

## Dealing with endogeneity in data envelopment analysis applications

Daniel Santín

*Department of Applied Economics VI, Universidad Complutense de Madrid, 28223, Madrid, Spain.*

dsantin@ccee.ucm.es

Gabriela Sicilia\*

*Department of Economics and Public Finance, Universidad Autónoma de Madrid, 28049, Madrid, Spain.*

gabriela.sicilia@uam.es

### Abstract

Although the presence of the endogeneity is frequently observed in economic production processes, it tends to be overlooked when practitioners apply data envelopment analysis (DEA). In this paper we deal with this issue in two ways. First, we provide a simple statistical heuristic procedure that enables practitioners to identify the presence of endogeneity in an empirical application. Second, we propose the use of an instrumental input DEA (II-DEA) as a potential tool to address this problem and thus improve DEA estimations. A Monte Carlo experiment confirms that the proposed II-DEA approach outperforms standard DEA in finite samples under the presence of high positive endogeneity. To illustrate our theoretical findings, we perform an empirical application on the education sector.

Keywords: Data envelopment analysis (DEA), Endogeneity, Simulation, Education

---

\* Corresponding author: Gabriela Sicilia. Tel: +34 914976160. E-mail: [gabriela.sicilia@uam.es](mailto:gabriela.sicilia@uam.es)  
Postal address: Department of Economics and Public Finance, Cantoblanco Campus Mod-6 Off.601, 28049 Madrid, Spain.

## 1. Introduction

Data envelopment analysis (DEA) has been applied in a wide range of contexts to empirically measure the technical efficiency of a set of Decision Making Units for *benchmarking* their performance, identifying the best practices and/or correcting inefficient behaviours among other purposes (Siford, 1997; Gattoufi et al. 2004a, 2004b; Emrouznejad et al. 2008). The main reasons of its widespread use are its flexibility and the few assumptions needed about the implicit technology that relates inputs to outputs (only some axiomatic assumptions, *i.e.* monotonicity and concavity) or about the distribution of inefficiency. Basically, DEA consists in applying a linear optimization program in order to obtain a production frontier that includes all the observed efficient units, along with their linear combinations, and leaves the rest of (inefficient) DMUs below it. Thus, the estimated technical efficiency score for each observed DMU in the sample (*i.e.* the distance to the ‘best-practice frontier’) is a relative measure calculated using all the production units that are compared.

Since the publication of the seminal papers by Charnes, Cooper and Rhodes (1978) and Banker, Charnes and Cooper (1984) several methodological extensions of the technique have been developed over the last three decades to improve its robustness (Emrouznejad et al., 2008, Cook and Seiford, 2009; Liu, et al. 2013). These contributions have been developed in many directions, For example, to deal with the selection of appropriate inputs and outputs to include in DEA specifications (Smith, 1997; Nataraja and Johnson, 2011; Cook et al., 2014), to deal with inclusion of non-discretionary inputs (Banker and Morey, 1986; Ruggiero, 1996; Muñiz et al. 2006; Cordero et al. 2009), to deal with fuzzy data (Liu and Chuang, 2009; Liu, 2011; Hatami-Marbini et al., 2012), to improve DEA estimations under the presence of noise (Gstach, 1998; Ruggiero, 2006; Simar, 2007), or to deal with complex input and output structures (Wang and Chin, 2010; Lotfi et al., 2011; Yu and Shi, 2014; Jianfeng, 2015), among others.

However, a key concern that plays a major role in the economics and management literature has been overlooked in the DEA literature: the presence of endogeneity (Angrist and Pischke, 2014). In the statistical framework, endogeneity arises when the assumption that all inputs or covariates are uncorrelated with the error term does not hold (Greene, 2003; Wooldridge, 2012). In the context of the measurement of technical efficiency using non-parametric techniques, this

concept implies a correlation between an input and the efficiency term (Orme and Smith, 1996; Bifulco and Bretschneider, 2001; Peyrache and Coelli, 2009)<sup>2</sup>.

Although the presence of the endogeneity is frequently observed in economic production processes, DEA is commonly run under the implicit assumption that the technical efficiency of the evaluated set of DMUs is exogenously distributed, *i.e.* assuming that the efficiency term is uncorrelated with the inputs and outputs. Recently, based on previous research (Gong and Sickles, 1992; Orme and Smith, 1996; Bifulco and Bretschneider, 2001, 2003; Ruggiero, 2003, 2004), Cordero, et al. (2015) found that DEA estimates could be severely impaired if the efficiency term is highly and positively correlated with one input. Such positive correlation is frequently observed in production processes where DEA is in widespread use. For example, this is the case when larger firms have access to better managers (more efficient). Since they produce more outputs and consume more inputs than small firms, we observe a correlation between technical efficiency and input levels (Wilson, 2003). Also, a positive correlation between inputs and technical efficiency is often observed in the context of public services provision due to the presence of additional demand-side relationships (Mayston, 2003)<sup>3</sup>. Finally, the education sector is another good example where positive endogeneity very frequently arises due to the school self-selection problem (Webbink, 2005; Shottler et al. 2011; De Witte and López-Torres, 2015). In all these contexts, the empirical estimation of technical efficiency using DEA models could lead to misleading efficiency estimates and thus to inappropriate performance-based recommendations.

Based upon these findings, two key challenges now emerge: how to detect the presence of endogeneity and how to improve DEA estimations in empirical applications. Although these issues have received considerable attention in statistics and econometrics, few previous works have addressed them in the context of the estimation of technical efficiency using DEA.

Wilson (2003) explores a number of relatively 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 properties and low power in moderate sample sizes. Based on this work, Peyrache and Coelli (2009) propose a semi-parametric Hausmann-type asymptotic test for linear independence of all inputs and outputs

---

<sup>2</sup> Kousmanen and Johnson (2010, p.152) clarify the use of this definition demonstrating in their work that the DEA problem in the single-output setting can be interpreted as a nonparametric least-squares model under the assumption of an error term component  $u$  equal or less than zero  $u \leq 0$ .

<sup>3</sup> See Mayston (2016) for a very interesting, illustrative and exhaustive discussion about when endogeneity can arise in public policy issues and public services.

included in the DEA model. Using a Monte Carlo experiment, they show that this test has good size and power properties in finite samples. However, since the null hypothesis of these proposed tests include all the inputs and outputs, we cannot identify which inputs are endogenous or exogenous in the production process. Moreover, we cannot discern if inputs are positively or negatively correlated with efficiency since these are all two-sided tests. As Cordero et al. (2015) show, it is only the presence of medium or high positive correlation that damages DEA performance. Finally, as far as we know, only Mayston (2016) has proposed a potential method to correct DEA estimations under the presence of positive endogeneity caused by the existence of inter-relationships between inputs and the quality of the outputs. He proposed to adapt the standard DEA by estimating an Achievement Possibility Frontier which is defined as the maximum feasible output quality that DMUs can achieve given the exogenous environmental variables which it faces.

This paper aims to contribute to this literature in both directions. First, we propose a simple statistical heuristic procedure to identify the presence of a positive endogenous input in an empirical production problem. Second, we introduce a potential solution based on the use of an instrumental input to improve DEA estimations. Finally, we perform an empirical application to the education sector in order to illustrate how to apply our proposals and to empirically size up the impact of not controlling for endogeneity. To the best of our knowledge, this is the first research that provides new tools to deal with both problems: the identification and the correction of endogeneity in DEA empirical applications.

The paper is organized as follows. Section 2 introduces a strategy to identify endogenous settings. Section 3 describes the II-DEA method and reports a Monte Carlo experiment to test its performance in finite samples. Section 4 describes the empirical application. Section 5 outlines the conclusions and provides some directions for future research.

## **2. Heuristic for identifying endogenous inputs in DEA**

From a microeconomic viewpoint and assuming that inputs are exogenous, the expected correlation coefficient between the  $p$  inputs  $x_{ik}$  and the DEA estimated efficiency scores  $\hat{\theta}_i$  for a set of  $n$  DMUs,  $n = 1, 2, \dots, N$ , should be close to zero. Therefore, our proposed heuristic method is based on comparing these expected correlation coefficients under the assumption that a

setting is exogenous or endogenous in order to classify each input type included in the DEA model.

To do this, we first run the Monte Carlo experiment described in Cordero et al. (2015, p. 515) using the same data generation process (DGP). Therefore, to emulate the production technology, we assume both a Cobb-Douglas and a *translog* technology in a single output setting with three inputs<sup>4</sup>. The Cobb-Douglas is defined as

$$\ln y_i = \alpha_1 \ln x_{1i} + \alpha_2 \ln x_{2i} + \alpha_3 \ln x_{3i} \quad (1)$$

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 are  $\alpha_1 = 0.3$ ,  $\alpha_2 = 0.35$  and  $\alpha_3 = 0.35$ . The *translog* production function introduced by Christensen *et al.* (1971) is defined as

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

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, fulfilling the monotonicity and concavity assumptions, within the bounds imposed by the inputs distribution. In this baseline scenario we randomly and independently generate three input vectors  $x_1$ ,  $x_2$  and  $x_3$  using a uniform distribution over the interval  $[5, 50]$  for  $N$  DMUs,  $i = 1, 2, \dots, N$ . To simulate the efficiency level, we randomly and independently generate  $N$  values of  $u_i$  using a half-normal distribution  $u_i \approx |N(0; 0.25)|$  and compute the true technical efficiency level for each DMU as  $\theta_i = \exp(-u_i)$ , where  $0 \leq \theta_i \leq 1$ . In order to represent a more realistic set, we simulate a small random error term  $v_i$  following a normal distribution  $v_i \approx N(0; 0.04)$ , representing the random statistical perturbation in the production function. Finally we compute the observed output as:  $\hat{y}_i = \exp(\ln y_i) \cdot \exp(v_i) \cdot \theta_i$  using Equations (1) and (2) to compute  $\ln y_i = f(\cdot)$  respectively.

<sup>4</sup> Although the Cobb-Douglas is the most commonly used function in the operational research and economics literature, the *translog* is a more realistic production process, since the constant input-output elasticities of the Cobb-Douglas technology are difficult to assume in many real world applications. Thus, for robustness, we run our simulations for both functions.

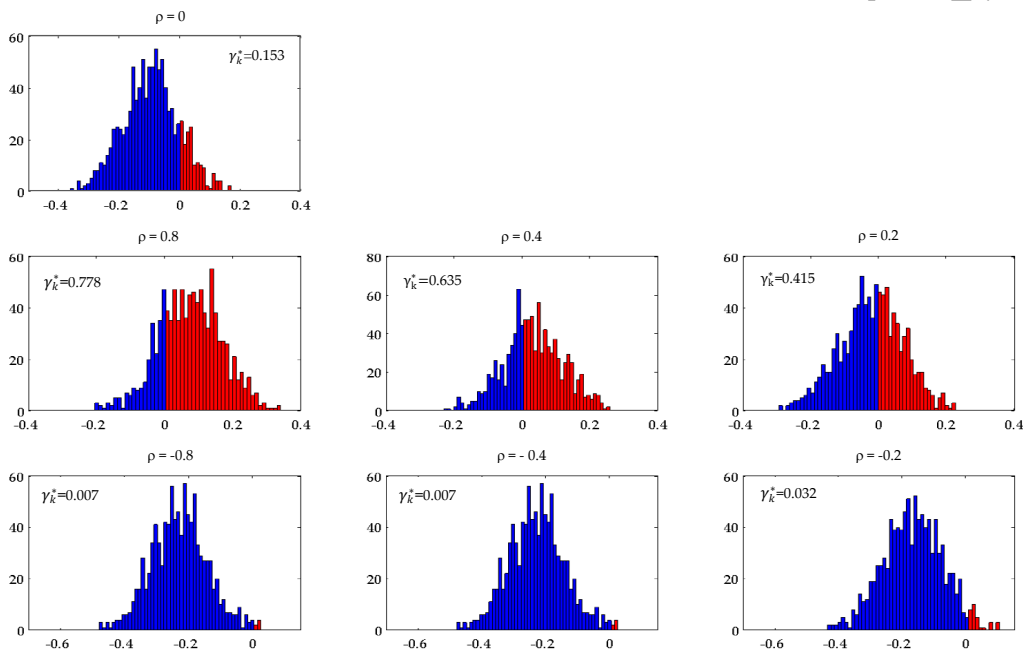
The remaining six endogenous scenarios were developed by a similar DGP, but taking into account the existence of endogeneity. Therefore, in each endogenous dataset we replicate the baseline DGP abovementioned, but substitute the exogenous input  $x_3$  by an endogenous input  $X^e$  which is also uniformly distributed between [5,50] and correlated with the true efficiency  $\theta_i$ . We vary the sign (positive and negative) and the intensity (low, medium and high) of the Pearson's correlation coefficient  $\rho$  between the true efficiency and the endogenous input to obtain the six endogenous scenarios  $\rho = [0.8; 0.4; 0.2; -0.2; -0.4; -0.8]$  and the exogenous one  $\rho = 0$ .

All scenarios were replicated using the Cobb-Douglas and the *Translog* production function, thus 14 scenarios were simulated assuming  $N = 100$ . For each dataset, we estimate the efficiency scores through an output oriented DEA. In particular we run the CCR-DEA (BCC-DEA) for the Cobb-Douglas (*translog*) production function and compute Pearson's correlation coefficient between the estimated and the real efficiency. Figures 1 and 2 show the results for each scenario, where the  $x$ -axis plots the estimated correlation coefficient values in each simulation and the  $y$ -axis counts the frequency of such correlations. Blue (red) bars denote negative (positive) correlations.

Firstly, we observe that the distribution of the estimated correlation coefficients  $\rho$  between the inputs and the estimated efficiency scores  $\hat{\theta}_i$  differ considerably in each scenario. Secondly, when the simulated correlation coefficient between the real efficiency and the input is negative, we find that almost all computed correlations are negative (blue bars in Figures 1 and 2). Thirdly, the frequency of the positively estimated correlation coefficients increases as long as the intensity of the simulated positive correlation grows. Finally, we identify only a few positive correlations compared with very frequent negative correlations in the exogenous scenario (top graph in Figures 1 and 2).

SCRIPT

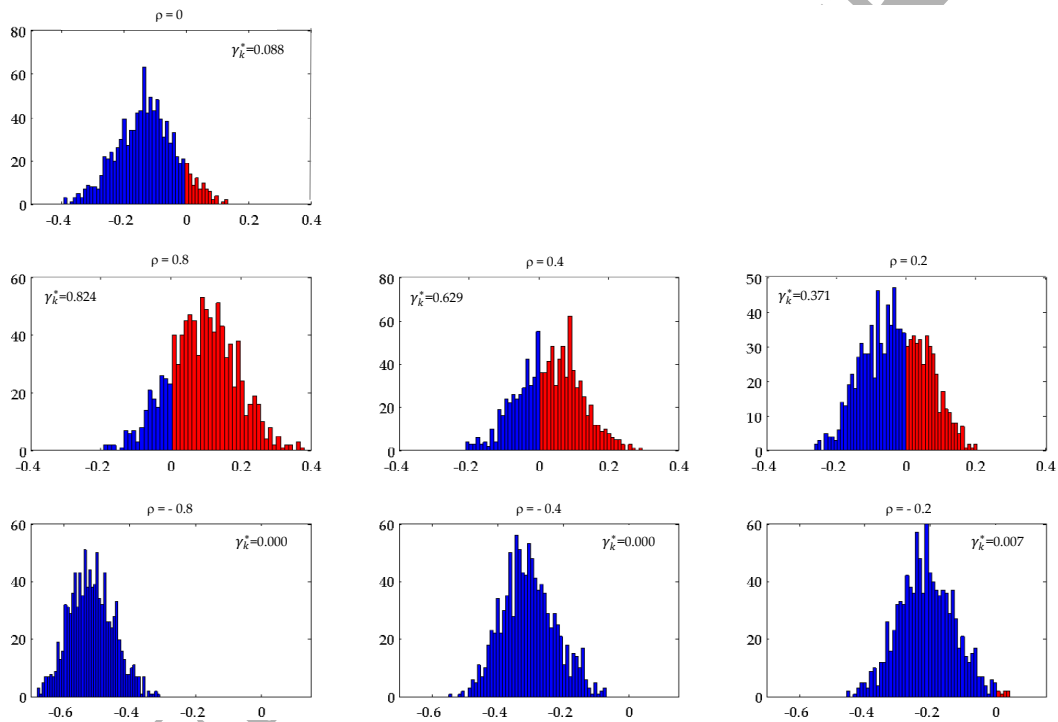
**Figure 1** Monte Carlo input classification using the proposed heuristic method to identify endogenous inputs (Cobb Douglas)



Note: Monte Carlo replicates  $B = 1,000$ . Sample size  $N = 100$ ,  $\rho$  indicates the simulated correlation between the input and the technical efficiency,  $\gamma_k^*$  is the indicator function. The CCR-DEA was estimated following an output orientation.

ACCEPTED

**Figure 2** Monte Carlo inputs classification using the proposed heuristic method to identify endogenous inputs (*Translog*)



Note: Monte Carlo replicates  $B=1,000$ , Sample size  $N=100$ ,  $\rho$  indicates the simulated correlation between the input and the technical efficiency,  $\gamma_k^*$  is the indicator function. The BCC-DEA was estimated following an output orientation.

Taking into account these Monte Carlo results, we build a heuristic method to classify each input type included in the DEA model based on the expected correlation coefficients. In practice, with empirical data we proceed in six steps as follows:

1. Randomly draw a bootstrap sample with replacement  $\chi^* = \{(X_{ib}^*, Y_{ib}^*) \mid i = 1, \dots, n\}$  from the empirical dataset  $\chi = \{(X_i, Y_i) \mid i = 1, \dots, n\}$ .

2. Compute the efficiency score using DEA. In this case we estimate  $0 \leq \hat{\theta}_{ib}^* = \frac{1}{\hat{\phi}_{ib}^*} \leq 1 \mid i = 1, \dots, n$  using DEA<sup>5</sup>

$$\varphi_i^* = \max_{\lambda, \varphi} \{ \varphi_i \mid \varphi y_i \leq Y \lambda; x_i \geq X \lambda; n1' \lambda = 1; \lambda \geq 0; \forall i = 1, \dots, n \} \quad (3)$$

3. For each input  $k = 1, \dots, p$ , compute the Pearson correlation coefficient between the estimated efficiency score  $\hat{\theta}_{ib}^*$  and the input  $x_{ik}^*$

$$\rho_{kb}^* = \text{corr}(x_{ik}^*, \hat{\theta}_{ib}^* \mid i = 1, \dots, n \ k = 1, \dots, p) \quad (4)$$

4. Repeat steps 1 to 3  $B$  times in order to obtain a bootstrap set of correlations  $\{\rho_{kb}^*, b = 1, \dots, B\}$  for each input  $k$ .

5. For each input  $x_{ik}^*$ , compute

$$\gamma_k^* = \frac{1}{B} \sum_{b=1}^B [I_{[0,1]}(\rho_{kb}^*)] \quad k = 1, \dots, p \quad (5)$$

where  $I_{[0,1]}(\rho^*)$  is an indicator function defined by

$$I_{[0,1]}(\rho^*) = \begin{cases} 1 & \text{if } 0 \leq \rho_k^* \leq 1; \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

6. Finally, classify the inputs according to the proposed heuristic as follows:

If  $\gamma_k^* < 0.25$ , input  $k$  is an *exogenous* or *negative endogenous* input.

<sup>5</sup> The DEA model in step 2 should be estimated following an input or output orientation and assuming constant (CCR-DEA model) or variable (BCC-DEA model) returns to scale according to the empirical problem. For illustration purposes we show the output-oriented BCC-DEA that is the model that will be run in the empirical application developed in Section 4.

If  $0.25 \leq \gamma_k^* < 0.5$ , input  $k$  is a *positive low endogenous* input.

If  $0.5 \leq \gamma_k^* < 0.75$ , input  $k$  is a *positive medium endogenous* input.

If  $0.75 \leq \gamma_k^*$ , input  $k$  is a *positive high endogenous* input.

Figures 1 and 2 also show the resulting  $\gamma_k^*$  after simulating the heuristic behavior. As we can observe  $\gamma_k^*$  is practically zero in all endogenous scenarios and slightly positive in the exogenous one. For positive endogeneity the most noticeable result from the Monte-Carlo experiment is that a higher frequency of positively estimated correlation coefficients between the real efficiency and the input (a higher  $\gamma_k^*$  value) indicates a higher intensity in the endogeneity. The cut-off points defined in step 6 to characterize the endogeneity are a *rule of thumb* orientation based on the quartiles to label the degree of endogeneity in empirical data<sup>6</sup>. The heuristic is explicitly designed to detect moderate and high positive endogeneity, since, according to Cordero et al. (2015), this problem can significantly bias DEA performance.

### 3. Strategy for improving DEA estimates

#### 3.1. Instrumental Input DEA

In order to improve DEA estimates when there is a positive and significant correlation between one input and true efficiency, we propose a semi-parametric strategy that introduces the well-known instrumental variable (IV) approach (see, e.g., Greene, 2003 or Wooldridge, 2012) into the conventional DEA model specification, which we call ‘instrumental input DEA’. This proposal shares the same intuitive idea as the IV strategy, namely, the DEA specification includes only the exogenous part of the endogenous input, that is, the part that is uncorrelated with technical efficiency.

To do this, we proceed as follows. Consider a single-output multi-input productive dataset

$\chi = \{(X, Y) \mid i = 1, \dots, n\}$ , where one input has a significant positive correlation with the true

---

<sup>6</sup> As a referee suggests, an alternative idea would be to define the heuristic modifying Equation 5 and only counting statistically different from zero correlation coefficients between the input and the estimated efficiency after each loop. This valuable extension of the analysis is left for future research.

efficiency term (hereafter the endogenous input  $X^e$ ). As in the classic IV approach, the first step is to find a good instrumental input  $G$  that must simultaneously satisfy two basic conditions:

- i) Relevance: the instrument  $G$  must be significantly correlated with the endogenous input  $X^e$ , *i.e.*  $E(X^e | G) \neq 0$ ;
- ii) Exogeneity: the instrument  $G$  must be uncorrelated with the true efficiency term, *i.e.*  $E(\theta | X^e) = 0$ .

The first condition can be checked in empirical applications by testing the significance of the parameter  $\tau$  in the estimated regression  $X^e = \alpha + \beta X + \tau G + \xi$ . If  $H_0 : \tau = 0$  is not rejected, 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 there being no causal relationship between the instrumental input  $G$  and the output variable  $Y$ . In other words,  $G$  should not have a partial effect on  $Y$  (beyond the effect mediated by the endogenous input), and  $G$  should also be uncorrelated with the omitted variables that affect  $Y$  (if the omission of relevant variables is the cause of the endogeneity). As Wooldridge says; “we must maintain this condition by appealing to economic behavior or introspection” (Wooldridge, 2012 p. 514). The II-DEA procedure is implemented according to two simple steps:

1. Regress the endogenous input ( $X^e$ ) over the instrumental input ( $G$ ) and the other exogenous inputs in order to isolate the exogenous component of the endogenous input that is uncorrelated with true efficiency:

$$X^e = \alpha + \beta_1 x_1 + \dots + \beta_{k-1} x_{k-1} + \tau G + \xi \quad (7)$$

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

2. Replace the endogenous input ( $X^e$ ) by the estimated exogenous variable  $X^e$  in the input matrix and run a standard DEA linear program (*e.g.* Eq. 3).

### 3.2. Monte Carlo simulations

#### 3.2.1 MC experimental design

In order to test the performance of the II-DEA strategy in finite samples, we use again synthetic data generated in a Monte Carlo experiment. We follow the same experimental design explained in Section 2 with the baseline scenario without endogeneity (exogenous scenario) and the six endogenous scenarios. In order to run the II-DEA we generate the instrumental input  $G$  also uniformly distributed between  $[5,50]$ , uncorrelated with the true efficiency level  $E(\theta|G) = 0$  and moderately correlated with the endogenous input  $E(X^e|G) \cong 0.25$ <sup>7</sup>. In Each Monte Carlo scenario we generate data and run the conventional DEA and the proposed II-DEA strategy to compare their performance.

All scenarios were replicated using the Cobb-Douglas and the *Translog* production function, thus 14 scenarios were simulated. For each dataset, we estimate the efficiency scores  $\hat{\theta}_i$  by running an output oriented conventional DEA and the II-DEA strategy in order to compare their performance<sup>8</sup>. To make the results more reliable, we undertake a Monte Carlo experiment where  $B$ , the number of replicates is 1,000<sup>9</sup>, consequently, all measures are computed in each replication and finally averaged to obtain the results below. Results from Monte Carlo simulations are reported in Tables 1 to 3 for a sample size of 100 DMUs in the 14 scenarios (the exogenous and the six endogenous under the Cobb-Douglas and the *translog* technology).

Finally, to examine the effect of sample size on the results, we also analyze the performance of the II-DEA strategy under different sample sizes but only for those scenarios where the performance of the standard DEA is significantly damaged and where it makes sense to correct for the presence of endogeneity (medium and high positive endogeneity). Results are presented in Tables 4 to 6 for a sample size of 50, 100 and 300 DMUs in 6 scenarios (the exogenous, medium and high positive endogeneity under the Cobb-Douglas and the *translog* technology).

<sup>7</sup> In Appendix A we describe the procedure to generate the endogenous input  $X^e$  and the instrumental input  $G$  with a specific desired Pearson's correlation coefficient between them and with the true efficiency  $\theta$ .

<sup>8</sup> As expected, CCR-DEA estimates outperform BCC-DEA for data generated with the Cobb-Douglas, where the elasticity of scale is constant, and *vice versa* for data derived from the *translog* technology where the elasticity of scale is variable. This result justifies that we run the DEA and II-DEA under constant returns to scale for the Cobb-Douglas and under variable returns to scale for the *Translog*. It is worth noting one more time the importance of making a correct choice of the assumed returns to scale before conducting an empirical DEA efficiency analysis.

<sup>9</sup> Simulations were carried out using MATLAB 7.9.0 software.

### 3.2.2 MC results

We compare the adequacy of the DEA and II-DEA in each scenario through two measures. Firstly, we test how accurately they rank observations by computing the Spearman's rho correlation coefficients between the true efficiency and estimated score pairs in each simulated scenario. The higher the correlation coefficient the better the technique identifies the true efficiency distribution. Secondly, we compute the Mean Absolute Deviation (MAD) by averaging the mean absolute deviations of the estimated scores from the true efficiency level after each replication  $MAD = \sum_{i=1}^N \frac{|\hat{\theta}_i - \theta_i|}{N}$ . A low MAD implies that, on average, the estimates are near to the true efficiency values, and hence small values are preferred. Table 1 summarizes and points out the following straightforward results.

We find that the II-DEA model clearly outperforms conventional DEA in the worst scenario, *i.e.* under a high and positive correlation between efficiency and one input ( $\rho = 0.8$ ), in terms of Spearman's rank correlation coefficient between true technical efficiency and DEA and II-DEA computed scores. However, while DEA tends to overestimate mean efficiency II-DEA tends to underestimate it. This result is reflected in terms of MAD. Under the Cobb-Douglas DGP the average estimated efficiency by DEA is closer to true efficiency than the estimation using the II-DEA approach and *vice versa* under the *translog* assumption. Secondly, we confirm the results reported by Cordero et al. (2015) with respect to conventional DEA being robust if the correlation is negative or moderately positive ( $\rho = 0.2$ ). Therefore, the instrumentation of endogenous inputs leads to considerably worse results in these cases. In the wake of these results, our discussion from now on is confined to the results for the scenarios where  $\rho = 0.4$  and  $\rho = 0.8$ .

Tables 2 and 3 show DEA and II-DEA ability to place DMUs in the appropriate quintile which allow us to identify where the main improvements of II-DEA strategy lie compared with DEA. The results show that the II-DEA outperforms DEA in the worst scenario, ( $\rho = 0.8$ ), because it is able to correctly identify the most inefficient units. Note that, in this scenario, DMUs with a low input level are the most inefficient. Therefore, this finding confirms that the proposed method can deal with the misidentification of the true frontier in this region. As the DEA estimates are relative measures, this improvement also affects the ability of this technique to correctly identify the most efficient units. In this regard, 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% (4.5%) assuming VRS (CRS) to almost zero.

Under the assumption of a moderate correlation between the true efficiency and one input ( $\rho = 0.4$ ), both the DEA and II-DEA techniques show similar results. This implies that the proposed method is not powerful enough to reduce 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  $X_i^e$  does not seem to provide enough information to correctly identify the true frontier at low input levels.

**Table 1** Accuracy measures for conventional DEA and II-DEA estimates in Monte Carlo simulations (Cobb-Douglas and *translog*)

	Cobb-Douglas DGP						<i>Translog</i> DGP					
	Spearman correlation		Estimated efficiency		MAD		Spearman correlation		Estimated efficiency		MAD	
	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA
$\rho \cong -0.8$	0.695	-0.270	0.898	0.849	0.058	0.136	0.708	0.128	0.957	0.893	0.097	0.127
$\rho \cong -0.4$	0.774	0.146	0.887	0.794	0.049	0.133	0.765	0.362	0.895	0.846	0.073	0.109
$\rho \cong -0.2$	0.778	0.289	0.885	0.777	0.048	0.134	0.757	0.439	0.893	0.831	0.071	0.105
$\rho \cong 0$	0.778	---	0.884	---	0.049	---	0.729	---	0.893	---	0.072	---
$\rho \cong 0.2$	0.754	0.586	0.886	0.750	0.051	0.138	0.675	0.605	0.898	0.810	0.078	0.100
$\rho \cong 0.4$	0.715	0.693	0.890	0.744	0.055	0.139	0.612	0.657	0.905	0.804	0.085	0.099
$\rho \cong 0.8$	0.520	0.881	0.911	0.732	0.073	0.141	0.342	0.760	0.936	0.794	0.116	0.097

*Note:* Mean values after 1,000 replications. Sample size  $n = 100$ . In the Cobb-Douglas DGP scenarios DEA and II-DEA were estimated under CRS and in the *translog* DGP scenarios DEA and II-DEA were estimated under VRS.

Finally, Tables 4, 5 and 6 show results from the Monte Carlo simulations under different samples sizes. First, Table 4 confirms, in line with previous research (Orme and Smith, 1996; Zhang and Bartels, 1998, Pedraja-Chaparro et al., 1999, Perelman and Santín, 2009, Krüger, 2012), that as long as sample size increases, the Spearman's correlation coefficient between the

estimated efficiency and the true efficiency increases and the estimated mean technical efficiency decreases in all scenarios. Second, the same conclusions discussed in the analysis with a sample size of 100 arise in terms of the Spearman's correlation and the MAE assuming ( $\rho = 0.8$ ). In this case the II-DEA model outperforms by far conventional DEA in terms of the Spearman's correlation coefficient but at the cost of a higher MAE due to underestimating efficiency values.

**Table 2** Monte Carlo results for the ability of conventional DEA and II-DEA to correctly assign observations to quintiles (Cobb-Douglas)

	% Assigned to the correct quintile		% Assigned to a quintile two or more quintiles away from actual quintile		% Assigned to bottom quintile actually in bottom quintile		% Assigned to top quintile actually in top quintile		% Assigned to bottom quintile actually in the top two quintiles		% Assigned to top quintile actually in the bottom two quintiles	
	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA
$\rho \cong -0.8$	43.5	14.7	16.9	56.3	68.3	12.3	47.8	7.1	1.2	72.2	12.2	42.2
$\rho \cong -0.4$	48.5	21.9	12.1	43.6	74.3	25.5	52.6	25.2	0.4	41.1	7.3	22.2
$\rho \cong -0.2$	48.8	25.2	11.8	38.1	74.9	32.2	52.3	32.0	0.4	31.0	7.0	16.7
$\rho \cong 0$	48.5	---	11.9	---	74.9	---	52.1	---	0.3	---	7.0	---
$\rho \cong 0.2$	46.6	35.6	13.3	24.6	72.4	50.4	49.9	48.1	0.5	11.5	8.5	7.8
$\rho \cong 0.4$	44.1	41.1	15.7	18.4	69.0	58.2	48.0	54.7	0.8	6.2	10.6	5.0
$\rho \cong 0.8$	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  $n = 100$ . DEA and II-DEA were estimated under CRS.

**Table 3** Monte Carlo results for the ability of conventional DEA and II-DEA to correctly assign observations to quintiles (*translog*)

	% Assigned to the correct quintile		% Assigned to a quintile two or more quintiles away from actual quintile		% Assigned to bottom quintile actually in bottom quintile		% Assigned to top quintile actually in top quintile		% Assigned to bottom quintile actually in the top two quintiles		% Assigned to top quintile actually in the bottom two 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  $n = 100$ . DEA and II-DEA were estimated under VRS.

Third, under a moderate endogeneity ( $\rho = 0.4$ ), the standard DEA method increasingly dominates the II-DEA as the sample size increases. II-DEA only performs slightly better than conventional DEA assuming the translog and in terms of Spearman's correlation coefficient. However, we think that this result does not provide enough evidence to instrument the endogenous input under a moderate endogeneity. Finally, Tables 5 and 6 show that the outperformance achieved by II-DEA over DEA under an intense positive endogeneity ( $\rho = 0.8$ ) stems from the higher ability to correctly rank DMUs in their correct quintile.

From these results, we derive two conclusions. Firstly, it would appear that under a moderate endogeneity level in one input ( $\rho = 0.4$ ) II-DEA does not provide enough evidence to substitute the use of a conventional DEA. Second, we show that II-DEA is able to improve standard DEA estimations for building a performance ranking of DMUs under a high level of endogeneity ( $\rho = 0.8$ ).

**Table 4** Accuracy measures for conventional DEA and II-DEA estimates in Monte Carlo simulations by sample size (Cobb-Douglas and *translog*)

	Cobb-Douglas DGP						Translog DGP					
	Spearman correlation		Estimated efficiency		MAD		Spearman correlation		Estimated efficiency		MAD	
	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA	DEA	II-DEA
$\rho \cong 0$												
$n = 50$	0.718	---	0.906	---	0.060	---	0.612	---	0.934	---	0.080	---
$n = 100$	0.779	---	0.884	---	0.048	---	0.726	---	0.893	---	0.073	---
$n = 300$	0.832	---	0.854	---	0.042	---	0.765	---	0.882	---	0.048	---
$\rho \cong 0.4$												
$n = 50$	0.656	0.661	0.913	0.776	0.067	0.124	0.499	0.560	0.943	0.854	0.090	0.097
$n = 100$	0.715	0.693	0.890	0.744	0.055	0.139	0.612	0.657	0.905	0.804	0.085	0.099
$n = 300$	0.776	0.719	0.859	0.703	0.047	0.144	0.656	0.697	0.890	0.781	0.057	0.109

$\rho \cong 0.8$ 

n = 50	0.467	0.849	0.931	0.763	0.084	0.123	0.241	0.665	0.963	0.843	0.109	0.095
n = 100	0.520	0.881	0.911	0.732	0.073	0.141	0.342	0.760	0.936	0.794	0.116	0.097
n = 300	0.579	0.900	0.883	0.706	0.063	0.148	0.362	0.827	0.919	0.772	0.082	0.101

Note: Mean values after 1,000 replications. In the Cob-Douglas DGP scenarios DEA and II-DEA were estimated under CRS and in the *translog* DGP scenarios DEA and II-DEA were estimated under VRS.

ACCEPTED MANUSCRIPT

**Table 5** Monte Carlo results for the ability of conventional DEA and II-DEA to correctly assign observations to quintiles by sample size  
(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
<b><math>\rho \cong 0</math></b>												
n = 50	44.2	---	15.5	---	71.4	---	45.4	---	0.7	---	10.7	---
n = 100	48.5	---	11.9	---	74.9	---	52.1	---	0.3	---	7.0	---
n = 300	52.7	---	8.5	---	77.6	---	58.1	---	0.2	---	3.5	---
<b><math>\rho \cong 0.4</math></b>												
n = 50	40.2	39.5	19.2	19.9	65.3	57.3	42.1	51.1	1.3	6.8	14.0	6.8
n = 100	44.1	41.1	15.7	18.4	69.0	58.2	48.0	54.7	0.8	6.2	10.6	5.0
n = 300	48.6	41.7	12.0	17.0	73.0	59.0	54.4	57.9	0.3	5.0	6.9	2.9
<b><math>\rho \cong 0.8</math></b>												
n = 50	31.7	53.3	28.9	7.2	47.5	76.1	36.6	62.0	5.6	0.3	24.1	2.4
n = 100	34.4	57.0	26.2	5.1	50.1	77.8	41.4	67.6	4.5	0.1	21.6	1.2
n = 300	36.8	59.0	23.0	3.8	53.4	78.7	45.1	70.7	3.0	0.1	18.5	0.4

Note: Mean values after 1,000 replications. DEA and II-DEA were estimated under CRS

**Table 6** Monte Carlo results for the ability of conventional DEA and II-DEA to correctly assign observations to quintiles by sample size (*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
<b><math>\rho \cong 0</math></b>												
n = 50	39.3	---	21.6	---	66.4	---	38.6	---	1.5	---	15.6	---
n = 100	47.2	---	13.3	---	74.8	---	46.9	---	0.2	---	12.3	---
n = 300	49.0	---	11.4	---	75.0	---	50.9	---	0.4	---	10.7	---
<b><math>\rho \cong 0.4</math></b>												
n = 50	34.2	37.6	27.9	23.1	55.5	57.0	36.1	39.8	3.6	6.7	20.3	19.5
n = 100	40.1	43.1	19.8	17.1	64.8	61.6	42.1	46.5	0.7	4.0	18.6	16.8
n = 300	42.9	43.9	16.8	15.8	68.0	62.6	43.3	47.8	0.7	3.8	17.4	9.8
<b><math>\rho \cong 0.8</math></b>												
n = 50	25.0	43.8	40.9	16.1	33.0	69.2	30.0	43.4	15.3	0.4	29.3	18.1
n = 100	28.0	51.8	34.8	10.0	40.8	75.7	32.8	46.9	8.2	0.1	30.3	15.6
n = 300	29.5	56.0	32.9	6.7	42.9	77.4	31.6	48.1	9.1	0.1	32.8	9.5

Note: Mean values after 1,000 replications. DEA and II-DEA were estimated under VRS.

## 4. Empirical application

To illustrate how the heuristic and the II-DEA can be used empirically and to size up the impact of not considering the endogeneity problem in a real setting, we perform an empirical analysis of the education sector where the endogeneity problem is frequently observed due to school self-selection (Webbink, 2005).

### 4.1. Data

We focus on the educational efficiency of public secondary schools in Uruguay, where 88% of high school students attend public institutions (Education Observatory, National Administration of Public Education). Historically, Uruguay has occupied a leading position in Latin America in terms of educational achievement, according to the main standard indicators and international studies. However, it is also one of the most unequal countries in terms of student achievement and socioeconomic background. This leads us to suspect that there is endogeneity related to the school socioeconomic context. In other words, not only do schools in more disadvantaged settings have pupils with a poorer socioeconomic status and poorer educational resources, but they are also significantly less efficient than schools in more advantaged contexts.

We analyze 71 Uruguayan public schools included in the Programme for International Student Assessment (PISA) 2012 database (OECD, 2013). As PISA 2012 focuses on mathematics, we select the schools' average result in mathematics (Maths) as the output of the educational process. Regarding educational inputs, we select three variables to represent the classical inputs required to carry out the learning process (for an extensive revision see Worthington, 2001 and De Witte and López-Torres, 2015). Firstly, we use the quality of material educational resources at the school (SCMATEDU), an index built from the school principal's responses to seven questions related to the scarcity or lack of laboratory equipment, institutional materials, computers, Internet, educational software, library materials and, finally, audiovisual resources, in order to capture the school's physical capital. The higher the index, the better the quality of the school's material. Secondly, we include the proportion of fully certified teachers (PROPCERT). This reflects the quality of teachers, and therefore the school's human capital. The index is constructed as the ratio of the total number of certified teachers (with a teaching degree) to the total number of teachers in the school. Finally, we include the school's socioeconomic context, computed as the average students' index of economic, social and cultural status (ESCS) developed by PISA analysts to indicate students' socioeconomic status (for further details see OECD (2014)).

Table 5 reports descriptive statistics for all selected variables mentioned above, including the instrumental input that we introduce below in Section 4.2. DEA-VRS was estimated following an output orientation according to Equation 1.

## 4.2. Results

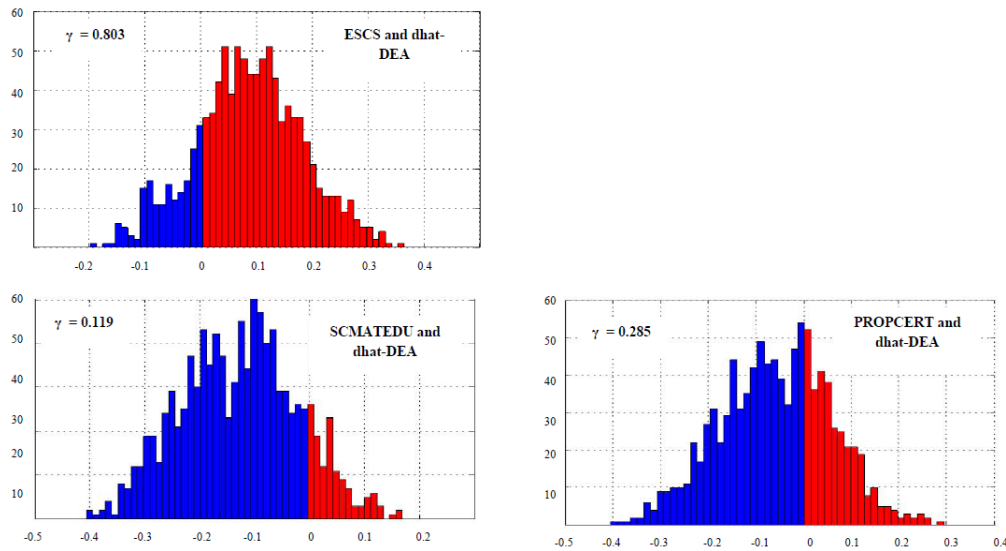
We use the heuristic explained in Section 2 to classify the three inputs (Figure 3). SCMATEDU is classified as exogenous or negatively correlated with the true efficiency term, PROPCERT appears to have a low positive correlation with true efficiency and, as we suspected, the school socioeconomic level (ESCS) appears to be a highly positive endogenous input. As a result, we decide to instrument this variable and correct our estimations.

**Table 7** Descriptive statistics of output, inputs and explanatory variables of efficiency in the empirical application to education sector

Variable	Mean	Std. Dev.	Min	Max
<i>Output</i>				
Maths	382.7	44.2	270.9	466.5
<i>Inputs</i>				
ESCS	2.20	0.42	1.35	3.29
SCMATEDU	4.50	1.11	2.30	6.57
PROPCERT	0.52	0.20	0.15	0.94
<i>Instrumental Input</i>				
ACCINT	0.86	0.08	0.60	1.00

Source: Own elaboration from PISA 2012 data.

**Figure 3** Input classification using the proposed heuristic to identify endogenous inputs using conventional DEA



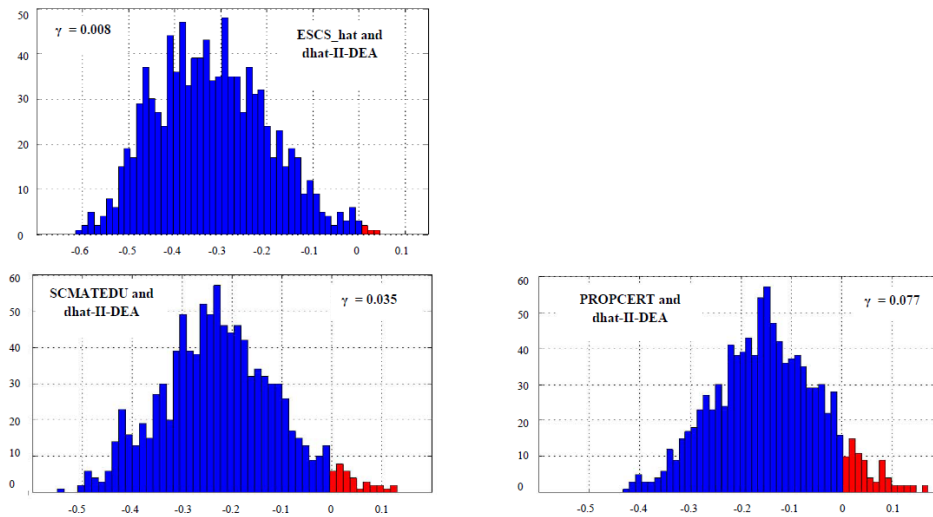
Note: Bootstrap replicates  $B = 2,000$ . Sample size  $N=71$ ,  $\rho$  indicates the simulated correlation between the input and the technical efficiency,  $\gamma_k^*$  is the indicator function. The CCR-DEA was estimated following an output orientation.

The first step for applying the II-DEA is to find a good instrument. As mentioned in Section 3, a good instrument should be correlated with the endogenous input (ESCS) but uncorrelated with true efficiency. In empirical applications, this means that there should be no clear causal relationship between the instrumental

input variable and the *Maths* output variable. Accordingly, we find an instrumental input that meets both conditions: the “percentage of students at the school that have had access to Internet before the age of thirteen” (ACCINT hereafter). The correlation between this variable and the school socioeconomic levels is 0.2 (similar to the correlation assumed in the Monte Carlo experiment performed in Section 3). After running Equation 7, we cannot reject the null hypothesis  $H_0 : \tau = 0$  concluding that the proposed instrument is significant at the 1% level. Furthermore, there is no clear evidence in the literature as to access to the Internet or ICTs alone leading to better academic results. The final effect will depend on how they are used and on parental monitoring and supervision (Angrist and Lavy 2002, Goolsbee and Guryan 2006).

Based on the above results, we enact the II-DEA procedure described in Section 3, first calculating  $\widehat{ESCS}$  through Equation 5. Again, we use the heuristic to classify and verify whether endogeneity persists and now find that all inputs are classified as being exogenous or having a negative correlation with the efficiency term (Figure 4).

**Figure 4** Input classification using the proposed heuristic to identify endogenous inputs using II-DEA



Note: Bootstrap replicates  $B = 2,000$ . Sample size  $N=71$ .  $\rho$  indicates the simulated correlation between the input and the technical efficiency,  $\gamma_k$  is the indicator function. The CCR-DEA was estimated following an output orientation.

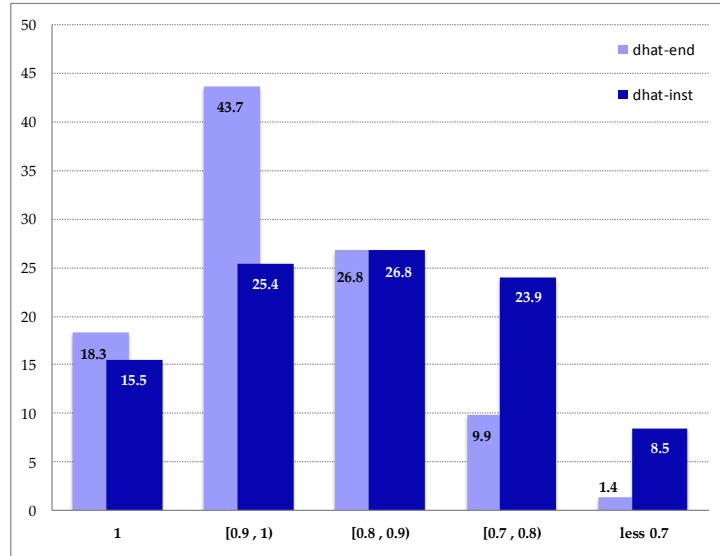
Table 6 and Figure 5 illustrate the distribution of efficiency scores  $0 \leq \hat{\theta}_i \leq 1$ , estimated by the output-oriented standard DEA and the II-DEA models under VRS (Eq.1). Figure 5 shows that when we control by endogeneity, only 15.5% of schools behave efficiently, and, on average, educational results could be improved by 17% given the available resources. We find that a quarter of the evaluated schools could increase academic achievements by up to 11% if they were fully efficient, and 26.8% of the schools could raise their educational results between 11% and 25%. Moreover, almost one third of the evaluated schools could improve their outcomes by over 25% to reach the frontier.

**Table 8** Descriptive statistics of the estimated efficiency scores using DEA and II-DEA

Efficiency	Mean	Std- Dev.	Min.	Max.	Q1	Q2	Q3
dhat-end	0.916	0.077	0.681	1.000	0.863	0.931	0.985
dhat-inst	0.869	0.102	0.610	1.000	0.795	0.879	0.963

Source: Author's estimates using PISA 2012 data

Notes: Efficiency scores are equal or greater than one. Q1, Q2 and Q3 are the percentile 0.25; 0.50 and 0.75 of the estimated efficiency distribution.

**Figure 5** Estimated efficiency scores distribution using standard DEA and II-DEA

Source: Own estimates using PISA 2012 data.

Note: dhat-end are the estimated efficiency scores using the standard DEA and dhat-inst are the estimated efficiency scores using the II-DEA method. Values equal to one represent fully efficient units. Higher values of the score imply more inefficiency.

#### 4.3. Standard DEA vs II-DEA results

For the purposes of comparison, we also ran the conventional DEA model under the endogenous scenario (*i.e.* using the ESCS as an input). We observe that, in this case, not only is the average efficiency overestimated, but also the distribution of all estimated efficiency scores is shifted to the left (Table 8 and Figure 5). These findings are consistent with the results of the Monte Carlo experiment reported in Section 3. Therefore, potential improvements in public school educational outcomes are considerably underestimated if endogeneity is not accounted for.

Table 9 provides the mean ESCS level and the mean estimated efficiency scores for both models (DEA and II-DEA) by quintiles according to the ESCS endogenous input level. Thanks to this complementary analysis, we can evaluate the differences between both specifications at different points of the distribution. This is, therefore, a helpful tool for locating the main damage done by endogeneity in this empirical

application. The last column of Table 9 reports the mean bias computed as the average of the absolute difference between the efficiency score estimated by the II-DEA model and the standard DEA, respectively:

$$bias_i = \left| \hat{\theta}_{i,II-DEA} - \hat{\theta}_{i,DEA} \right| \quad (6)$$

We confirm in our application that endogeneity has a major effect on the efficiency measure of schools in a more disadvantaged socioeconomic context. Schools in the bottom quintile by socioeconomic context show the greatest bias (0.142), whereas bias is not at all significant (0.003) for schools positioned in the top quintile. In other words, if our estimations take into account the endogeneity issue, schools from the bottom quintile could improve their results by on average 26.6%, whereas, if this issue is not taken into account, the potential improvement is no more than 7%. For schools in the top quintile (from most advantaged contexts), the potential improvements are 7% in both scenarios. We also observe that the more advantaged the average socioeconomic school context, the greater the average school efficiency estimated by II-DEA is.

**Table 9** Mean ESCS, DEA and II-DEA estimated efficiency scores and bias by quintiles (according the school's ESCS)

	Mean ESCS	Mean dhat-inst	Mean dhat-end	Mean  Bias
<b>Quintiles by ESCS</b>				
Bottom quintile	1.68	0.93	0.79	0.142
4th quintile	1.92	0.89	0.82	0.068
3rd quintile	2.13	0.91	0.88	0.029
2nd quintile	2.40	0.91	0.91	0.002
Top quintile	2.82	0.93	0.93	0.003

Source: Author's estimates using PISA 2012 data

## 5. Conclusions

Although endogeneity is currently one of the biggest concerns in microeconomics, it has not so far received much attention in the DEA literature. Recently, Cordero et al. (2015) was the bearer of good and bad news for DEA practitioners: the technique appears to be robust to the presence of negative and slightly positive endogeneity, but a significant positive correlation between one input and the true efficiency level severely biases DEA performance. Unfortunately, as such a positive correlation is frequently observed in different production contexts, the key issue is to deal with this problem in empirical applications.

In this paper, we propose a simple statistical heuristic procedure that will allow practitioners to identify potential endogenous inputs when they suspect that there might be some significant positive endogeneity in

their setting. In addition, we propose the instrumental input DEA strategy as a potential solution for dealing with this problem and improving DEA estimations. Monte Carlo simulations show that the II-DEA strategy outperforms standard DEA if there is a high positive correlation, further driven by its ability to correctly identify the most inefficient DMUs. Its application to the education sector reveals that, if endogeneity is not accounted for, potential efficiency improvements could be seriously underestimated due to the misidentification of the most inefficient DMUs as benchmarks. This leads to inaccurate performance-based recommendations.

In conclusion, the potential existence of endogeneity can no longer be ignored in DEA estimations. Presumably, new theoretical and empirical contributions will be developed to address this issue in the near future. In this vein, this paper should be construed as an initial step to encourage academics and practitioners to continue contributing to this line of research, since more research is still needed in different directions. However, this exceeds the scope of this paper. Although the experimental Monte Carlo design tries to replicate a general production setting and is in line with several previous studies, results must be interpreted cautiously as they depend on the parameters and assumed functional forms and cannot be directly generalized to all contexts. In this regard, we think that the derivation of the asymptotic properties of the proposed II-DEA estimator and research about how other non-parametric DEA-rooted efficiency techniques (free disposal hull, order-m, order-alpha, total factor productivity indexes based on DEA, STONED, conditional efficiency models) can be affected by endogeneity are two of the potentially most promising future contributions with respect to this issue.

## References

- Angrist, J., & Lavy, V. (2002). New Evidence on Classroom Computers and Pupil Learning\*. *The Economic Journal*, 112(482), 735-765.
- Banker, R. D., Charnes, A., & 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., & Morey, R. C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research*, 34(4), 513-521.
- Bifulco, R., & Bretschneider, S. (2001). Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review*, 20(5), 417-429.
- Bifulco, R., & Bretschneider, S. (2003). Response to comment on estimating school efficiency. *Economics of Education Review*, 22(6), 635-638.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444.
- Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA)—Thirty years on. *European Journal of Operational Research*, 192(1), 1-17.

- Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, 44, 1-4.
- Co-operation, O. f. E., & Development. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*.
- Co-operation, O. f. E., & Development. (2014). *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised Edition, February 2014): Student Performance in Mathematics, Reading and Science*: OECD Publishing.
- Cordero, J. M., Santín, D., & Sicilia, G. (2015). Testing the accuracy of DEA estimates under endogeneity through a Monte Carlo simulation. *European Journal of Operational Research*, 244(2), 511-518.
- Cordero-Ferrera, J. M., Pedraja-Chaparro, F., & Salinas-Jiménez, J. (2008). Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs. *Applied Economics*, 40(10), 1323-1339.
- De Witte, K., & López-Torres, L. (2015). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*.
- Emrouznejad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-economic planning sciences*, 42(3), 151-157.
- Gattoufi, S., Oral, M., Kumar, A., & Reisman, A. (2004a). 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.
- Gattoufi, S., Oral, M., & Reisman, A. (2004b). Data envelopment analysis literature: A bibliography update (1951–2001). *Journal of Socio-Economic Planning Sciences*, 38(2-3), 159-229.
- Gong, B.-H., & 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.
- Goolsbee, A., & Guryan, J. (2006). The impact of Internet subsidies in public schools. *The Review of Economics and Statistics*, 88(2), 336-347.
- Greene, W. (2003). *Econometric Analysis*. New Jersey, Prentice Hall.
- Gstach, D. (1998). Another approach to data envelopment analysis in noisy environments: DEA+. *Journal of Productivity Analysis*, 9(2), 161-176.
- Hatami-Marbini, A., Emrouznejad, A., & Tavana, M. (2011). A taxonomy and review of the fuzzy data envelopment analysis literature: two decades in the making. *European Journal of Operational Research*, 214(3), 457-472.
- Jianfeng, M. (2015). A two-stage DEA model considering shared inputs and free intermediate measures. *Expert Systems with Applications*, 42(9), 4339-4347.
- 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., & Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. *Operations Research*, 58(1), 149-160.

- Liu, J. S., Lu, L. Y., Lu, W.-M., & Lin, B. J. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, *41*(1), 3-15.
- Liu, S.-T. (2011). A fuzzy modeling for fuzzy portfolio optimization. *Expert Systems with Applications*, *38*(11), 13803-13809.
- Lotfi, F. H., Noora, A. A., Jahanshahloo, G. R., & Reshadi, M. (2011). One DEA ranking method based on applying aggregate units. *Expert Systems with Applications*, *38*(10), 13468-13471.
- Mayston, D. (2003). Measuring and managing educational performance. *Journal of the Operational Research Society*, *54*(7), 679-691.
- Mayston, D. J. (2016). Data envelopment analysis, endogeneity and the quality frontier for public services. *Annals of Operations Research*, 1-19.
- Muñiz, M., Paradi, J., Ruggiero, J., & Yang, Z. (2006). Evaluating alternative DEA models used to control for non-discretionary inputs. *Computers & operations research*, *33*(5), 1173-1183.
- Nataraja, N. R., & 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., & Smith, P. (1996). The potential for endogeneity bias in data envelopment analysis. *Journal of the Operational Research Society*, 73-83.
- Pedraja-Chaparro, F., Salinas-Jiménez, J., & Smith, P. (1999). On the quality of the data envelopment analysis model. *Journal of the Operational Research Society*, *50*(6), 636-644.
- Perelman, S., & Santín, D. (2009). How to generate regularly behaved production data? A Monte Carlo experimentation on DEA scale efficiency measurement. *European Journal of Operational Research*, *199*(1), 303-310.
- Peyrache, A., & Coelli, T. (2009). Testing procedures for detection of linear dependencies in efficiency models. *European Journal of Operational Research*, *198*(2), 647-654.
- Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, *90*(3), 553-565.
- 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. (2006). Measurement error, education production and data envelopment analysis. *Economics of Education Review*, *25*(3), 327-333.
- Schlotter, M., Schwerdt, G., & Woessmann, L. (2011). Econometric methods for causal evaluation of education policies and practices: a non-technical guide. *Education Economics*, *19*(2), 109-137.
- Seiford, L. M. (1997). A bibliography for data envelopment analysis (1978-1996). *Annals of Operations Research*, *73*, 393-438.
- 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.

- Smith, P. (1997). Model misspecification in data envelopment analysis. *Annals of Operations Research*, 73, 233-252.
- Wang, Y.-M., & Chin, K.-S. (2010). Some alternative DEA models for two-stage process. *Expert Systems with Applications*, 37(12), 8799-8808.
- 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.
- Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics*, 9(3), 245-268.
- Yu, Y., & Shi, Q. (2014). Two-stage DEA model with additional input in the second stage and part of intermediate products as final output. *Expert Systems with Applications*, 41(15), 6570-6574.
- Zhang, Y., & Bartels, R. (1998). The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. *Journal of Productivity Analysis*, 9(3), 187-204.

## Appendix A

In this appendix we describe the procedure used to compute the endogenous input  $X^e$  and the instrumental input  $G$  in order to obtain both variables with the same distribution as the exogenous inputs, but correlated between them by  $\rho_{G,X^e} = 0.25$ , and with the true technical efficiency  $\theta$  by  $\rho_{G,\theta} = 0$  and with a specific value of  $\rho_{X^e,\theta}$  in each of the six endogenous scenario. In each case we follow the next procedure:

1. Select the desired Pearson's correlation coefficient  $\rho_{X^e,\theta}$  between the endogenous input  $X^e$  and the true technical efficiency  $\theta$ .
2. Draw a random matrix  $A_{N \times 3} = (a_1, a_2, a_3)$  from a multivariate normal distribution

$$N(0; \Sigma), \text{ where } \Sigma = \begin{bmatrix} 1 & \rho_{X^e,\theta} & 0.25 \\ \rho_{X^e,\theta} & 1 & 0 \\ 0.25 & 0 & 1 \end{bmatrix}.$$

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):  $B' = [ID_{a_1} \ a_1]$ .
5. Generate an independent vector  $x_{(N \times 1)}$  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  $x_s$ :  $C = [ID_{a_1} \ a_1 \ x_s]$ .
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,  $X^e = x_{ID}$ .
9. Match ID with the vector  $a_2$  obtaining:  $D = [ID \ a_2]$ . Sort D by  $a_2$  in an ascending order (the ID variable will be unsorted):  $D' = [ID_{a_2} \ a_2]$ .
10. Randomly and independently generate  $N$  values of  $u_i$  using a half-normal distribution  $u_i \approx |N(0; 0.25)|$ . Then, compute the vector  $e_i = \exp(-u_i)$  and sort this variable in an ascending order obtaining  $e_s$ .
11. Compute a new H matrix by merging  $D'$  with  $e_s$ :  $H = [ID_{a_2} \ a_2 \ e_s]$ .
12. Sort H by the ID variable in an ascending order:  $H' = [ID \ a_{2\_ID} \ e_{ID}]$ .
13. The latter vector of  $H'$  ( $e_{ID}$ ), will be defined as the true technical efficiency level  $\theta = e_{ID}$ . 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. Match the ID with the vector  $a_3$  obtaining:  $K = [ID \ a_3]$ . Sort  $K$  by  $a_3$  in an ascending order (the ID variable will be unsorted):  $K' = [ID_{a_3} \ a_3]$ .
15. Generate an independent vector  $r_{(N \times 1)}$  from a uniform distribution over the interval [5, 50] and sort it in an ascending order obtaining  $r_s$ .
16. Compute a new  $L$  matrix by merging  $K'$  with  $r_s$ :  $L = [ID_{a_3} \ a_3 \ r_s]$ .
17. Sort  $L$  by the ID variable in an ascending order:  $L' = [ID \ a_{3-ID} \ r_{ID}]$ .
18. The latter vector of  $L'$  ( $r_{ID}$ ), will be defined as the instrumental input,  $G = r_{ID}$ .

We vary the sign (positive and negative) and the intensity (high, medium or weak) of the  $\rho_{X^e, \theta}$  in order to obtain the six endogenous scenarios. Table A.1 summarizes the main descriptive statistics of the correlation coefficients that have been actually obtained in each scenario.

**Table A.1** Descriptive statistics of the correlation coefficients between the endogenous input, technical efficiency and the instrumental input after the Monte Carlo experiments

	$\rho \cong -0.8$	$\rho \cong -0.4$	$\rho \cong -0.2$	$\rho \cong 0.2$	$\rho \cong 0.4$	$\rho \cong 0.8$
<u>Cobb-Douglas technology</u>						
Mean	<b>-0.809</b>	<b>-0.422</b>	<b>-0.235</b>	<b>0.231</b>	<b>0.426</b>	<b>0.811</b>
Std. deviation	0.033	0.084	0.096	0.095	0.083	0.036
Mean	<b>0.002</b>	<b>0.004</b>	<b>-0.004</b>	<b>0.001</b>	<b>-0.002</b>	<b>0.002</b>
Std. deviation	0.105	0.104	0.100	0.105	0.102	0.099
Mean	<b>0.233</b>	<b>0.234</b>	<b>0.239</b>	<b>0.241</b>	<b>0.231</b>	<b>0.234</b>
Std. deviation	0.095	0.097	0.094	0.095	0.097	0.095
<u>Translog technology</u>						
Mean	<b>-0.813</b>	<b>-0.425</b>	<b>-0.230</b>	<b>0.239</b>	<b>0.427</b>	<b>0.812</b>
Std. deviation	0.033	0.085	0.096	0.097	0.081	0.034
Mean	<b>0.000</b>	<b>0.002</b>	<b>-0.001</b>	<b>0.000</b>	<b>-0.005</b>	<b>-0.004</b>
Std. deviation	0.100	0.100	0.103	0.100	0.101	0.100
Mean	<b>0.233</b>	<b>0.233</b>	<b>0.235</b>	<b>0.235</b>	<b>0.237</b>	<b>0.231</b>
Std. deviation	0.097	0.099	0.100	0.097	0.091	0.097

Note: Mean values after 1,000 replications. Sample size  $n = 100$ .