Dealing with the Endogeneity Problem in Data Envelopment Analysis

José Manuel Cordero and Daniel Santín and Gabriela Sicilia

University of Extremadura, Complutense University of Madrid

April 2013

Dealing with the Endogeneity Problem in Data Envelopment Analysis

José Manuel Cordero

University of Extremadura, Spain

Daniel Santín

Gabriela Sicilia[1]

Complutense University of Madrid, Spain

April, 2013

Abstract

Endogeneity, and the distortions on the estimation of economic models that it causes, is a familiar problem in the econometrics literature. Although non-parametric methods like data envelopment analysis (DEA) are among the most used techniques for measuring technical efficiency, the effects of endogeneity on such efficiency estimates have received little attention. The aim of this paper is twofold. First, we further illustrate the endogeneity problem and its causes in production processes like the correlation between one input and the efficiency level. Second, we use synthetic data generated in a Monte Carlo experiment to analyze how different levels of positive and negative endogeneity can impair DEA estimations. We conclude that although DEA is robust to negative endogeneity, a high positive endogeneity level, *i.e.*, a high positive correlation between one input and the true efficiency level, significantly and severely biases DEA performance.

Keywords: Technical efficiency, DEA, Endogeneity, Monte Carlo.

---

[1] Corresponding author.
E-mail addresses:
José Manuel Cordero: jmcordero@unex.es
Daniel Santín: dsantin@ccee.ucm.es
Gabriela Sicilia: gabriels@ucm.es

1. Introduction

The evaluation of the technical efficiency of decision-making units (DMUs) is a crucial component of management decision-making for saving resources, monitoring DMU activity in order to detect better and worse performers and improving results. Public service providers are naturally interested in efficiency assessments since they face both increasing demands and financial constraints. Completely unknown production technologies or the frequent use of multiple proxy variables to approximate the real output as a result of the special characteristics of public sector production, which is not fixed upon profit maximization, complicate the estimation of accurate efficiency measures (Bowlin, 1986). In these contexts, non-parametric techniques, and especially data envelopment analysis (DEA), are the most commonly applied methods for measuring technical efficiency. This technique does not assume a particular functional form for the underlying production technology or the distribution of inefficiency and also it can easily handle multiple inputs and outputs providing very useful information about benchmark units.

Within this framework, one of the key microeconomic assumptions for estimating a production frontier is monotonicity. This assumption implies that the output will be constant or increase, but never decrease, when more units of an input are added. In practice, this assumption implies a straightforward positive causal relationship from inputs to outputs as long as efficiency is exogenously distributed across DMUs. Selecting the appropriate input and output variables to include in the model is the most critical decision that practitioners will have to undertake in order to obtain reliable efficiency scores[2]. Regarding this point, the literature has identified several issues that can be detrimental to DEA estimates, such as the inclusion of irrelevant variables in the model (Smith, 1997; Simar and Wilson, 2001), the omission of relevant inputs (Galagedera and Silvapulle, 2003; Ruggiero, 2005) or the presence of correlation between inputs (Smith, 1997; Pedraja et al., 1999).

[2] This choice tends to depend on users' expertise and judgement, as well as data availability. The literature provides different methods that can be used to guide variable selection. See Natarajan and Johnson (2011) for a good survey of such methods.

However, there is another major concern, namely, the presence of endogeneity in the production process, which is frequently overlooked when practitioners apply DEA even though it is an issue that has received plenty of attention in econometrics (Wooldridge, 2002). There are several potential sources of endogeneity, such as when a two-way causal relationship is observed between inputs and outputs. The idea behind this concept is that some inputs are not exogenous to the production model, but are determined within the model. For instance, it is quite common in sectors like education or health insurance where better providers are sometimes in a position to select their consumers and vice versa (Parry, 1996). For example, this is the process, usually known as cream skimming, by which more motivated parents generally choose the best schools and people in a better state of health tend to have more insurance benefits. Also, this problem can arise in the opposite direction, for example, when the worst producers receive more resources (input) to improve their observed poor results (outputs).

As we mentioned above, the distortions in the estimation of economic models potentially caused by such endogeneity have been widely studied in the econometrics literature. However, its effects on efficiency measures calculated using non-parametric techniques like DEA have not yet been analyzed in depth. There are only a handful of studies that have tested the performance of DEA under endogeneity using different experimental designs (Orme and Smith, 1996; Bifulco and Bretschneider, 2001, 2003; Ruggiero, 2003, 2004). Using alternative simulation strategies and data generation processes, all the above research concludes that DEA estimates can be biased if there is a certain level of correlation between one input and true efficiency.

In this paper we set out to generalize the effect of endogeneity on DEA estimations in order to overcome some of the limitations of previous research into this issue by expanding the analysis in various directions. Firstly, most of the above studies basically focus on comparing the performance of DEA with alternative methods to measure technical efficiency instead of determining how the presence of endogeneity affects DEA estimates[3]. Secondly, they all use experimental designs based on a Cobb-Douglas production function which fail to capture the potential nonlinear effects of inputs on the output variable[4]. Thirdly, none perform a Monte Carlo experiment to get results. Finally, almost all previous studies examine the effect of negative endogeneity only. The sole remarkable exception is Ruggiero (2004), who analyses the effect of one nondiscretionary input that positively influences both the output and technical efficiency simultaneously.

The aim of this paper is to analyze whether endogeneity can bias the results of DEA in order to make practitioners using this technique aware of the accuracy of their estimates. For this purpose, we simulate different types of endogeneity in synthetic data by examining the effect of both the negative and the positive correlation between one input and the true efficiency level on DEA estimated scores. To do this, we perform a Monte Carlo experiment using a more flexible *translog* production function simulating different intensities and signs of correlations. The article is organized as follows. Section II introduces some basic concepts about endogeneity and its potential effects on DEA estimates. Section III describes the methodology used to generate the synthetic data in our Monte Carlo experimental design. Section IV presents the main results of the analysis. The paper concludes with a discussion of the main implications of our findings for practitioners using DEA to measure technical efficiency in different contexts and some directions for future research.

2. The endogeneity issue

Endogeneity is said to exist in an economic model when an explanatory variable is correlated with the error term. This phenomenon can arise as the result of several problems: measurement errors, omitted variables in the model specification or sample selection errors, although perhaps the most common cause is the presence of two-way causal relationships between the dependent and independent variables. In the

---

[3] For example, Bifulco and Bretschneider (2001, 2003) and Ruggiero (2003) compared DEA with corrected ordinary least squares.

[4] This can be a significant weakness in complex frameworks such as the public sector education or health provision.

production economics field, this framework can be represented as follows. Let us assume the following production function:

$$Y_i = F(x_{1i}x_{2i} \dots x_{ki}) * \theta_i \qquad (1)$$

where i denotes the ith production unit, Y_i is the level of the observed output, x_{ki} are the k productive inputs and θ_i is the technical efficiency, ranging from 0 to 1. The endogeneity issue arises in association with efficiency measurement when the correlation between at least one input (x_{ki}) and efficiency (θ_i) is different from zero: $corr\ (x_{ki}, \theta_i) \neq 0$.

The education sector is a good example to illustrate this problem[5]. In this context, it is claimed that schools with better academic outcomes tend to attract relatively more advantaged students from a high socioeconomic background. If parent motivation is correlated with socio-economic level, such pupils (and thus the school they attend) will tend to obtain better academic results for two reasons. Firstly, socio-economic level is an essential input for producing educational output. Secondly, motivation has a positive effect on school efficiency. Consequently, we will observe that schools whose students are from a high socio-economic background are more prone to be fully efficient. This mechanism results in a positive correlation between socio-economic background and technical efficiency. A similar thing applies to the quality of teachers in public education systems where the teachers that come top in civil servant entrance exams get first choice of school. Schools that have higher examination pass rates, a lower ratio of repeating and disadvantaged pupils, better facilities and students from higher income families are likely to attract higher quality teachers. Again, the positive two-way causal relationship between the input and the output occurs because highly qualified and motivated teachers self-select into the best schools and also have a positive effect on student results at those schools.

[5] Mayston (2003) reported a detailed set of possible sources of endogeneity in educational contexts.

However, the endogeneity problem can also arise in the opposite direction when a direct negative feedback from outputs to resources is observed. This applies 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, Levacic and Vignoles, 2002). If lower results are due to a high inefficiency, then the reverse causality implies allocating more resources to inefficient schools causing a negative correlation between resources (input) and efficiency.

In this context, the use of econometric techniques like OLS will produce biased and inconsistent estimates of parameters, as one of the most important econometric assumptions (non-correlation between the error term and independent variables) does not hold in practice. Actually, the problem of self-selection has been argued to be the basis for multiple theoretical and empirical critiques of the findings using conventional econometric techniques in the field of economics. As a result, multiple methods have been developed in the literature to deal with this problem (Schlotter *et al.*, 2011).

The endogeneity problem has also been a factor considered recently in the estimation of technical efficiency with parametric frontier techniques. 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 stochastic frontier analysis and apply it to measure state health system performance. Perelman and Santín (2011) address the endogeneity problem of school choice in Spain with instrumental variables. Finally, Mayen et al. (2010), Bravo et al. (2012)[6] and Crespo-Cebada et al. (2013) apply propensity score matching to American dairy farms, farmers in Honduras and education in Spain, respectively.

Nevertheless, the potential distortions of the measurement of technical efficiency using nonparametric techniques caused by endogeneity have received much less attention in

---

[6] This paper also applies the Greene's procedure (2010).

the literature. In principle, it might seem that this technique should not be influenced by this problem, since it constructs a boundary around feasible combinations of inputs and outputs without assuming a parametric functional form (Orme and Smith, 1993). However, the existence of interrelationships between inputs, outputs and the level of efficiency can also change the set of observed points, so DEA estimates can also been distorted by endogeneity bias, as we explain below.

The DEA methodology was originally proposed by Charnes, Cooper and Rhodes (1978) and Banker, Charnes and Cooper (1984)[7]. The output-oriented problem of calculating the efficiency of a production process under variable returns to scale (DEA-BCC model) can be specified as

$$Max\ \delta_0 \qquad (2)$$

$$s.t. \quad \sum_{i=1}^{N} \lambda_i y_{ri} \geq \delta y_{r0}$$

$$\sum_{i=1}^{N} \lambda_i x_{ki} \leq x_{k0}$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

$$i = 1, \dots, N \quad r = 1, \dots., s \quad k = 1, \dots, m$$

where x_k denotes input k, y_r stands for output r and i is the DMU. The λ vector contains the virtual weights of each DMU determined by the problem solution. If $\delta_i = 1$, the analyzed unit is fully efficient, whereas $\delta_i > 1$ indicates that the ith DMU is inefficient, δ_i being the distance between the *i*th unit and the estimated frontier. The technical efficiency score of the *i*th DMU is defined by $\hat{\theta}_i = 1/\delta_i$, which is bounded between zero and one, where $\hat{\theta}_i = 1$ represents an efficient unit and the efficiency of the observed units decreases proportionally to the value of $\hat{\theta}_i$.

[7] The first model assumes constant return to scale and is known as the CCR model. The second model assumes variable return to scale and is identified as the BCC model.

As mentioned in the introduction, one of the main implicit assumptions of DEA is monotonicity. According to Färe and Primont (1995), a multi-input multi-output production technology can be characterized by the technology set S as

$$S = \{(x, y): x \text{ can produce } y\}$$

Monotonicity assumes that if $x_0 \geq x_1$, then $y_0 \geq y_1$. In practice, this property implies that there is a positive causal relationship between a vector of inputs which is able to produce a vector of outputs that we denote x and y, respectively. Consequently, most researchers using this method implicitly assume that this is a univocal relationship from inputs to outputs, and therefore the level of technical efficiency is exogenous and independent of the input and output levels. To illustrate this idea, Figure 1 represents a single-input (X) / single-output (Y) production setting in which true efficiency θ_i is exogenously distributed, *i.e.*, there is no correlation between the level of input X and the efficiency θ_i. 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 spread along the production frontier.

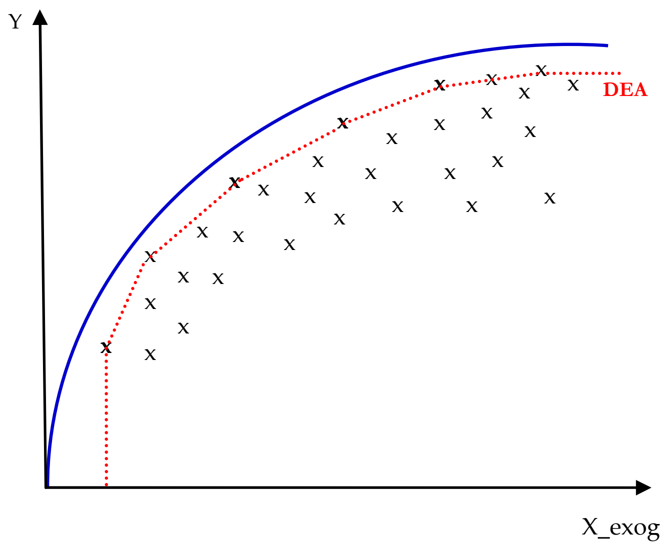

Figure 1. True frontier and DEA-BCC estimates in an exogenous scenario.

However, as noted in above Section 2, we may well find, in real-world production processes, a correlation between the true efficiency and the level of input that is significantly different from zero: $corr(x, u) \neq 0$. This correlation can be either positive or negative, as mentioned previously in the examples of different educational settings. Figure 2 illustrates the situation where endogeneity is positive.

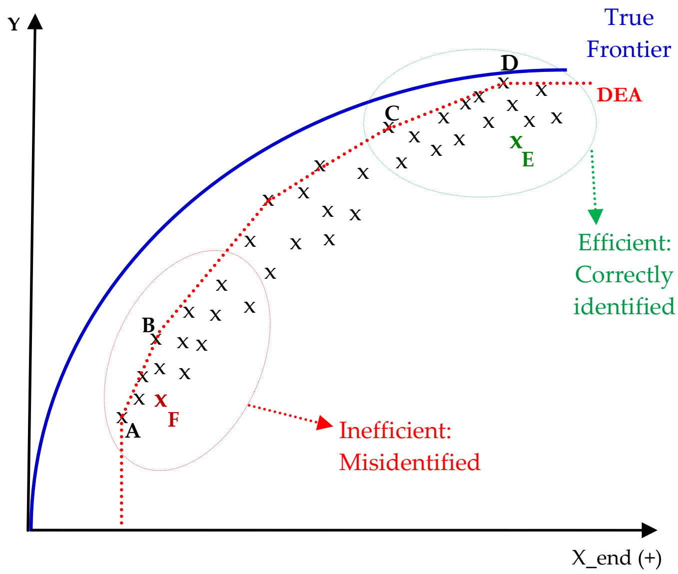

Figure 2. True frontier and DEA-BCC estimates under positive and high correlation between the true efficiency level and one input. The higher the input level, the greater efficiency.

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 the true frontier, DMUs with higher levels of input (and outputs), e.g., dots C and D in Figure 2, are fully efficient units, whereas DMUs with lower input levels are less efficient. However, as DEA estimates efficiency scores based on observed data, the frontier built by DEA will find and classify some DMUs that have a low input level and are really highly inefficient as efficient. This is the case of dots A and B in Figure 2, which are actually very far away from the true frontier but are identified by DEA as efficient units. Consequently, the border estimated by DEA will be far removed from the true one in the lower input frontier region. This means that efficiency improvement targets will 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. 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 policy making.

On the other hand, the existence of a significant negative correlation between the input level and the true efficiency can also bias DEA estimates. Nonetheless, this kind of endogeneity makes more sense from a microeconomic perspective. At a constant output level, a high input rate implies less technical efficiency. Figure 3 illustrates this scenario where efficient units with low input levels and inefficient units with high input levels are all correctly identified since the true frontier is properly identified by DEA at almost all the input level range.

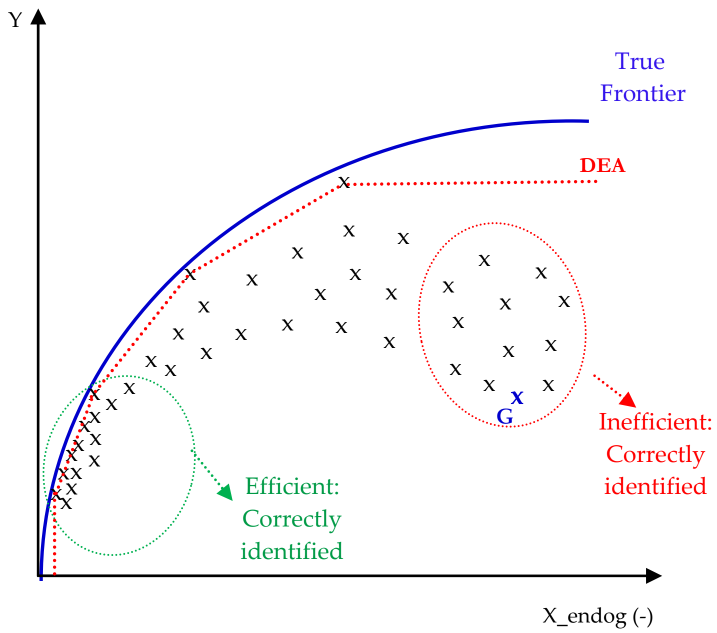

Figure 3. True frontier and DEA-BCC estimates under negative and high correlation between the true efficiency level and one input. The higher the input level, the greater inefficiency.

It is worth to note, that at the region of high input level the estimated frontier shifts down from the true one, but inefficient DMUs that use more input remain far away from the DEA frontier to be identified as inefficient producers compared with other DMUs. For instance, unit G is highly inefficient, and, although the DEA frontier at that point is slightly removed from the real one, the distance between G and the estimated frontier is still large enough in terms of output for it to be recognized as an inefficient producer. Thus, under negative endogeneity the efficiency scores estimated by DEA appear to conform better to the true relative positions, and thus the estimated ranking is not significantly different from the true one.

The potential implications of these problems in quantitative terms have to be measured from a theoretical point of view. The aim of this paper is precisely to alert DEA practitioners to how the presence of either type of endogeneity in the production process can potentially bias the results. On this ground, we test DEA performance under endogeneity of different signs and intensity in Section 3. To do this, we simulate endogeneity through a significant positive or negative correlation between the true technical efficiency and one input.

3. Experimental design and data generation process

In order to illustrate the ideas developed above, we perform a Monte Carlo experiment applied to seven scenarios. Firstly, we use a data generation process (DGP) to create a baseline dataset without endogeneity (the *exogenous* scenario). Secondly, we simulate six alternative settings taking into account the correlation between the true efficiency (θ_i) and one observed input (the *endogenous* scenarios). Results from each endogenous scenario are then compared to the baseline scenario in order to measure the effects that endogeneity has on DEA estimations. All datasets were defined in a single output framework with three inputs. The first decision to be made in the DGP as part of the experimental design was to choose the functional form for the production function.

3.1. The production function

Almost all previous studies in the literature have simulated data using the Cobb-Douglas production function. For the sake of comparability, we also draw data from a Cobb-Douglas function with a single output and three inputs:

$$y_i = x_{1i}^{\alpha}\, x_{2i}^{\beta}\, x_{3i}^{\gamma}, \tag{3}$$

where y_i represents the output, and x_1, x_2 and x_3 are the observed inputs. The input weights assigned in this research were α=0.3, β=0.35 and γ=0.35, assuming constant returns to scale[8]. Although this functional form is the most commonly used in economics and operational research, the assumption of constant input-output elasticities is a significant drawback. This means that regardless the scale of production, the marginal effects of inputs on outputs are the same. Therefore, the Cobb-Douglas production function fails to capture potential nonlinear effects of those resources. Since the main aim of this work is to test the accuracy of DEA in an experimental setting that reproduces a more realistic context, we also considered a more flexible technology in our experimental design, namely, the *translog* production function introduced by Christensen et al. (1971):

$$ln\, y = \beta_0 + \sum_{k=1}^{K} \beta_k * lnx_k + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{J} \beta_{kj} * lnx_k * lnx_j, \tag{4}$$

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

Firstly, we verified the monotonicity property, where all marginal products must be nonnegative in a single-output case. For the *translog* production function, this implies:

$$\partial y/\partial x_k = y/x_k * \partial lny/\partial lnx_k = y/x_k * \{\beta_k + \textstyle\sum_{i=1}^{N} \beta_{ki} * lnx_k\} = y/x_k * \varepsilon_k \geq 0 \;\; \forall k \tag{5}$$

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

As the average product y/x_k is always positive, monotonicity implies that all input-output elasticities ε_k must be nonnegative across the entire input range.

Secondly, we checked for concavity in all inputs, which implies that all marginal products must be nonincreasing, *i.e.*, the law of diminishing marginal productivity (Coelli *et al.*, 2005). For the *translog* production function, all inputs must satisfy the following expression across the entire simulated data range:

$$\partial^2 y/\partial\, x_k^2 = \{\beta_{kk} + \varepsilon_k^2 - \varepsilon_k\} * y/x_k^2 \leq 0 \quad \forall k \qquad (6)$$

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. In the field of education, for example, if the initial school input endowments are all doubled, it would be reasonable to expect a less-than-double increase in students' test scores, particularly at high levels of educational attainment (Essid *et al.*, 2013).

3.2. DGP for the baseline scenario

The baseline scenario represents the exogenous case, where no inputs are correlated with the true technical efficiency. It is simulated according to the following procedure:

1. Randomly and independently generate three input vectors x_{1i}, x_{2i} and x_3 using a uniform distribution over the interval [5, 50] for N DMUs, n=1,...,N.
2. Calculate the efficient output level as $y_i = F(x_{1i}, x_{2i}, x_{3i})$ using Eq. (3) or Eq. (4), respectively.
3. Draw a random error term v_i from $N(0; 0.04)$, which represents 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 random disturbances. As demonstrated in previous studies, the larger the measurement error, the poorer the performance of DEA (Bifulco and Bretschneider, 2001). Therefore, we chose a small measurement error

in order to generate some noise but not so much as to distort the analysis of endogeneity.

4. Compute the observed output as $\hat{y}_i = y_i \,.\, exp(v_i) \,.\, \theta_i$, where θ_i is the true technical efficiency level generated in the 13th step of the endogenous dataset procedure.

3.3. DGP for the endogenous scenarios

The remaining six scenarios were developed using a similar DGP, albeit taking into account the existence of endogeneity, which was modeled by Pearson's correlation coefficient between the true technical efficiency θ_i and one observed input. Therefore, we substitute the exogenous input x_3 for an endogenous input x_{end} in each dataset. In order to compute the endogenous input with the same distribution as the exogenous inputs (x_1, x_2, x_3) and with a specific level of correlation with θ_i, we enact, for each endogenous scenario, the following procedure:

1. Select the desired Pearson's correlation coefficient between x_{end} and θ ($\rho_{x_{end},\theta}$).
2. Draw a random matrix $A = (a_1, a_2)$ from a multivariate normal distribution $N(0; \Sigma)$, with the following variance-covariance matrix $\Sigma = \begin{bmatrix} 1 & \rho_{x_{end},\theta} \\ \rho_{x_{end},\theta} & 1 \end{bmatrix}$.
3. Compute an identification number variable (ID) from 1 to *N*.
4. Match the ID with the vector a_1 to get $B = [ID \quad a_1]$. Sort B by a_1 in ascending order (the ID variable will be unsorted): $B' = [ID_{a1} \quad a_{1s}]$.
5. Generate an independent vector $x_{(n\times 1)}$ from a uniform distribution over the interval [5, 50] and sort in ascending order to get x_s.
6. Compute a new C matrix by merging B$'$ with x_s : $C = [ID_{a1} \quad a_{1s} \quad x_s]$.
7. Sort C by the ID variable in ascending order: $C' = [ID \quad a_{1s\_ID} \quad x_{end}]$.
8. The latter vector of C$'$, (x_{end}), will be used as the endogenous input.
9. Match ID with the vector a_2 to get $D = [ID \quad a_2]$. Sort D by a_2 in ascending order (the ID variable will be unsorted): $D' = [ID_{a2} \quad a_{2s}]$.
10. Randomly and independently generate n values of u_i using a half-normal distribution $u_i \sim |N(0; 0.20)|$. Then compute the vector $e = exp(-u)$ and sort this variable in ascending order to get e_s.
11. Compute a new E matrix by merging D$'$ with e_s: $E = [ID_{a2} \quad a_{2s} \quad e_s]$.

12. Sort E by the ID variable in ascending order: $E' = [ID \quad a_{2s\_ID} \quad e_{s\_ID}]$.
13. The latter vector of E', (e_{s_ID}), will be used as the true technical efficiency level for each unit: θ_i. The generated average efficiency in each experiment ranges from 0.829 to 0.883 with a standard deviation of from 0.078 to 0.118, respectively.
14. Use the exogenous inputs x_1 and x_2 generated in the baseline scenario and the endogenous input x_{end} to compute the efficient output as $y_{endi} = F(x_{1i}, x_{2i}, x_{endi})$ using Eq. (3) or Eq.(4), respectively.
15. Finally, calculate the observed output using the random term v_i computed in the baseline dataset and the true efficiency level θ_i computed in step 13: $\hat{y}_{endi} = y_{endi} \, . \, exp(v_i) \, . \, \theta_i$.

Two factors were varied in order to generate the six endogenous settings: the sign (negative or positive) and the intensity (high, medium or low) of the correlation coefficient between the true efficiency and the endogenous inputs ($\rho_{x_{end,\theta}}$).

Table 1 summarizes the main descriptive statistics of the correlation coefficients that were actually obtained in each simulated scenario[9].

Table 1. Descriptive statistics for the correlation between the true technical efficiency and the endogenous input after the Monte Carlo experiment in each simulated scenario.

	Negative correlation			**Positive correlation**		
	HIGH	MEDIUM	LOW	HIGH	MEDIUM	LOW
				Cobb-Douglas		
Mean	-0.860	-0.472	-0.282	0.859	0.464	0.281
Std. Deviation	0.028	0.080	0.091	0.027	0.079	0.093
				Translog		
Mean	-0.864	-0.465	-0.279	0.863	0.472	0.284
Std. Deviation	0.026	0.078	0.093	0.027	0.077	0.095

[9] The correlation coefficients were computed in each experiment and then averaged to obtain the final measures presented in Table 1.

All scenarios were replicated using the Cobb-Douglas and the *translog* production functions for a sample size of 100 DMUs[10]. Finally, for each dataset, we estimate the efficiency scores, $\widehat{\theta_i}$, by running an output oriented DEA model under constant and variable returns to scale (CRS and VRS), as proposed by Charnes *et al.* (1978) and Banker *et al*. (1984), respectively. As a result, 28 scenarios were analyzed (the exogenous scenario, six types of endogeneity, two production technologies and two types of return to scale). In order to make the results more reliable, we ran a Monte Carlo experiment, where B, the number of replicates, was 1,000[11]. Consequently, all measures were computed in each replication and, finally, averaged to yield the results reported in Section 4.

4. Monte Carlo experiment results

4.1. Accuracy measures

In order to test the adequacy of DEA under endogeneity, we present a set of accuracy measures. Firstly, we are interested in measuring how accurately DEA ranks observations. For this purpose, we compute Spearman's rho (ρ_S) correlation coefficients between the true efficiency and estimated score pairs. The higher the correlation coefficient ρ_S, the better able DEA is to identify the true efficiency distribution. The first two columns of Table 2 list these coefficients for the DEA-CRS and DEA-VRS models under different endogenous scenarios compared to the exogenous baseline assuming data from a Cobb-Douglas production function. Table 3 contains equivalent results for a *translog* DGP.

Secondly, we are interested in testing how well the model estimates the true level of efficiency. For this purpose, we average the estimated efficiency scores (mean estimated efficiency) which we compare with the true mean efficiency. If the mean

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

[11] Simulations were carried out using MATLAB 7.6.0 software.

estimated efficiency is greater (smaller) than the true mean efficiency, DEA overestimates (underestimates) the true efficiency level. Finally, we also calculate the mean absolute error (MAE), $MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{\theta}_i - \theta_i\right|$, which is the result of computing and averaging the sum of absolute deviations of DEA estimated scores from the true efficiency level for each observation. A low MAE implies that, on average, the estimates are near to the true efficiency values, and hence small values are preferred.

All Monte Carlo results are reported in Table 2 and Table 3 for data generated from the Cobb-Douglas and *translog* production technologies, respectively, assuming both CRS and VRS.

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. We then examined how well the technique was able to place observations in the appropriate quintile. Using this complementary measure we can evaluate technique accuracy at different points of the distribution, and hence it is a useful tool for locating the main drawbacks of the technique. For example, for research aiming to identify best practices, we will be more interested in the percentage of top quintile observations that DEA correctly places in the top quintile rather than in the accuracy of the overall ranking. Results from this analysis are reported in Table 4 and Table 5 for the Cobb-Douglas and the *translog* DGPs, respectively, assuming CRS and VRS.

Table 2. DEA estimate accuracy measures (Cobb Douglas).

Intensity and sign	**Spearman's correlation**		**Average estimated efficiency**		**MAE**	
	CRS	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**
Exogenous	0.783	0.672	0.883	0.912	0.048	0.065
Negative endogeneity						
High	0.674	0.534	0.901	0.955	0.061	0.103
Medium	0.766	0.680	0.888	0.918	0.050	0.068
Low	0.778	0.688	0.885	0.913	0.049	0.065
Positive endogeneity						
High	0.465	0.236	0.933	0.947	0.094	0.099
Medium	0.702	0.538	0.893	0.921	0.056	0.075
Low	0.744	0.606	0.902	0.898	0.066	0.186

Results from Monte Carlo experiments. N=100. B=1000.

Table 3. DEA estimate accuracy measures (translog).

Intensity and sign	**Spearman's correlation**		**Average estimated efficiency**		**MAE**	
	CRS	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**
Exogenous	0.666	0.729	0.816	0.894	0.083	0.072
Negative endogeneity						
High	0.711	0.688	0.799	0.922	0.091	0.097
Medium	0.717	0.762	0.835	0.897	0.116	0.074
Low	0.701	0.755	0.814	0.893	0.091	0.072
Positive endogeneity						
High	0.227	0.265	0.947	0.945	0.154	0.123
Medium	0.548	0.594	0.845	0.907	0.086	0.087
Low	0.601	0.659	0.829	0.900	0.084	0.079

Results from Monte Carlo experiments. N=100. B=1000.

4.2. Baseline scenario results

The results confirm that DEA performs well in the exogenous case regardless of the production function or the assumed returns to scale. However, as expected, DEA-CRS estimates outperform DEA-VRS for data generated with a Cobb-Douglas production function and vice versa for data derived from the *translog* DGP. For example, under the Cobb-Douglas assumption, Spearman's correlation coefficients between the true and estimated efficiency are 0.783 under CRS and 0.672 under VRS, whereas these figures are 0.666 and 0.729, respectively, for the *translog* DGP. This finding highlights the importance of making a correct choice of the assumed returns to scale before conducting a DEA efficiency analysis. In view of this evidence, we will now refer to - DEA-CRS for the results estimated from the Cobb-Douglas and to DEA-VRS results for the *translog* scenarios.

Results from Table 4 and Table 5 also confirm that DEA performs adequately in the exogenous case. Almost 50% of observations are placed in the correct quintile, and only about one out of every eight units are assigned to quintiles that are two or more quintiles removed from where they should be. We find from columns 5 to 12 that DEA's major weakness lies in its ability to correctly identify the most efficient DMUs. Whereas about three quarters of the most inefficient units are correctly assigned to the bottom quintile, this proportion drops to around 50% for units properly placed in the top quintile. Also, the percentage of units which should really be in the top two quintiles and are actually placed in the bottom quintile is close to zero; but this figure rises to 7% and to 12.5% for observations that should really be in the bottom quintile and are assigned to the top quintile for CRS and VRS, respectively. This evidence should be taken into account especially if DEA is conducted for the purpose of performance-based reforms. In this case, some units which are not actually benchmarks would be identified as such.

4.3. Endogeneity effects

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 CRS and VRS estimates are similar, being more pronounced in the case of VRS than for CRS. For instance, Spearman's correlation coefficient between the true efficiency and DEA estimates is 0.783 under CRS and 0.729 under VRS in the baseline scenario, whereas these correlations drop to 0.465 and 0.265, respectively, when high and positive endogeneity is introduced. This performance can be explained by the fact that the technique is more sensitive to changes in the data distribution under VRS than under CRS assumptions. Given that VRS is a more realistic and frequent assumption in real-world applications and conclusions are similar for both types of returns to scale, we will only discuss results for DEA-VRS more in depth (Table 3 and Table 5)[12].

The main finding from our simulations is that positive and high endogeneity is the worst possible scenario, shattering DEA performance. As the intensity of positive endogeneity decreases to medium, DEA results improve but remain still poor compared with the exogenous scenario. Finally, for low positive endogeneity, DEA estimations are very close to baseline scenario values. As Table 3 shows, the exogenous dataset simulations yield a Spearman's correlation coefficient of 0.729 between estimated and actual efficiency, which drops to 0.265 in the presence of high and positive endogeneity and to 0.594 in the case of medium positive endogeneity. The MAE underscores this result. The value of MAE in the positive and high endogeneity scenario climbs to 0.123, which is significantly greater than the 0.072 calculated in the exogenous baseline scenario. Average estimated efficiency (column 4 of Table 3) is another way to observe the effects on the estimated efficiency level. It reveals that DEA overestimates the true mean technical efficiency under both types of endogeneity (negative and positive), particularly when the input and the efficiency are highly correlated.

[12] Results are also reported for CRS.

The damage caused by endogeneity can also be evaluated by means of the proportion of units assigned to the correct quintile by DEA. According to Table 5, the DMUs correctly assigned to the correct quintile drops from almost 47.3% in the exogenous setting to 25.7% (38.9%) under high (medium) and positive endogeneity. Additionally, as Table 5 shows, the proportion of units assigned two or more quintiles from actual triplicates the baseline percentage (increasing from 13.4% to 38.4%) in the case of high positive endogeneity. The quintile analysis indicates that the decline in DEA performance is further driven by the fact that, under positive and high endogeneity, the technique identifies several of what are really the most inefficient units as efficient. When there is a high and positive correlation between true efficiency and one input, only one third of units assigned by DEA to the bottom quintile were actually in the bottom quintile, whereas the equivalent percentage was as high as 74.7% in the exogenous scenario. In addition, the percentage of DMUs placed in the top quintile when they should actually have been in the bottom two quintiles triples the baseline scenario values. These results confirm that, at low levels of endogenous input, the frontier estimated by DEA (which is driven by the data shape) is located further from the true frontier, identifying quite a lot of inefficient units as very efficient, as illustrated in Figure 2.

As we discussed 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 such endogeneity also degrades DEA's ability to correctly identify the most efficient DMUs. For instance, the proportion of units correctly assigned to the top quintile drops from 48% to 31%. Furthermore, whereas no units that should actually be ranked in the top two quintiles are found to be assigned to the bottom quintile in the absence of endogeneity, 12.6% of DMUs are wrongly ranked in the bottom quintile under high positive endogeneity.

Table 4. Measures of how accurately DEA-CRS and DEA-VRS assign observations to quintiles (Cobb Douglas).

Sign and intensity of endogeneity	**% Assigned to the correct quintile**		**% Assigned to a quintile two or more quintiles removed from the one to which they belong**		**% Assigned to bottom quintile and actually in bottom quintile**		**% Assigned to top quintile and actually in top quintile**		**% Assigned to bottom quintile and actually ranked in the top two quintiles**		**% Assigned to top quintile and actually ranked in the bottom two quintiles**	
	CRS	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**
Exogenous	48.8	42.7	11.6	17.4	74.9	69.8	52.9	42.9	0.3	0.6	6.7	14.2
NEGATIVE												
High	42.6	33.7	18.2	26.8	67.0	55.2	47.0	36.2	1.6	7.0	13.7	19.5
Medium	48.0	41.7	12.5	17.4	73.6	68.4	51.8	42.0	0.5	1.2	8.1	12.8
Low	48.7	42.7	11.9	16.9	74.4	69.9	52.3	42.4	0.3	0.7	7.1	12.9
POSITIVE												
High	32.2	24.9	28.9	39.3	45.2	31.4	39.7	30.7	6.1	14.6	24.5	35.0
Medium	43.3	36.1	16.4	24.4	67.8	58.9	46.8	37.7	0.9	2.1	11.6	21.6
Low	45.8	39.5	14.1	20.9	71.8	65.2	49.6	40.3	0.6	1.1	8.7	18.1

Means values after 1000 replications. N=100.

Table 5. Measures of how accurately DEA-CRS and DEA-VRS assign observations to quintiles (*translog*).

Sign and intensity of endogeneity	**% Assigned to the correct quintile**		**% Assigned to a quintile two or more quintiles removed from the one to which they belong**		**% Assigned to bottom quintile and actually in bottom quintile**		**% Assigned to top quintile and actually in top quintile**		**% Assigned to bottom quintile and actually ranked in the top two quintiles**		**% Assigned to top quintile and actually ranked in the bottom two quintiles**	
	CRS	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**	**CRS**	**VRS**
Exogenous	39.6	47.3	20.5	13.4	64.8	74.7	45.7	48.0	2.5	0.1	7.8	11.2
NEGATIVE												
High	42.7	41.6	16.5	17.6	64.3	70.3	54.0	42.3	0.7	1.8	10.1	11.4
Medium	42.7	48.1	16.7	11.8	66.6	75.6	51.2	48.0	1.4	0.2	7.0	9.3
Low	41.6	48.3	18.0	12.0	66.4	76.2	49.2	47.7	1.7	0.2	6.9	10.4
POSITIVE												
High	24.9	25.7	40.3	38.4	34.0	34.2	26.4	31.1	22.2	12.6	32.1	33.2
Medium	33.9	38.9	26.9	20.7	56.1	62.7	37.4	41.0	5.7	0.9	14.1	19.6
Low	36.3	42.9	24.4	17.1	60.4	69.3	41.4	43.8	4.6	0.4	10.6	16.6

Means values after 1000 replications. N=100.

Finally, under low and medium negative endogeneity, Monte Carlo simulations evidence that DEA estimates remain robust. Estimations appear to be slightly impaired only in the scenario where negative endogeneity is high. These results are similar to findings by Bifulco and Bretschneider (2003), who conclude, for the same measurement error as our simulation, that DEA performance is not substantially affected by the negative endogeneity. This finding can be explained by the fact that negative endogeneity correlates input and efficiency in the same way as DEA assumes to plot the frontier (*i.e.*, the higher the input level, the lower the efficiency score). Consistently, estimates are unaffected by this kind of endogeneity.

In summary, we can conclude from our results that DEA-CRS and DEA-VRS provide accurate efficiency measures in all scenarios, save when there is a medium or high positive correlation between one input and the true efficiency. Note again that DEA estimates will be far removed from the actual efficiency values in the presence of a high positive endogeneity regardless of the assumed functional form. This result is worthy of mention since such endogenous scenarios are similar to the ones that are likely to be found in public sector efficiency analysis applications (due to a two-way causality or an omitted variable). Therefore, this evidence suggests that, unless the presence of endogeneity is taken into account, the use of DEA models to estimate technical efficiency could, in such cases, impair efficiency estimates and thus result in inappropriate performance-based recommendations.

5. Concluding remarks

The endogeneity problem and the distortions in the estimation of efficiency and production models that it causes is a growing concern in the frontier analysis field. Most reported research is starting to apply conventional econometric approaches to deal with endogeneity in the estimation of production frontiers and efficiency using parametric techniques. Although non-parametric methods like DEA are the most used techniques for measuring technical efficiency, the effects of endogeneity on efficiency estimates have received little attention in the literature.

In this paper, we analyze the extent to which endogeneity in the production process can bias DEA estimations. For this purpose, we simulate different levels of negative and positive endogeneity through the correlation between one input and the true efficiency using synthetic data generated in a Monte Carlo experiment. We conduct the analysis using a Cobb-Douglas production technology, as well as a more flexible *translog* specification.

We conclude that DEA is robust to the presence of negative and low positive endogeneity. However, a high and medium positive endogeneity, *i.e.*, a high positive correlation between one input and the true efficiency level, significantly biases DEA performance. For example, while for the *translog* production function in the exogenous scenario DEA-VRS yields a Spearman's correlation coefficient of 0.729 between the estimated and actual efficiency, this correlation drops to 0.594 in the case of medium positive endogeneity and to 0.265 in the presence of high positive endogeneity.

DEA's ability to properly assign DMUs to their correct quintile also decreases when endogeneity is introduced. Whereas DEA correctly identifies 47.3% of DMUs in the exogenous setting, this percentage falls to 38.9% and 25.7% under medium and high positive endogeneity, respectively. The quintile analysis denotes that the decline in DEA performance is further driven by the misidentification of efficient DMUs. For instance, under high positive endogeneity, DEA triplicates, with respect the baseline scenario, the proportion of DMUs assigned two or more quintiles away from actual and the proportion of DMUs placed in the top quintile when they were actually in the bottom two quintiles.

Following from the above, the main result of this paper is to caution DEA practitioners concerning the accuracy of their estimates when they suspect that there is a high correlation between one of their inputs and technical efficiency. The assessment of exactly how common this issue is in empirical frontier analysis is beyond the scope of this research and should be tested in real-world problems. We think that more research is still needed in three directions. First, the main sources of endogeneity affecting the estimation of production frontiers should be thoroughly conceptualized from a theoretical point of view. Second, it is necessary to define a method to test for endogeneity in empirical problems. Finally, a technique should be developed to deal with endogeneity in order to improve DEA estimations.

References

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, 1078–1092.

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

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

Bowlin, W. (1986). "Evaluating Performance in Governmental Organizations". *The Government Accountant´s Journal*, 35, 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.

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 Production Function”. *Econometrica*, 39, 255-256.

Crespo-Cebada, E., Pedraja-Chaparro 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

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., and Battese, G. E. (2005). “An introduction to efficiency and productivity analysis”. *Springer*.

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.

Färe, R., and Primont, D. (1995). “Multi-output production and duality: theory and applications”. Kluwer Academic Pub.

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

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

Levacic, 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., Alexander C.E. (2010). “Technology adoption technical efficiency: organic and conventional dairy farms in United States”. *American Journal of Agricultural Economics,* 92 (1), 181-195.

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

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

Orme, C. and Smith, P. (1996). "The Potential for Endogeneity Bias in Data Envelopment Analysis". *Journal of the Operational Research Society*, 47, 73-83.

Parry, Taryn R. (1996). "Will Pursuit of Higher Quality Sacrifice Equal Opportunity in Higher Education? An Analysis of the Education Voucher System in Santiago". *Social Science Quarterly*, 77(4), 821-841.

Pedraja-Chaparro, F., Salinas-Jimenez, J. and Smith, P. (1999). "On the Quality of the Data Envelopment Analysis Model". *Journal of Operational Research Society*, 50, 636-644.

Perelman, S. and Santín, 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.

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

Ruggiero, J. (2004). "Performance Evaluation when Non-Discretionary Factors Correlate with Technical Efficiency". *European Journal of Operational Research*, 159, p.250-257.

Ruggiero, J. (2005). "Impact assessment of input omission on DEA". *International Journal of Technology and Decision Making*, 4(3), 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. and Wilson, P.W. (2001). "Testing restrictions in nonparametric efficiency models". *Communications in Statistics, Simulation and Computation*, 30, 159–184.

Smith, P. (1997). "Model misspecification in data envelopment analysis". *Annals of Operations Research*, 73 (0), 233–252.

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