Err. | z | P>z |
|---|---|---|---|---|
| constant | 90.80 | 2.52 | 35.97 | 0.000 |
| TEACHgen* | 0.926 | 0.380 | 2.440 | 0.015 |
| TEACHcertified* | 0.794 | 1.024 | 0.780 | 0.438 |
| TEACHgraduated* | 1.327 | 1.111 | 1.190 | 0.233 |
| TEACHexp5* | 0.541 | 0.696 | 0.780 | 0.437 |
| TEACHexp10* | 0.139 | 0.591 | 0.240 | 0.814 |
| TEACHexp30* | 0.807 | 0.392 | 2.060 | 0.039 |
| TEACHschold* | -0.782 | 0.498 | -1.570 | 0.116 |
| TEACHtutor* | 0.892 | 0.401 | 2.230 | 0.026 |
| CLASSIZE | -0.097 | 0.048 | -2.040 | 0.042 |
| PGIRLS | -0.846 | 1.390 | -0.610 | 0.543 |
| EARLYSCH | -0.055 | 0.459 | -0.120 | 0.905 |
| PMONOPARENTAL | -6.675 | 1.774 | -3.760 | 0.000 |
| PQUARTER4 | -5.855 | 1.595 | -3.670 | 0.000 |
| SCHpublic** | -0.043 | 0.419 | -0.100 | 0.919 |
| SCHrural** | -0.052 | 0.435 | -0.120 | 0.905 |
| SCHcity** | 2.271 | 0.606 | 3.750 | 0.000 |
| PPALfemale** | 0.083 | 0.343 | 0.240 | 0.809 |
| PPALexp5** | 0.009 | 0.352 | 0.030 | 0.979 |
| /sigma | 5.429 | 0.121 | 44.98 | 0.000 |

*Notes*: Sample size = 1,295 (all classrooms evaluated in EGD) // Dependent variable: Classroom estimated efficiency scores.// Truncated regression model with 64 right-truncated observations at value 100. // Bootstrap replications = 2000.// *Reference categories for teacher's dummies variables are: teacher men, non certified, non graduated,more than 5 year experience, more than 10 years experience, less than 30 years experience, more than 5 years in the school, non tutor respectively. // **Reference categories for school dummies variables are: private school, non rural, non big city, principal gender male and principal experience more than five years respectively

# Concluding Remarks and Future Research

*"One never notices what has been done;*
*one can only see what remains to be done."*
Marie Curie

This section summarizes and discusses the general contributions of this Ph.D. dissertation and exposes some possible future research directions.

The three chapters included in this research provide new insights about how endogeneity affects the estimation of educational technical efficiency and suggest some approaches to deal with this problem. Although this is a very well-known and widespread econometrics problem frequently observed in numerous economic processes, its presence has been overlooked in the context of technical efficiency estimation, partly due to the inexistence of easy alternatives to deal with it. In this regard, this research makes a novel contribution not only by investigating the potential effects of this problem on DEA estimates, but also by providing methodological solutions to tackle this issue in empirical research.

Chapter 1 presented strong evidence to conclude that, although DEA is robust to negative endogeneity (Bifulco and Bretshneider, 2001, 2003 and Ruggiero, 2003), the existence of significant positive endogeneity severely impairs DEA performance. This evidence takes greater significance since, unfortunately, high positive endogenous scenarios are likely to be found in several public sector production processes, especially in education provision. Furthermore, the Monte Carlo simulations revealed that this decline in DEA performance is further driven by the misidentification of the most inefficient units with low level of the endogenous input. As technical efficiency estimates are relative measures, the most efficient units (from which we should learn the best practices) are also misidentified.

It is worth noting that, in the education sector, this misidentification not only has considerable effects on the design of educational policies, but it also reinforces the educational inequalities already associated with the endogeneity problem. The most inefficient schools operate in the most disadvantaged contexts and it is crucial that they are correctly identified so effective policies and practices can be implemented to correct their inefficient behaviour and reverse their current situation.

Drawing on this evidence, as practitioners we wondered how we could deal with this problem in an empirical application when we suspect the presence of endogeneity. This requires both to identify the problem and to correct it. From the Monte Carlo simulations we provided a simple heuristic method to identify the presence of endogenous inputs, which performs correctly in all simulated scenarios. The power of this heuristic relies on the DGP used in the Monte Carlo experiments. Although we try to simulate a simple and flexible DGP that replicates a general production setting, more research would be necessary to generalize the validity of the heuristic method. In addition to the identification of endogenous environments, getting insights from the Instrumental Variable approach in econometrics, we proposed a novel strategy to tackle the endogeneity problem in the estimation of technical efficiency: the '*Instrumental Input DEA*'. The Monte Carlo simulations actually showed that this strategy could accurately deal with the presence of endogenous inputs in the estimation of technical efficiency, therefore allowing us to correctly identify the most inefficient units.

In addition to the theoretical analysis, this research also provides evidence from two empirical applications where the endogeneity problem is present. In Chapter 2 we applied the strategies proposed in Chapter 1 to data from Uruguayan public secondary schools. Using the heuristic method we detected that the socio-economic background of the school was positive and highly correlated with schools' efficiency, and consequently we performed the II-DEA strategy to estimate schools technical efficiency. In Chapter 3, based on the impact evaluation literature, we dealt with the endogeneity problem from an alternative approach. We used data from a natural experiment in Spanish primary schools to estimate teacher's efficiency. Based on a random assignment of students into classrooms within schools we exploited the exogenous variation in technical efficiency between teachers to assess their performance.

Beyond the particular empirical findings for each educational context discussed in chapters 2 and 3, both analysis provide strong evidence that taking or not into account the endogeneity problem can lead to radically different educational public policy recommendations to improve the provision of schooling.

To conclude, this Ph.D. dissertation provides novel answers to important questions but naturally, it also raises other questions and opens the door to new lines of future research. First, the most immediate extension would be to analyse the effects of endogeneity in parametric frontier techniques. Secondly, although the experimental Monte Carlo design tried to replicate a general production setting that is in line with several previous studies, the effectiveness of the proposed heuristic method and the II-DEA strategy depend on the parameters and the functional form assumed. In this vein, to derive the asymptotic properties of both strategies could be a potential contribution.

Third, from the pioneer work of Charnes et al. (1978) and Banker et al. (1981) several extensions of the DEA model have been developed to improve its robustness (for example to deal with the presence of outliers, special types of data or non-discretionary inputs in the model). In this sense, it is expected that the same problems affecting the DEA performance could affect the performance of these extensions. Thus, a natural and attractive future line of research could be to extent the analysis conducted in this research to other non-parametric efficiency techniques (Free Disposal Hull, order-m, order-alpha, total factor productivity indexes based on DEA, conditional efficiency models and so on).

Finally, both strategies to tackle the endogeneity problem proposed in this research are rooted in the causal inference literature. In this direction, it would be a promising future research line to attempt to combine other existing impact evaluation techniques (e.g. differences in differences, discontinuity regressions or propensity score matching) with non-parametric frontier methods to measure efficiency.

# Abstract

## Essays on the estimation of educational technical efficiency under endogeneity

### Introduction

The evaluation of technical efficiency in the Public Sector has gained growing attention over the last decades. Public services providers have a natural interest in efficiency assessments since they face up increasing demands of quantities and quality together with financial constraints. Within this framework, the measurement of educational technical efficiency is one of the current major concerns as the education expenditure is one of the largest public budget items and the public sector is usually the main provider of education in most modern countries.

Given that the investment in quality education is essential to ensure sustainable development and economic growth (Barro and Lee, 1996, 2012; Hanushek and Kimbo, 2000; De la Fuente, 2011; Hanushek and Woessmann, 2012a, 2012b), several countries in the last decades have significantly increased their public educational budget. However, these efforts have not always been translated into better academic achievements. This fact has led researchers and policy-makers to wonder why these additional investments in educational resources do not lead to improvements in the quality of education. Although the answer is not evident, this fact alerts about the presence of great inefficiencies in schooling production and has spurred the interest in measuring these inefficiencies and explaining their main sources, with the ultimate goal of correcting these behaviours.

The educational production has, like most public sector production processes, some special characteristics that complicate the estimation of accurate efficiency measures (i.e. the completely unknown production technology, the lack of prices information or the frequent use of multiple proxy variables to approximate the real output). In this sense, non-parametric techniques and particularly the DEA model proposed by Charnes, Cooper and Rhodes (1978) and Banker, Charnes and Cooper (1984) are the most commonly applied methods for measuring educational technical efficiency (Worthington, 2001). This is mainly because of two reasons: its flexibility allows to adapting it to the stated particularities of this sector, and the results of this technique can be easily translated to stakeholders and politicians.

However, there is a major concern frequently observed in educational production processes which has been overlooked in the context of the technical efficiency estimation: the endogeneity problem. In statistical terms, this phenomenon implies the presence of a significant correlation between one input and the error term, and it can arise as the result of multiple sources (e.g. measurement errors, unobserved heterogeneity, the omission of relevant variables in the model specification or the presence of simultaneity). In the context of the estimation of technical efficiency with frontier techniques, this problem of endogeneity implies the presence of a significant correlation between at least one input and the efficiency term (Peyrache and Coelli, 2009).

In the education provision framework, the most common source of endogeneity is the educational self-selection. Students are not exogenously assigned to schools but their allocation depends on decisions made by parents, teachers and schools' principals. Indeed, this problem has been one of the focuses of attention in econometrics along the last three decades. Endogeneity has been argued to be the basis for multiple theoretical and empirical critiques of traditional findings and multiple methods have been developed in the literature to deal with this problem (Webbink 2005, Schlotter et al. 2011).

However, this widespread acknowledgement in the context of econometrics of the existence of the self-selection or the endogeneity problem is ignored when we move into the world of the efficiency estimation. There are only a handful of studies that using alternative simulation strategies have tested the performance of DEA under some kind of endogeneity (Gong and Sickles, 1992; Orme and Smith, 1996; Bifulco and Bretschneider, 2001, 2003; Ruggiero, 2003, 2004). Consequently, this problem is still an unknown and incipient issue in the literature of the estimation of frontiers using DEA and thus it is frequently overlooked when practitioners apply this technique.

## Objectives and Results

Based on this background, the present Ph.D. dissertation aims to contribute theoretically and empirically to understand the extent to which the endogeneity problem, a major concern frequently observed in educational production processes, affects the estimation of technical efficiency using the Data Envelopment Analysis (DEA) technique. Furthermore, this research combines insights from impact evaluation literature and non-parametric frontier techniques in order to provide potential solutions to deal with this problem in educational empirical applications and obtain more accurate efficiency estimates.

Chapter 1 analyses theoretically to which extent does the presence of endogeneity in the production process affect DEA estimates in finite samples, so practitioners performing this technique can be aware of the accuracy of their estimates. To do this, we firstly illustrate the endogeneity problem and its implications for the efficiency estimation from a conceptual perspective. Secondly, using synthetic data generated in a Monte Carlo experiment we evaluate how different levels of positive and negative endogeneity can affect DEA performance. We conclude that, although DEA is robust to the negative endogeneity (Bifulco and Bretshneider, 2001, 2003 and Ruggiero, 2003), a significant positive endogeneity severely impair DEA performance.

Building upon this evidence, the question that arises is: how can we deal with this problem in an empirical application if we suspect for the presence of endogeneity? This requires both to identify the problem and to correct it. From the Monte Carlo simulations we provided a simple heuristic method to identify the presence of endogenous inputs, which performs correctly in all simulated scenarios. In addition, getting insights from the Instrumental Variable approach in econometrics, we proposed a novel strategy to tackle the endogeneity problem in the estimation of technical efficiency: the '*Instrumental Input DEA*'. The Monte Carlo simulations actually showed that this strategy could accurately deal with the presence of endogenous inputs in the estimation of technical efficiency, therefore allowing us to correctly identify the most inefficient units.

In Chapter 2 we applied the strategies proposed in Chapter 1 to data from Uruguayan public secondary schools. Using the heuristic method we detected that the socio-economic background of the school was positive and highly correlated with schools' efficiency, and consequently we performed the II-DEA strategy to estimate schools technical efficiency. Beyond estimating the efficiency potential improvements for each school and identifying the better and the worst performers, we aim to explore the explanatory factors of the efficient behaviours. Thus, once we have estimated the II-DEA efficiency scores we regress them on several contextual variables related to students and schools characteristics. The results of this second stage allow us to draw conclusions about which educational policies and practices would be desirable to design and promote in order to improve the quality of education.

The II-DEA strategy proposed in the first chapter and implemented in Chapter 2 requires finding a good instrument. This is not an easy task and, in some contexts, it may not even be possible to find one. In the third chapter, taking again insights from the impact evaluation literature we provide an alternative strategy to deal with the endogeneity problem in the estimation of educational technical efficiency.

In Chapter 3, based on the impact evaluation literature, we dealt with the endogeneity problem from an alternative approach. We used data from a natural experiment in Spanish primary schools to estimate teacher's efficiency. Based on a random assignment of students into classrooms within schools we exploited the exogenous variation in technical efficiency between teachers to assess their performance. This strategy allows us to obtain an unbiased measure of the true teacher's effect on students' achievement and to explore the main drivers of teachers' efficiency. As in the previous chapter, we also perform the analysis without taking into account the presence of self-selection to empirically quantify the effect of this problem in terms of educational public recommendations.

## Conclusions

In conclusion, this research provides new insights about how endogeneity affects the estimation of educational technical efficiency and suggests some approaches to deal with this problem. Chapter 1 presented strong evidence to conclude that, although DEA is robust to negative endogeneity (Bifulco and Bretshneider, 2001, 2003 and Ruggiero, 2003), the existence of significant positive endogeneity severely impairs DEA performance. This evidence takes greater significance since, unfortunately, high positive endogenous scenarios are likely to be found in several public sector production processes, especially in education provision. Furthermore, the Monte Carlo simulations revealed that this decline in DEA performance is further driven by the misidentification of the most inefficient units with low level of the endogenous input. As technical efficiency estimates are relative measures, the most efficient units (from which we should learn the best practices) are also misidentified.

Furthermore, the Monte Carlo simulations revealed that this decline is further driven by the misidentification of the most inefficient units with low level of the endogenous input. As technical efficiency estimates are relative measures, the most efficient units (from which we should learn the best practices) are also misidentified. It is worth to note, that in the education sector this misidentification has not only considerable effects on the design of educational policies but also it reinforces the educational inequalities already caused by the endogeneity problem. The most inefficient schools operate in most disadvantaged context and thus, they should implement effective policies and practices to correct their inefficient behaviour and reverse their current situation.

In addition to the theoretical analysis, this research also provides evidence from two empirical applications where the endogeneity problem is present. Beyond the particular empirical findings for each educational context discussed in chapters 2 and 3, both analysis provide strong evidence that taking or not into account the endogeneity problem can lead to radically different educational public policy recommendations to improve the provision of schooling.

To conclude, this Ph.D. dissertation provides novel answers to important questions but naturally, it also raises other questions and opens the door to new lines of future research. First, the most immediate extension would be to analyse the effects of endogeneity in parametric frontier techniques. Secondly, although the experimental Monte Carlo design tried to replicate a general production setting that is in line with several previous studies, the effectiveness of the proposed heuristic method and the II-DEA strategy depend on the parameters and the functional form assumed. In this vein, to derive the asymptotic properties of both strategies could be a potential contribution.

Third, from the pioneer work of Charnes et al. (1978) and Banker et al. (1981) several extensions of the DEA model have been developed to improve its robustness (for example to deal with the presence of outliers, special types of data or non-discretionary inputs in the model). In this sense, it is expected that the same problems affecting the DEA performance could affect the performance of these extensions. Thus, a natural and attractive future line of research could be to extent the analysis conducted in this research to other non-parametric efficiency techniques (Free Disposal Hull, order-m, order-alpha, total factor productivity indexes based on DEA, conditional efficiency models and so on).

Finally, both strategies to tackle the endogeneity problem proposed in this research are rooted in the causal inference literature. In this direction, it would be a promising future research line to attempt to combine other existing impact evaluation techniques (e.g. differences in differences, discontinuity regressions or propensity score matching) with non-parametric frontier methods to measure efficiency.

# Resumen

## Ensayos sobre la estimación de la eficiencia técnica bajo la presencia de endogeneidad

### Introducción

El estudio de la medición de la eficiencia técnica en el sector público ha crecido notoriamente en las últimas décadas. Los proveedores de servicios públicos tienen un interés natural en medir la eficiencia, producto de las crecientes demanda de servicios y restricciones presupuestarias que éstos enfrentan diariamente. En este contexto, la medición de la eficiencia técnica educativa es una de las principales preocupaciones actuales dado que el gasto en educación es una de las mayores partidas del presupuesto público y que el sector público por lo general es el principal proveedor de la educación en la mayoría de los países modernos.

Teniendo en cuenta que la inversión en educación de calidad que realiza un país es esencial para asegurar su desarrollo y crecimiento económico sostenible (Barro y Lee, 1996, 2012; Hanushek y Kimko, 2000; De la Fuente, 2011; Hanushek y Woessmann, 2012a, 2012b), varios países en las últimas décadas han aumentado considerablemente su presupuesto público educativo. Sin embargo, estos esfuerzos no siempre se han traducido en mejores logros académicos. Este hecho ha llevado a los investigadores y los responsables políticos a preguntarse por qué estas inversiones adicionales en recursos educativos no dan lugar a mejoras en la calidad de la educación. A pesar de que la respuesta no es evidente, este hecho alerta sobre la presencia de ineficiencias en la producción educativa. Por tanto, no sorprende el gran interés en la medición de estas ineficiencias y en intentar explicar sus principales fuentes para corregir estos comportamientos.

La producción educativa, al igual que la producción del sector público, tiene características especiales (por ejemplo, el desconocimiento de la tecnología de producción, falta de información de los precios o el carácter multidimensional de la producción) que dificultan la estimación de la eficiencia (Bowlin, 1986). En este sentido, las técnicas no paramétricas y en particular el modelo DEA propuesto por Charnes, Cooper y Rhodes (1978) y Banker, Charnes y Cooper (1984) han sido los métodos más comúnmente aplicados para medir la eficiencia técnica educativa

(Worthington, 2001). La razón principal radica en su flexibilidad que permite adaptarse a las particularidades del sector mencionadas anteriormente, y que los resultados de esta técnica puede ser fácilmente traducido a los diversos agentes y políticos involucrados en el proceso educativo.

Sin embargo, existe un problema mayor observado en los procesos de producción educativa y que ha sido pasado por alto en el contexto de la estimación de la eficiencia técnica: la presencia de endogeneidad. En términos estadísticos, este fenómeno implica la presencia de una correlación significativa entre uno de los inputs y el término de error. En el contexto de la estimación de la eficiencia técnica, el problema de endogeneidad implica la presencia de una correlación significativa entre uno de los inputs y el término de eficiencia (Peyrache y Coelli, 2009).

En el ámbito educativo la causa más frecuente de endogeneidad está asociada a la auto-selección escolar. En general los alumnos no son asignados aleatoriamente a los colegios, sino que por el contrario, su distribución depende de las decisiones de padres, profesores y directores. En efecto, este problema ha sido uno de los principales focos de atención principal de la econometría en las últimas tres décadas. El problema de la endogeneidad ha sido la base de múltiples críticas teóricas y empíricas a los resultados tradicionales en economía de la educación y múltiples métodos han sido desarrollados en la literatura para poder hacer frente a este problema (Webbink 2005, Schlotter et al. 2011).

Sin embargo, este amplio reconocimiento de la existencia de la auto-selección escolar o el problema de endogeneidad es ignorado cuando nos movemos al mundo de la estimación de la eficiencia. Existen escasos estudios previos que utilizando estrategias de simulación alternativas han testeado el desempeño de DEA bajo la presencia de algún tipo de endogeneidad (Gong y Sickles, 1992; Orme y Smith, 1996; Bifulco y Bretschneider, 2001, 2003; Ruggiero, 2003, 2004). Por tanto, este problema sigue siendo un tema desconocido e incipiente en la literatura de la estimación de fronteras utilizando DEA y por ende es un problema frecuentemente ignorado por los investigadores al aplicar esta técnica.

## Objetivos y Resultados

En base a estos antecedentes, la presente Tesis Doctoral tiene como objetivo contribuir, teórica y empíricamente, a entender hasta qué punto el problema de endogeneidad, uno de los principales problemas observado frecuentemente en los procesos de producción educativos, afecta a la estimación de la eficiencia técnica mediante el Análisis Envolvente de Datos (DEA). Asimismo, esta investigación combina ideas de la literatura de evaluación de impacto con las técnicas de medición de eficiencia no paramétricas con el fin de aportar potenciales soluciones para hacer frente a este problema en aplicaciones empíricas educativas y obtener así estimaciones de la eficiencia más precisas.

El Capítulo 1 analiza teóricamente en qué medida la presencia de endogeneidad en el proceso de producción puede afectar a las estimaciones DEA en muestras finitas, de modo que los investigadores que aplican esta técnica conozcan la precisión de sus estimaciones. Para ello, en primer lugar se ilustra desde un punto de vista conceptual el problema de la endogeneidad y sus implicaciones en la estimación de la eficiencia. En segundo lugar, utilizando datos generados en un experimento de Monte Carlo evaluamos cómo diferentes niveles de endogeneidad positiva y negativa pueden afectar al desempeño de DEA.

A partir de los resultados hallados previamente, la siguiente pregunta que surge es ¿Cómo podemos hacer frente a este problema en una aplicación empírica cuando sospechamos de la presencia de endogeneidad? Esto implica responder dos cuestiones: cómo identificar el problema y cómo enfrentarlo. A partir de las simulaciones de Monte Carlo se propone un método heurístico sencillo que permite identificar correctamente la presencia de inputs endógenos en todos los escenarios simulados. Además, a partir de la técnica de Variables Instrumentales (VI) ampliamente utilizada en econometría, ofrecemos una nueva estrategia para abordar el problema de endogeneidad en la estimación de la eficiencia técnica: el "Instrumental Input DEA". Las simulaciones de Monte Carlo evidencian que esta estrategia propuesta permite abordar adecuadamente la presencia de los inputs endógenos en la estimación de la eficiencia técnica ya que identifica correctamente las unidades más ineficientes.

En el capítulo 2 se aplican las estrategias propuestas en el Capítulo 1 a datos de colegios públicos de educación secundaria en Uruguay. Utilizando el método heurístico detectamos que el nivel socio-económico medio de los colegios está alta y positivamente correlacionado con la eficiencia técnica de los mismos, y por lo tanto aplicamos la estrategia II-DEA para estimar la eficiencia técnica de los colegios controlando por endogeneidad. Más allá de la estimación de las potenciales mejoras de eficiencia para cada colegio y de identificar a los mejores y peores, el objetivo es explorar los factores explicativos de los comportamientos eficientes. Por lo tanto, una vez que han sido estimados los índices de eficiencia II-DEA éstos se regresan sobre diversas variables contextuales que caracterizan a los estudiantes y a los colegios. Los resultados de esta segunda etapa permiten extraer conclusiones acerca de cuáles políticas y prácticas educativas serían deseables de diseñar y promover con el fin de mejorar la calidad de la educación.

La estrategia II-DEA propuesta en el primer capítulo e implementada en el capítulo 2 requiere encontrar un buen instrumento lo cual no es una tarea fácil y en algunos contextos, incluso no es posible encontrar uno. En el tercer capítulo, tomando nuevamente ideas de la literatura de evaluación de impacto se proporciona una estrategia alternativa para tratar el problema de endogeneidad en la estimación de la eficiencia técnica educativa.

En el capítulo 3 se utilizan datos de un experimento natural en las escuelas de educación primaria en España para estimar la eficiencia de los maestros. En base a la asignación aleatoria de los estudiantes a las clases dentro de los colegios explotamos la variación exógena de la eficiencia técnica entre los maestros para evaluar su desempeño. Esta estrategia nos permite obtener una medida objetiva del verdadero efecto del maestro sobre los logros de los estudiantes y explorar los principales factores que explican la eficiencia de los docentes.

## Conclusiones

En conclusión, esta investigación proporciona nuevos conocimientos sobre cómo el problema de la endogeneidad afecta la estimación de la eficiencia técnica educativa y provee algunas estrategias para hacer frente a este problema.

El Capítulo 1 evidencia que a pesar de que DEA es robusto a la presencia de endogeneidad negativa (Bifulco y Bretshneider, 2001, 2003 y Ruggiero, 2003), la existencia de una endogeneidad positiva y significativa perjudica gravemente el desempeño de DEA. Estos resultados tienen especial relevancia, ya que, lamentablemente, los escenarios de endogeneidad positiva y alta son los que se encuentran con mayor probabilidad en varios procesos de producción del sector público y sobre todo en la provisión de educación.

Por otra parte, las simulaciones de Monte Carlo revelan que este deterioro en la técnica es impulsado principalmente por la identificación errónea de las unidades más ineficientes con bajos niveles del input endógeno. Dado que las estimaciones de la eficiencia técnica son medidas relativas, esta correcta identificación implica que también se identifiquen incorrectamente a las unidades más eficientes (de las que deberíamos aprender las mejores prácticas). Vale la pena destacar, que en el sector de la educación esta identificación errónea no sólo tiene efectos considerables en el diseño de políticas educativas sino que también refuerza las desigualdades educativas ya causadas por la presencia de endogeneidad. Los centros educativos más ineficientes operan en contextos más desfavorecidos y, por tanto, deberían ser los primeros en implementar políticas y prácticas educativas efectivas para corregir sus comportamientos ineficientes y revertir su situación actual.

Adicionalmente a este ejercicio teórico, los capítulos 2 y 3 proporcionan evidencia de dos aplicaciones empíricas en el que el problema de endogeneidad está presente. Más allá de los resultados concretos de cada contexto educativo analizado (que se discuten en cada capítulo), ambos análisis proporcionan evidencia robusta de que el tomar o no en consideración el problema de endogeneidad conduce a resultados radicalmente diferentes en términos de las recomendaciones de política educativa pública para mejorar la calidad de la enseñanza.

Para finalizar, esta Tesis Doctoral proporciona nuevas respuestas a preguntas relevantes, pero, naturalmente, también plantea nuevas interrogantes y abre las puertas a diversas líneas de investigación futuras. En primer lugar, la contribución más inmediata sería extender el análisis de los potenciales efectos de la endogeneidad sobre las técnicas de frontera paramétricas. En segundo lugar, aunque el diseño experimental de las simulaciones Monte Carlo intenta replicar un contexto de producción general y está en línea con la mayoría de los estudios previos, la eficacia del método heurístico propuesto y la estrategia II-DEA depende de los parámetros y la forma funcional asumidos. En este sentido, derivar las propiedades asintóticas de ambas estrategias serían contribuciones prometedoras ya que permitirían generalizar las conclusiones de la presente investigación.

En tercer lugar, desde el trabajo pionero de Charnes et al. (1978) y Banker et al. (1981) se han desarrollado diversas extensiones del modelo DEA para mejorar su robustez (por ejemplo, ante la presencia de valores atípicos, de datos con características especiales o para incluir inputs no discrecionales en el modelo). En este sentido, es de esperar que los mismos problemas que afectan el desempeño de DEA pudieran afectar el desempeño de estas extensiones. Por lo tanto, una línea de investigación futura natural y atractiva sería extender el análisis realizado en esta investigación a otras técnicas de eficiencia no paramétricas (por ejemplo FDH, orden-m, orden-alfa, índices de productividad total de factores basados en DEA, modelos de eficiencia condicionada).

Por último, ambas estrategias propuestas en esta investigación para abordar el problema de endogeneidad están basadas en la literatura inferencia causal. En este sentido, sería interesante hacer el esfuerzo de adaptar otras técnicas de evaluación de impacto existentes (diferencias en diferencias, regresiones de discontinuidad, Propensity Score Matching, etc.) al contexto de la medición de la eficiencia utilizando métodos de frontera no paramétricos.