European Journal of Operational Research 290 (2021) 657–670

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Methodology for calculating critical values of relevance measures in variable selection methods in data envelopment analysis

Jeyms Villanueva-Cantillo[a], Manuel Munoz-Marquez[b,*]

[a] Faculty of Basic and Biomedical Sciences, Universidad Simón Bolívar, 59th Avenue # 59 - 65, Barranquilla 080002, Colombia
[b] Statistics and Operational Research Department, Universidad de Cádiz, 11510 Puerto Real, Cádiz, Spain

ARTICLE INFO

Article history:
Received 3 April 2019
Accepted 12 August 2020
Available online 19 August 2020

Keywords:
Data envelopment analysis
Variable selection
Critical values
Monte Carlo simulations

ABSTRACT

The selection of input and output variables is a key step in evaluating the relative efficiency of decision-making units (DMUs) in data envelopment analysis (DEA). In this paper, we present a methodology based on Monte Carlo simulations and bootstrapping for calculating the critical values of relevance measures in variable selection methods in DEA. Additionally, we define a set of metrics to study the methods' performance when using such critical values. We conducted an extensive simulation study, applying the proposed methodology to two variable selection methods in 28 single-output model specifications (i.e., different number of inputs and DMUs in the DEA model) under multiple scenarios, varying factors related to the functional form of the production function, the probability of an input being relevant in the model, the probability distribution of the inputs, and the theoretical efficiencies of the DMUs. The simulation study shows that (i) our proposed methodology yields consistent results for the two methods studied, in terms of the generated critical values and the performance metrics, and (ii) for most model specifications, the critical values can be estimated with a linear model with a high adjusted R^2, using factors related to the input probability distribution and the probability of an input being relevant as independent variables. Furthermore, we describe and compare the performance of the two methods studied, provide guidelines for using our methodology and the results presented in this paper, and propose suggestions for future research.

1. Introduction

Data envelopment analysis (DEA) is a mathematical programming-based methodology developed by Charnes, Cooper, and Rhodes (1978) to measure the relative efficiency of a set of n homogeneous decision-making units (DMUs) that use the same set of m inputs to produce the same set of s outputs. DEA calculates the efficiency of each DMU based on two sets of weights, one for the inputs and another for the outputs, chosen in such a way that each DMU achieves the maximum feasible efficiency.

The initial number of variables included in a DEA model is frequently very large (Wagner & Shimshak, 2007), and DEA formulation does not provide guidelines for the selection of inputs and outputs. Sexton, Silkman, and Hogan (1986) indicated that care must be taken with variable selection when conducting a DEA analysis, and presented two related principles: (i) the efficiency scores cannot decrease if we add a variable to the model, and (ii) variable selection can affect the shape and location of the efficient frontier, altering efficiency scores. Moreover, Dyson et al. (2001) noted that DEA loses its discriminatory power as the number of variables included in the model increases, i.e., many DMUs of the sample with an efficiency score equal to one and higher average efficiency of the DMUs. Variable selection methods are, therefore, valuable tools for DEA as they help to determine which variables should be included in a DEA model. Furthermore, since DEA results are affected by the number of inputs and outputs, Liu, Lu, and Lu (2016) identified variable selection as a coherent research subarea in the DEA literature, reporting 38 papers dedicated to this subject.

Many variable selection methods proposed in the literature are based on relevance measures defined by the researchers. These relevance measures can be calculated for each variable in the model to estimate their importance. Alternatively, some methods use relevance measures defined from the efficiencies estimated by DEA

* Corresponding author.
E-mail addresses: jvillanueva9@unisimonbolivar.edu.co (J. Villanueva-Cantillo), manuel.munoz@uca.es (M. Munoz-Marquez).

https://doi.org/10.1016/j.ejor.2020.08.021

as test statistics in a hypothesis test to determine whether a variable is relevant or not. However, Sirvent, Ruiz, Borraś, and Pastor (2005) noted that problems could arise in this type of hypothesis testing since the test statistics' null distribution cannot be determined analytically; hence, they suggested the use of simulation studies to analyze their performance. Some methods determine a variable to be relevant in the model if its relevance measure is greater than some threshold, but there are no guidelines for choosing appropriate thresholds based on the characteristics of the data (Nataraja & Johnson, 2011).

To address this issue, we present in this paper a methodology based on Monte Carlo simulations and bootstrapping for calculating the critical values of relevance measures to test the relevance of variables in single-output DEA models. Our methodology determines the critical values using simulated data and a hypothesis test based on a nominal size of the following type I error: Eliminating a relevant variable from the model. These critical values thus have a statistical basis and can be used as thresholds in variable selection methods based on a single relevance measure without relying on assumptions to apply certain statistical tests. The core of the proposed methodology is a data generation process that differs from those in the current literature (Adler & Yazhemsky, 2010; Jitthavech, 2016; Nataraja & Johnson, 2011; Ruggiero, 2005; Sirvent et al., 2005) since (i) it handles cases where the DEA model has multiple irrelevant inputs, as suggested by Jenkins and Anderson (2003), while the latter only considered the case with one irrelevant input in the model; (ii) it weights multiple production functions with different contributions of the inputs as their exact contribution is generally unknown in real-world applications, whereas the latter included only one production function with fixed contributions; and (iii) we can adjust the values of the parameters used in its formulation according to the characteristics of each DEA application.

Furthermore, we implemented and evaluated our proposed methodology in two existing variable selection methods through an extensive simulation study covering multiple DEA model specifications in several different scenarios. Most previous simulations studies in the literature have analyzed variable selection methods in DEA models with up to five inputs and 50 or more DMUs. However, Toloo, Barat, and Masoumzadeh (2015) and Toloo and Tichý (2015) reviewed multiple DEA applications where 75% of such applications were associated with 50 or fewer DMUs. On this basis, we increased the number of inputs to eight and limited the number of DMUs to a maximum of 50 in our simulation study, to calculate critical values for the two chosen variable selection methods using relatively common DEA model specifications. We defined the scenarios by considering multiple factors, including (i) the form of the production function, suggested by Nataraja and Johnson (2011) and Jitthavech (2016); (ii) the distribution of the inputs, also suggested by Nataraja and Johnson; (iii) the efficiency distribution of the DMUs, frequently explored in previous papers. We selected these factors to investigate the effects of varying the data characteristics on the simulation results, particularly on the critical values. Additionally, we defined three new metrics to describe and compare the performance of variable selection methods when using the critical values generated by our methodology, in addition to the traditional size and power estimates.

The rest of this paper is organized as follows: Section 2 presents our literature review and details the two variable selection methods studied in this research. In Section 3, we develop the proposed methodology and define performance metrics for the variable selection methods. Section 4 describes our simulation study of the proposed methodology for the two chosen methods. Section 5 presents a summary and discussion of the simulation results. Finally, Section 6 presents our conclusions, guidelines for using both the proposed methodology and the simulation study results, and suggestions for further research.

2. Literature review

The literature documents a wide variety of methods for addressing the problem of variable selection in DEA. We have grouped these into two classes, as described in the following subsections.

2.1. Methods strictly based on statistical or mathematical criteria

The first class contains the methods that rely only on statistical or mathematical criteria to obtain the final set of DEA model variables from an initial set.

Ueda and Hoshiai (1997) and Adler and Golany (2001, 2002) developed methods based on principal component analysis to reduce the dimensionality of the data, whereas Kao, Lu, and Chiu (2011) and Lin and Chiu (2013) used independent component analysis. Toloo et al. (2015) proposed two different mixed-integer linear programming models to reduce the number of variables to satisfy the well-known rule of thumb, $n \geq \max\{3(m+s), m \times s\}$; Toloo and Tichý (2015) modified and extended these models to make them more flexible. Fanchon (2003), Ruggiero (2005), Sharma and Yu (2015), and Li, Shi, Yang, and Liang (2017) employed regression techniques as the core of their variable selection methods. Daraio and Simar (2007) introduced a method that minimizes the sum of squared residuals regarding the original data based on linear combinations of variables. Morita and Avkiran (2009) presented a method based on factorial designs and the Mahalanobis distance. Amirteimoori, Despotis, and Kordrostami (2012) proposed an approach founded on linear combinations of highly correlated variables; Toloo and Babaee (2015) subsequently pointed out the drawbacks of this method and provided an enhanced version. Limleamthong and Guillén-Gosálbez (2018) proposed a bilevel mixed-integer programming approach for variable selection in a DEA model.

2.2. Methods based on relevance measures

The second class contains variable selection methods based on relevance measures. The most common approach is to define the relevance measure in terms of the DEA estimated efficiencies before and after eliminating a variable from the model, as in the following methods. Banker (1996) proposed three different test statistics based on the estimated efficiencies to perform two *F*-tests and one Smirnov test. Simar and Wilson (2001) also used estimated efficiencies to define six test statistics for evaluating the relevance of variables using Monte Carlo simulations and bootstrapping. Pastor, Ruiz, and Sirvent (2002) defined two parameters, the proportion p_0 of DMUs in which the relative change in their efficiencies is greater than a tolerance value φ, and used those parameters in a binomial test. Based on a simulation study, they reported $p_0 = 0.15$ and $\varphi = 1.1$ to be associated with a good performance of their method. Sirvent et al. (2005) formulated a similar method, using only the relative change in the estimated efficiencies to perform a *t*-test. In addition, Sirvent et al. (2005) evaluated their method and that proposed by Pastor et al. (2002) in multiple simulation experiments, varying the parameters of each test. Their results indicated that the estimated size and power of the tests (methods) depend on the parameters' values and the characteristics of the data. Li and Liang (2010) defined a Shapley value from the estimated efficiencies as the basis of their method to measure the relevance

Table 1
Backward stepwise approach of ACE.

Step	Description
1.	Calculate the estimated efficiencies of the DMUs by including all potential input and output variables in the DEA model.
2.	Calculate new sets of estimated efficiencies by excluding each input and then each output at a time. If the model has one input or output, then it is not possible to exclude such variables. If the model has one input and one output, then stop.
3.	Calculate the average change in the estimated efficiencies of the DMUs associated with each tested variable in *step 2*.
4.	Eliminate the variable associated with the smallest average change in the estimated efficiencies from the model and go to *step 1*.

of variables in DEA models. Madhanagopal and Chandrasekaran (2014) employed a genetic algorithm and used the parameters suggested by Pastor et al. (2002) to iteratively add variables to an initial model with one input and one output. Jitthavech (2016) used the change in the number of efficient DMUs in the model after the elimination of a variable to perform a binomial or McNemar test.

Jenkins and Anderson (2003) introduced a different approach to defining the relevance measure. Their method eliminates variables that contain minimum information based on partial covariance. However, they did not determine any threshold for the total variance proportion lost by the elimination of variables from the model.

We shall now introduce the two methods studied in this paper. First, Wagner and Shimshak (2007) proposed a method based on the average change in the estimated efficiencies of the DMUs when a variable is dropped or added to the model. We refer to this method as ACE (average change in efficiencies). Table 1 details the backward stepwise approach of ACE, i.e., dropping one variable at a time, studied in the current work. Wagner and Shimshak (2007) do not provide any threshold for the average change in efficiencies to determine whether a variable is relevant. Nevertheless, the authors suggest that: “alternative stopping rules can be developed, and in most cases would be desirable. Some possible stopping rules include: (1) when the average difference in efficiency scores exceeds some maximum level, (2) when the change in any one efficiency score exceeds some maximum level, or (3) when the number of efficient DMUs falls below some minimum number.” The current paper focuses on the first of these suggestions, i.e., calculating critical values for the average change in efficiencies.

Since our methodology is designed to calculate the critical values of single relevance measures, we chose to study ACE from among the efficiency-based variable selection methods because (i) the average change in efficiencies is related to the parameter “relative change in efficiencies” in the method proposed by Pastor et al. (2002), and (ii) the variable selection criterion in Pastor’s method was found to be straightforward to management staff (Eskelinen, 2017). Therefore, we believe our methodology provides a more straightforward criterion because ACE is based on just one parameter rather than two. Nonetheless, the results of our research on ACE can provide insights into the behavior of the relative change in efficiencies, which could contribute to extending the proposed methodology to determine critical values for the two joint parameters defined by Pastor et al. (2002)

Second, Fernandez-Palacin, Lopez-Sanchez, and Munoz-Marquez (2017) developed the *alternative DEA* method for variable selection, henceforth referred to as ADEA, based on a relevance measure called *load*. ADEA defines the loads of the variables in terms of their contribution to the global efficiency of all the DMUs and calculates the loads by solving the following linear programming

Table 2
Backward stepwise approach of ADEA.

Step	Description
1.	Calculate the estimated efficiencies of the DMUs by using the input-oriented CCR DEA model with all potential input and output variables.
2.	Calculate the loads of the variables by solving model (1).
3.	Select the variable with the smallest load as the candidate variable, eliminate it from the model if that load is not greater than a specified threshold, and go to *step 1*; otherwise, stop.

model:

$$
\begin{aligned}
\max\quad & \hat{\alpha} \\
\text{s.t:}\quad & \sum_{i=1}^{m} \nu_{ij}x_{ij} = 1, \quad \forall j = 1,\dots,n \\
& \sum_{r=1}^{s} \mu_{rj}y_{re} \le \sum_{i=1}^{m} \nu_{ij}x_{ie}, \quad \forall e = 1,\dots,n, \forall j = 1,\dots,n \\
& \sum_{r=1}^{s} \mu_{rj}y_{rj} = h_j^*, \quad \forall j = 1,\dots,n \\
& \hat{\alpha}_i^{\text{in}} = \frac{m}{n}\sum_{j=1}^{n} \nu_{ij}x_{ij}, \quad \forall i = 1,\dots,m \\
& \hat{\alpha}_r^{\text{out}} = \frac{s}{h}\sum_{j=1}^{n} \mu_{ij}y_{ij}, \quad \forall r = 1,\dots,s \\
& \hat{\alpha}_i^{\text{in}} \ge \hat{\alpha}, \quad \forall i = 1,\dots,m \\
& \hat{\alpha}_r^{\text{out}} \ge \hat{\alpha}, \quad \forall r = 1,\dots,s \\
& \hat{\alpha}, \hat{\alpha}_i^{\text{in}}, \hat{\alpha}_r^{\text{out}} \ge 0, \quad \forall i = 1,\dots,m, \forall r = 1,\dots,s \\
& \nu_{ij}, \mu_{rj} \ge 0, \quad \forall i = 1,\dots,m, \forall r = 1,\dots,s, \forall j = 1,\dots,n,
\end{aligned}
\tag{1}
$$

where h_j^* is the estimated efficiency of the jth DMU by the input-oriented CCR DEA model (Charnes et al., 1978) and $h = \sum_{j=1}^{n} h_j^*$. The $\hat{\alpha}$ variables are called $\hat{\alpha}$-ratios, and the load of each input and output is given by the corresponding value of $\hat{\alpha}^{\text{in}}$ and $\hat{\alpha}^{\text{out}}$, respectively, in the solution of this problem. The value of $\hat{\alpha}^{\text{in}}$ is the proportion corresponding to each input of the sum of the virtual inputs associated with the efficiencies of the DMUs, multiplied by the number of inputs in the model. The value of $\hat{\alpha}^{\text{out}}$ is defined in a similar manner to $\hat{\alpha}^{\text{in}}$ for the outputs.

ADEA maximizes the minimum $\hat{\alpha}$-ratio of the variables in the model while preserving the initially estimated efficiencies of the DMUs with those variables. Consequently, for a given initial solution of model (1), ADEA increases the smallest $\hat{\alpha}$-ratio until it is equal to the second smallest $\hat{\alpha}$-ratio (if feasible); then, ADEA simultaneously increases these two $\hat{\alpha}$-ratios until they are equal to the third smallest $\hat{\alpha}$-ratio (if feasible) and so forth. It is worth noting that the sum of the $\hat{\alpha}^{\text{in}}$-ratios (resp. $\hat{\alpha}^{\text{out}}$-ratios) is equal to the number of inputs (resp. outputs) in the model; therefore, the loads of all the inputs or outputs in the model will be equal to one if they have the same relevance to ADEA.

We chose to study ADEA because it evaluates variables’ relevance based on their weights in DEA in one single linear programming model rather than on the estimated efficiencies of two nested DEA models, i.e., before and after the elimination of a variable. Moreover, Sirvent et al. (2005) highlighted two problems with these two sets of estimated efficiencies: They are not independent, and the estimated bias differs between them (being smaller for the DEA model with fewer variables).

ADEA uses a similar approach to ACE for variable selection, as described in Table 2. Fernandez-Palacin et al. (2017) applied ADEA in two real-world problems using Monte Carlo simulations to calculate the load thresholds as quantiles of the load distribution of a dummy variable included in the DEA model. Based on the results of these simulations, they suggest using 0.6 as a threshold for the application of ADEA.

Regarding variable selection method performance, many authors have used Monte Carlo simulations for performance evaluation. Similarly, Sirvent et al. (2005), Adler and Yazhemsky (2010), Nataraja and Johnson (2011), Kao et al. (2011), Jitthavech (2016), and Eskelinen (2017) compared sets of methods, reporting their strengths and weaknesses. The most common performance metrics are the estimated size and power of the method, and the correlation and deviation of the DEA estimated efficiencies from the theoretical efficiencies.

3. Proposed methodology

In this section, we introduce our proposed methodology for calculating critical values of single relevance measures in variable selection methods in DEA. These critical values can be used as thresholds in such methods to perform a backward stepwise algorithm by calculating for each step the relevance measure of each variable in the DEA model, and eliminating the variable with the lowest relevance measure (i.e., the candidate variable) if its value does not exceed the critical value.

We developed the proposed methodology for the single-output case in DEA, given that we defined and included several elements in its formulation and that the single-output case is the most widely studied in the literature. Nevertheless, we believe that most of the elements and results of the current work will be relevant in extending the methodology to the multi-output case. Consequently, we model the output y of a DMU as

$$y = f(\boldsymbol{x}) \cdot e^{-\tau}, \tag{2}$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)$ is the input vector, f is the production function that transforms the inputs into the output, and $e^{-\tau}$ is the efficiency term. Eq. (2) is widely used in the literature (See Banker, 1996; Nataraja & Johnson, 2011; Simar & Wilson, 2001; Sirvent et al., 2005), where $\tau \geq 0$ is called the *inefficiency parameter* and is equal to 0 if a DMU is efficient. For notation purposes, let $i \in \mathcal{I} = \{1, \ldots, m\}$.

3.1. Critical values calculation

Since our focus is the single-output case in DEA, the candidate variable is the input with the lowest relevance measure in the model. Accordingly, we formulate the following problem for hypothesis testing:

$$\begin{cases} H_0: & \text{The candidate input is relevant,} \\ H_1: & \text{The candidate input is irrelevant.} \end{cases} \tag{3}$$

For this hypothesis test, the test statistic θ is the relevance measure of the candidate input. Then, for a given nominal size α of type I error, we seek a critical value θ_α such that

$$P(\theta \leq \theta_\alpha | H_0 \text{ is true}) = \alpha. \tag{4}$$

To compute the critical value θ_α, we use Monte Carlo simulations to perform a data generation process (DGP) to obtain a simulated null distribution $\mathcal{S}$ of θ. Then, we can calculate θ_α as the α-quantile of $\mathcal{S}$. Let T be the number of simulation trials in the DGP and $\mathcal{T} = \{1, \ldots, T\}$.

Before performing a DGP for the hypothesis test in Eq. (3), we must first address two concerns: The first arises from the need to define the condition under which an input is relevant in the model. For simplicity's sake, we assume that the production function associates a unique parameter with each input, given that an input will theoretically be relevant if its respective parameter is not zero. An example of this type of function is the following Cobb-Douglas production function:

$$y = \prod_{i=1}^{m} x^{\lambda_i} \cdot e^{-\tau}, \tag{5}$$

where $\lambda_i \geq 0 \forall i \in \mathcal{I}$. Consequently, we introduce the *input parameters vector* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$, with $\sum_{i=1}^{m} \lambda_i = L$, $0 < L < \infty$ and $\lambda_i \geq 0 \forall i \in \mathcal{I}$, so that if any $\lambda_i > 0$, then the respective input x_i is relevant. However, since the effect of input variations on the output value can be imperceptible to variable selection methods for small λ values, we consider it necessary to select a non-zero minimum value for the λ parameters so that an input is considered relevant.

Definition 1. Let $\lambda_{\text{rel}} > 0$ be the *minimum relevant parameter value* that an input must have in a production function to be considered relevant and $m_r = m_r(\boldsymbol{\lambda}, \lambda_{\text{rel}}) = |\{i \in \mathcal{I} : \lambda_i \in \boldsymbol{\lambda}, \lambda_i \geq \lambda_{\text{rel}}\}|$ be the *number of relevant inputs* in a production function.

Therefore, based on Definition 1, and denoting the parameter λ of the candidate input as λ_{cand}, we will not reject H_0 if $\lambda_{\text{cand}} \geq \lambda_{\text{rel}}$ as λ is an increasing relevance measure. The selection of λ_{rel} depends on the effect size we wish to detect. For instance, consider the production function $y = x_1^{\lambda_1} x_2^{\lambda_2} e^{-\tau}$. If $\lambda_1 = 0.01$, then doubling the x_1 value will increase y by only 0.7%. In contrast, if $\lambda_1 = 0.05$ or 0.1, then the increase will be 3.5% or 7.2%, respectively, and these variations can be more perceptible to variable selection methods. Thus, we used $\lambda_{\text{rel}} = 0.1$ as the minimum relevant parameter value in this paper.

The second consideration is that we need observations of θ when the candidate input is relevant in order to construct the simulated null distribution $\mathcal{S}$. This is because the candidate input is determined by the variable selection method's relevance measure rather than the actual relevance (λ parameters) of the inputs in the production function. One possible approach to guarantee the relevance of candidate input in all simulation trials, independently of the method, is to use production functions with only relevant inputs. If we obtain $\mathcal{S}$ using this approach, we assume the candidate input is relevant only when all the model's inputs are relevant. However, since the relevance measure θ is not necessarily a strictly increasing function of λ, the previous assumption is not necessarily true, because a relevant input may still be selected as the candidate input even when there are irrelevant inputs in the model. Consequently, an appropriate DGP for the hypothesis test in Eq. (3) must satisfy the following conditions:

- C.1 *Inclusion of production functions with different numbers of relevant inputs.*
- C.2 *Utilization of different values for the non-zero input parameters in the production functions.*

The arguments supporting condition C.1 are: (i) we do not know how many inputs in the initial set are relevant, and (ii) we must examine scenarios where the candidate input may be relevant when there are irrelevant inputs in the model, as described above. In contrast, the supporting arguments for condition C.2 are: (i) we do not know the actual relevance of each relevant input, (ii) the effects of λ parameters of different magnitudes are likely to be different, and (iii) the effect of a specific λ value can be different in distinct $\boldsymbol{\lambda}$ vectors. The implementation of these conditions in the DGP is described in the next subsection.

A consequence of using a DGP that satisfies conditions C.1 and C.2 is that we obtain observations of θ for both relevant (H_0) and irrelevant (H_1) candidate inputs. Nevertheless, we can track which case is associated with each observation by storing both θ and λ_{cand} in each simulation trial, i.e., collecting a sample $S = \{(\theta^{(t)}, \lambda_{\text{cand}}^{(t)})\}_{t=1}^{T}$ from the DGP. Then, it is straightforward that $\mathcal{S} = \{\theta^{(t)} : (\theta^{(t)}, \lambda_{\text{cand}}^{(t)}) \in S, \lambda_{\text{cand}}^{(t)} \geq \lambda_{\text{rel}}, t \in \mathcal{T}\}$.

After conducting the DGP, we use bootstrapping to obtain $B = 1000$ pseudo-samples $\hat{S}_b$, $b = 1, \dots, B$, from S to find the critical value θ_α^b of the $\hat{S}_b$ empirical distribution associated with each pseudo-sample $\hat{S}_b$. Finally, we calculate the critical value θ_α as the mean of the pseudo-sample critical values, i.e.,

$$\theta_\alpha = \frac{1}{B} \sum_{b=1}^{B} \theta_\alpha^b. \tag{6}$$

3.2. Data generation process

In the literature, we observed the DGPs performed for single-output DEA models to have a common structure (Adler & Yazhemsky, 2010; Jitthavech, 2016; Nataraja & Johnson, 2011; Ruggiero, 2005; Sirvent et al., 2005). Therefore, we define a *standard single-output DGP* with such a structure as follows: first, define a production function to be used in all the simulation trials. Then, for each trial, perform the following steps: (i) randomly generate the inputs from a probability distribution F, (ii) randomly assign an efficiency equal to 1 (i.e., $\tau = 0$) to a ϕ proportion of the DMUs, (iii) calculate the efficiencies of the other DMUs by generating their inefficiency parameter τ from a probability distribution H, and (iv) calculate the output of each DMU using the predefined production function.

We must modify the standard DGP to implement conditions C.1 and C.2 defined in Section 3.1 since we need to include multiple production functions in the DGP simulation trials. These conditions are implemented as follows:

Definition 2. Let $Y = \{y_1, \dots, y_T\}$ be the set of production functions assigned to the T simulation trials of the DGP.

To implement condition C.1, we must define the percentage of functions in the set Y with $m_r = 1, \dots, m$ relevant inputs. For this purpose, we assume that one of the inputs in the model is relevant and that each of the other inputs has a fixed probability p_{rel} of being relevant. Therefore, the proportion of functions in Y with $m_r \in \mathcal{I}$ relevant inputs is determined by the binomial distribution in the following definition.

Definition 3. The proportion $\pi(m_r)$ of functions with m_r relevant inputs in Y is given by

$$\pi(m_r) = \binom{m-1}{m_r-1} p_{\text{rel}}^{m_r-1} (1 - p_{\text{rel}})^{m-m_r}, \tag{7}$$

where $m_r \in \mathcal{I}$, $0 < p_{\text{rel}} < 1$ and $m \in \mathbb{N} \setminus \{1\}$.

The selection of a suitable value for p_{rel} may depend on the real-world DEA application. In some cases, the initial set of variables may be chosen well, making the value of p_{rel} relatively high. In other cases, this initial set may contain several irrelevant variables, such that p_{rel} would be relatively low. Taking this into account, we included p_{rel} as a factor in the simulation study in Section 4 and selected $p_{\text{rel}} = 0.4$, 0.6, 0.8, and 0.9 to test multiple values for the probability of an input being relevant in a DEA model.

Furthermore, a consequence of condition C.1 is the need to define a set that contains functions with different numbers of relevant inputs, from which we can obtain Y by sampling with replacement.

Definition 4. Let $\hat{Y} = \{\hat{y}_1, \dots, \hat{y}_Q\}$ be the set of distinct production functions to be sampled with replacement to obtain the set Y, $\hat{\boldsymbol{p}} = (\hat{p}_1, \dots, \hat{p}_Q)$ be the probability vector to sample from $\hat{Y}$ and $q \in \mathcal{Q} = \{1, \dots, Q\}$.

Definition 5. For all $m_r \in \mathcal{I}$, let $\Omega(m_r)$ be the subset of $\hat{Y}$ that contains the functions with m_r relevant inputs.

Table 3
$\pi(m_r)$ values for $m = 4$ and $p_{\text{rel}} = 0.8$.

m_r	1	2	3	4
$\pi(m_r)$	$\frac{1}{125}$	$\frac{12}{125}$	$\frac{48}{125}$	$\frac{64}{125}$

Definition 6. For all $m_r \in \mathcal{I}$, let $\mathcal{Q}(m_r) = \{q \in \mathcal{Q} : y_q \in \Omega(m_r)\}$ be the index set of the functions in $\Omega(m_r)$.

According to condition C.2, the set $\hat{Y}$ must contain production functions with different input parameters vectors ($\boldsymbol{\lambda}$) for each possible number of relevant inputs. However, since the only constraint on the elements of $\boldsymbol{\lambda}$ is to have a positive finite sum L, and we do not know the actual relevance of each relevant input, there are infinite options for choosing these vectors. To deal with this situation, we limit the options to those formed with zeros and multiples of λ_{rel}. Moreover, since the inputs are generated from the same probability distribution in our DGP, the production functions with input parameter vectors that differ only in the order of their elements are expected to yield the same results. Therefore, we construct the set $\hat{Y}$ using only the $\boldsymbol{\lambda}$ vectors that have their elements arranged in ascending order and assign them a weight w equal to the number of possible permutations with repetition with their elements. For instance, the weight of the production function $y = x_1^{0.1} x_2^{0.2} x_3^{0.2} x_4^{0.5} e^{-\tau}$ is 12, since that is the number of different production functions that can be formed with the permutations of the vector (0.1,0.2,0.2,0.5). Based on Definitions 3–5, these weights are necessary to calculate the vector $\hat{\boldsymbol{p}}$ by appropriately distributing the probability $\pi(m_r)$ among the functions in the subset $\Omega(m_r)$ for all $m_r \in \mathcal{I}$. Consequently, we define the probability $\hat{p}_q \in \hat{\boldsymbol{p}}$ associated with the function $\hat{y}_q \in \hat{Y}$ as

$$\hat{p}_q = \frac{w_q}{\sum_{q \in \mathcal{Q}(m_r)} w_q} \times \pi(m_r), \tag{8}$$

where w_q, m_r and $\Omega(m_r)$ are the weight, the number of relevant inputs, and the subset of $\hat{y}_q$, respectively.

For illustration purposes, we present the set $\hat{Y}$ and the vector $\hat{\boldsymbol{p}}$ for $m = 4$ inputs, $\lambda_{\text{rel}} = 0.1$, $p_{\text{rel}} = 0.8$, and a Cobb-Douglas production function with constant returns to scale ($\sum_{i=1}^{m} \lambda_i = 1$) in Tables 3 and 4. In more detail, Table 3 indicates that for $p_{\text{rel}} = 0.8$ we assume the probabilities of having $m_r = 1, 2, 3$, or 4 relevant inputs in a given 4-input DEA model are 0.8%, 9.6%, 38.4%, and 51.2%, respectively. Then, these probabilities are distributed among the production functions in the respective subsets $\Omega(m_r)$, as shown in Table 4.

In summary, we generate the set Y by sampling with replacement from the set $\hat{Y}$ using the vector $\hat{\boldsymbol{p}}$ in the first stage of our DGP. Then, for each simulation trial $t = 1, \dots, T$, we generate the inputs and efficiencies of the DMUs following steps (i)-(iii) in the standard DGP and calculate the output of the DMUs using the assigned production function y_t.

3.3. Performance metrics

We define the following metrics to describe and compare the variable selection methods' performance when using the critical values generated by our methodology. To calculate each metric, we use bootstrapping one more time to obtain $D = 1000$ new pseudo-samples $\tilde{S}_d$ $(d = 1, \dots, D)$ from S. Then, we determine the value of the metric for each $\tilde{S}_d$. Finally, we set the mean of these pseudo-sample metrics as the final value of the performance metric.

Definition 7. The *estimated type I error rate* A_α associated with the critical value θ_α is given by

$$A_\alpha = \frac{1}{D} \sum_{d=1}^{D} \tilde{P}_d(\theta \leq \theta_\alpha | \lambda_{\text{cand}} \geq \lambda_{\text{rel}}). \tag{9}$$

Table 4
Set $\hat{Y}$ and vector $\hat{\boldsymbol{p}}$ for $m = 4$ inputs, $\lambda_{\text{rel}} = 0.1$, $p_{\text{rel}} = 0.8$ and Cobb-Douglas production function with constant returns to scale.

$\hat{y}_q$	w_q	$\hat{p}_q$
$\hat{y}_1 = x_1^{0.0}x_2^{0.0}x_3^{0.0}x_4^{1.0}e^{-\tau}$	4	$\frac{4}{4} \times \frac{1}{125} = 0.0080$
$\sum_{q\in\mathcal{Q}(1)} w_q =$	4	
$\hat{y}_2 = x_1^{0.0}x_2^{0.0}x_3^{0.1}x_4^{0.9}e^{-\tau}$	12	$\frac{12}{54} \times \frac{12}{125} = 0.0213$
$\hat{y}_3 = x_1^{0.0}x_2^{0.0}x_3^{0.2}x_4^{0.8}e^{-\tau}$	12	$\frac{12}{54} \times \frac{12}{125} = 0.0213$
$\hat{y}_4 = x_1^{0.0}x_2^{0.0}x_3^{0.3}x_4^{0.7}e^{-\tau}$	12	$\frac{12}{54} \times \frac{12}{125} = 0.0213$
$\hat{y}_5 = x_1^{0.0}x_2^{0.0}x_3^{0.4}x_4^{0.6}e^{-\tau}$	12	$\frac{12}{54} \times \frac{12}{125} = 0.0213$
$\hat{y}_6 = x_1^{0.0}x_2^{0.0}x_3^{0.5}x_4^{0.5}e^{-\tau}$	6	$\frac{6}{54} \times \frac{12}{125} = 0.0107$
$\sum_{q\in\mathcal{Q}(2)} w_q =$	54	
$\hat{y}_7 = x_1^{0.0}x_2^{0.1}x_3^{0.1}x_4^{0.8}e^{-\tau}$	12	$\frac{12}{144} \times \frac{48}{125} = 0.0320$
$\hat{y}_8 = x_1^{0.0}x_2^{0.1}x_3^{0.2}x_4^{0.7}e^{-\tau}$	24	$\frac{24}{144} \times \frac{48}{125} = 0.0640$
$\hat{y}_9 = x_1^{0.0}x_2^{0.1}x_3^{0.3}x_4^{0.6}e^{-\tau}$	24	$\frac{24}{144} \times \frac{48}{125} = 0.0640$
$\hat{y}_{10} = x_1^{0.0}x_2^{0.1}x_3^{0.4}x_4^{0.5}e^{-\tau}$	24	$\frac{24}{144} \times \frac{48}{125} = 0.0640$
$\hat{y}_{11} = x_1^{0.0}x_2^{0.2}x_3^{0.2}x_4^{0.6}e^{-\tau}$	12	$\frac{12}{144} \times \frac{48}{125} = 0.0320$
$\hat{y}_{12} = x_1^{0.0}x_2^{0.2}x_3^{0.3}x_4^{0.5}e^{-\tau}$	24	$\frac{24}{144} \times \frac{48}{125} = 0.0640$
$\hat{y}_{13} = x_1^{0.0}x_2^{0.2}x_3^{0.4}x_4^{0.4}e^{-\tau}$	12	$\frac{12}{144} \times \frac{48}{125} = 0.0320$
$\hat{y}_{14} = x_1^{0.0}x_2^{0.3}x_3^{0.3}x_4^{0.4}e^{-\tau}$	12	$\frac{12}{144} \times \frac{48}{125} = 0.0320$
$\sum_{q\in\mathcal{Q}(3)} w_q =$	144	
$\hat{y}_{15} = x_1^{0.1}x_2^{0.1}x_3^{0.1}x_4^{0.7}e^{-\tau}$	4	$\frac{4}{84} \times \frac{64}{125} = 0.0244$
$\hat{y}_{16} = x_1^{0.1}x_2^{0.1}x_3^{0.2}x_4^{0.6}e^{-\tau}$	4	$\frac{4}{84} \times \frac{64}{125} = 0.0244$
$\hat{y}_{17} = x_1^{0.1}x_2^{0.1}x_3^{0.3}x_4^{0.5}e^{-\tau}$	12	$\frac{12}{84} \times \frac{64}{125} = 0.0731$
$\hat{y}_{18} = x_1^{0.1}x_2^{0.1}x_3^{0.4}x_4^{0.4}e^{-\tau}$	6	$\frac{6}{84} \times \frac{64}{125} = 0.0366$
$\hat{y}_{19} = x_1^{0.1}x_2^{0.2}x_3^{0.2}x_4^{0.5}e^{-\tau}$	12	$\frac{12}{84} \times \frac{64}{125} = 0.0731$
$\hat{y}_{20} = x_1^{0.1}x_2^{0.2}x_3^{0.3}x_4^{0.4}e^{-\tau}$	24	$\frac{24}{84} \times \frac{64}{125} = 0.1463$
$\hat{y}_{21} = x_1^{0.1}x_2^{0.3}x_3^{0.3}x_4^{0.3}e^{-\tau}$	4	$\frac{4}{84} \times \frac{64}{125} = 0.0244$
$\hat{y}_{22} = x_1^{0.2}x_2^{0.2}x_3^{0.2}x_4^{0.4}e^{-\tau}$	4	$\frac{4}{84} \times \frac{64}{125} = 0.0244$
$\hat{y}_{23} = x_1^{0.2}x_2^{0.2}x_3^{0.3}x_4^{0.3}e^{-\tau}$	6	$\frac{6}{84} \times \frac{64}{125} = 0.0244$
$\sum_{q\in\mathcal{Q}(4)} w_q =$	84	

Definition 8. The *estimated power* P_α associated with the critical value θ_α is given by

$$P_\alpha = \frac{1}{D}\sum_{d=1}^{D}\hat{P}_d(\theta \le \theta_\alpha | \lambda_{\text{cand}} < \lambda_{\text{rel}}). \tag{10}$$

Definition 9. The *strong type I error rate* B_α associated with the critical value θ_α is given by

$$B_\alpha = \frac{1}{D}\sum_{d=1}^{D}\hat{P}_d(\theta \le \theta_\alpha \cap \lambda_{\text{cand}} > \lambda_{\text{rel}} | \lambda_{\text{cand}} \ge \lambda_{\text{rel}}). \tag{11}$$

B_α is the part of A_α corresponding with cases where an input with a parameter greater than λ_{rel} was dropped from the model. The ideal value of B_α is 0 since the type I error should ideally be associated with the elimination of inputs with the minimum relevant parameter value λ_{rel}.

Definition 10. The *tie-related type I error rate* W_α associated with the critical value θ_α is given by

$$W_\alpha = \frac{1}{D}\sum_{d=1}^{D}\hat{P}_d(\theta \le \theta_\alpha \cap \theta_{\text{rel}} = \theta_{\text{irrel}} | \lambda_{\text{cand}} \ge \lambda_{\text{rel}}), \tag{12}$$

where θ_{rel} and θ_{irrel} are the lowest relevance measures of the relevant and irrelevant inputs in the model, respectively.

W_α is the part of A_α that corresponds to the cases where a tie in θ occurred between relevant and irrelevant inputs. We randomly select the candidate input from the tied inputs.

Definition 11. The *type I error impact* $\boldsymbol{\Delta_\alpha}$ associated with the critical value θ_α is given by

$$\Delta_\alpha = \frac{1}{D}\sum_{d=1}^{D}\left(\frac{\sum_{t\in\mathcal{T}_d}\left(r_t^a - r_t^b\right)}{|\mathcal{T}_d|}\right) \tag{13}$$

where r^b and r^a are the Pearson correlation coefficients between the theoretical and DEA estimated efficiencies of the DMUs before and after eliminating the candidate input, respectively, and $\mathcal{T}_d = \{t \in \mathcal{T} : (\theta^{(t)}, \lambda_{\text{cand}}^{(t)}) \in \hat{S}_d, \theta^{(t)} \le \theta_\alpha, \lambda_{\text{cand}}^{(t)} \ge \lambda_{\text{rel}}\}$ is the index set of the simulation trials where a type I error is committed using the critical value θ_α in the pseudo-sample $\tilde{S}_d$.

The purpose of calculating Δ_α is to measure the impact of type I error on the correlation between the theoretical and estimated efficiencies of the DMUs.

4. Simulation study

A simulation study was performed to investigate the results of implementing the proposed methodology to calculate critical values for ACE and ADEA methods in different model specifications using multiple different DGP configurations. Specifically, for each method, we conducted 28 experiments corresponding to all possible model specifications from $m = 2, \ldots, 8$ inputs and $n = 20$, 30, 40, and 50 DMUs, considering that many DEA applications are associated with small sample sizes, i.e., $n \le 50$ (see Toloo et al., 2015; Toloo & Tichý, 2015). All the experiments consisted of 256 different scenarios defined by the four levels of p_{rel} and six additional factors with two levels each, and five replicates of $T = 1000$ simulation trials in each scenario. This meant that $2 \times 28 \times 256 \times 5 \times 1000 = 71{,}680{,}000$ data sets were generated.

Furthermore, we considered $\alpha = 5\%$ and $\alpha = 10\%$ as nominal sizes, and calculated the critical values and performance metrics of the methods for such values. Under the assumption that the production functions exhibit constant returns to scale (CRS), we used the input-oriented CCR DEA model (Charnes et al., 1978) to estimate the DMUs' efficiencies. The simulation was implemented in *R* (Team, 2017), using a package provided by Fernandez-Palacin et al. (2017) to perform the calculations needed for the ADEA method.

Next, we shall detail the other six factors used to generate the scenarios in each experiment. The factor levels were selected to investigate their effect on the critical values generated by our methodology. The first factor is the functional form f of the production function. Since each input must have only one parameter in f, we used the Cobb–Douglas (CD) function, i.e.,

$$f^{CD}(\boldsymbol{x}) = \prod_{i=1}^{m} x_i^{\lambda_i}, \tag{14}$$

and the constant elasticity of substitution (CES) function, i.e.,

$$f^{CES}(\boldsymbol{x}) = \left(\sum_{i=1}^{m}\lambda_i x_i^{\rho}\right)^{1/\rho}, \tag{15}$$

with $\rho = -1/9$. We imposed the constraint $\sum_{i=1}^{m}\lambda_i = 1$ on the input parameters due to the CRS assumption. We selected these functional forms to examine the effect of changing the elasticity of substitution of the inputs from 1 (CD function) to 0.9 (CES function).

The second and third factors are the shape and spread of the probability distribution of the inputs F, respectively. For the shape, we used uniform and lognormal distributions because (i) the former has negative excess kurtosis and the latter has positive excess kurtosis, and (ii) we assume that many real-world variable distributions will have identical or similar shapes. For the spread, we employed the coefficient of variation (CV) as a metric and used two values, 29% and 51%, to represent input sets with low and high variation, respectively. We selected the parameters of the F distributions used in the simulation study so that they all have a mean of 40. Table 5 shows the values of those parameters.

The fourth factor is the proportion ϕ of efficient DMUs in the sample; we used $\phi = 0.20$ and $\phi = 0.30$, as per Holland and Lee (2002) and Jitthavech (2016), respectively. The fifth factor is the shape of the inefficiency parameter distribution H for which we used exponential, $\text{Exp}(\lambda)$, and half-normal, $|N(0, \sigma^2)|$, distributions,

Table 5
Parameters of the F distributions used in the simulation study.

CV	Uniform		Lognormal	
	a	b	μ	σ
29%	20	60	3.6488	0.2829
51%	5	75	3.5752	0.4768

Table 6
Parameters of the H distributions used in the simulation study.

Mean efficiency	Exponential	Half-normal
0.80	$\lambda = 0.25$	$\sigma = 0.3$
0.85	$\lambda = 0.1765$	$\sigma = 0.2137$

Table 7
Coded factors levels in simulation study.

Code	f_{form}	F_{shape}	F_{spread}	ϕ	H_{shape}	H_{mean}
−1	CES	Lognormal	29%	0.20	Exponential	0.80
+1	CD	Uniform	51%	0.30	Half-normal	0.85

following the trend in the literature. Finally, the sixth factor is the theoretical mean efficiency of the non-efficient DMUs (H_{mean}). Here we selected the parameters for each H shape to obtain two levels of mean efficiency, 0.80 and 0.85, as in the works developed by Perelman and Santín (2009) and Nataraja and Johnson (2011), respectively. Table 6 shows the parameters of the H distributions used in the simulation study, and Table 7 shows the coded levels of the six factors described above.

5. Results and discussion

In this section, we present and discuss the results obtained from the simulation study. We focus on the results corresponding to $\alpha = 5\%$ nominal size as the same conclusions can be drawn for $\alpha = 10\%$. (Results for $\alpha = 10\%$ are reported in Appendix A.)

5.1. Critical values analysis

For each variable selection method, we calculated an ANOVA model with the critical value as the dependent variable and the following factors: Number of inputs, number of DMUs, probability of an input being relevant, and the six factors defined in Section 4. ANOVAs for ACE and ADEA methods are presented in Tables 8 and 9, respectively. The aim of these ANOVA models was to determine the significance and effect size of each factor on the critical values.

We can see in Tables 8 and 9 that all the factors were statistically significant at the 1% level. Nevertheless, we considered it necessary to evaluate the effect size of the factors on the critical values to evaluate their practical significance. For this purpose, we used the eta-squared effect size, i.e., $\eta^2 = \frac{\text{sum of squares of the factor}}{\text{total sum of squares}}$. The number of inputs was notably the factor with the greatest effect size, approximately $\eta^2 = 0.8$, for both ACE and ADEA. The other factors with a considerable effect on the critical values of both methods were the number of DMUs, and the shape and spread of the input distribution. Additionally, the factor related to the probability of an input being relevant had a noticeable effect on ADEA critical values ($\eta^2 = 0.1$). Taking this into account, we constructed a linear model of the critical values for each method and combination of the number of inputs and DMUs in terms of the other aforementioned factors. The forms of these linear models are shown in Eqs. (16) and (17) for ACE and ADEA, respectively, and the corresponding coefficients are shown in Table 10.

$$\hat{\theta}_{\alpha}^{\text{ACE}} = b_0 + b_1 F_{\text{shape}} + b_2 F_{\text{spread}} + \epsilon \tag{16}$$

$$\hat{\theta}_{\alpha}^{\text{ADEA}} = b_0 + b_1 F_{\text{shape}} + b_2 F_{\text{spread}} + b_3 p_{\text{rel}} + \epsilon \tag{17}$$

We highlight the following key points from the results in Table 10.

(i) All factors were statistically significant at the 5% level in all the linear models, except for F_{shape} and F_{spread} in some experiments in which their coefficient was zero.
(ii) The linear models had a high adjusted R^2 for ACE in the experiments with six inputs or less, except those with six inputs and 20 DMUs, and for ADEA in the experiments with three or more inputs. This indicates that we can use the linear models to estimate the critical values in terms of F_{shape}, F_{spread}, and p_{rel} in most of the DEA model specifications used in this paper.
(iii) In the experiments for ACE with 7–8 inputs and 20 DMUs, the intercepts and coefficients were zero or very close to zero; therefore, the critical values can be assumed to be zero for these experiments. Moreover, in the experiments for ADEA with 2 inputs, the critical values were equal to the intercept from a practical point of view, as the factor coefficients were very small compared to the intercept.
(iv) ACE critical values asymptotically decreased to zero for more inputs in the DEA model and increased with more DMUs. This behavior is reasonable since the average change in the estimated efficiencies when we remove a variable from the DEA model is expected to be smaller in higher dimensional spaces (with more variables) and larger in more constrained DEA models (with more DMUs).
(v) ADEA critical values asymptotically increased to one with more inputs in the DEA model and decreased with more DMUs. This tendency is also reasonable since the load of a variable is based on its weights in the solution of model (1), and having more variables (inputs) allows the smallest load to be bigger, while having more constraints (DMUs) limits the maximization of that smallest load.

Table 8
ANOVA of ACE critical values at $\alpha = 5\%$.

Source of variation	Sum of squares	df	Mean square	F	p-value
Number of inputs	16.8236	6	2.8039	38875.27	0.0000
Number of DMUs	0.4855	3	0.1618	2243.58	0.0000
p_{rel}	0.0009	3	0.0003	4.30	0.0049
f_{form}	0.0008	1	0.0008	10.66	0.0011
F_{shape}	0.1937	1	0.1937	2685.31	0.0000
F_{spread}	1.0046	1	1.0046	13929.03	0.0000
ϕ	0.0161	1	0.0161	223.69	0.0000
H_{shape}	0.0007	1	0.0007	9.96	0.0016
H_{mean}	0.0098	1	0.0098	135.38	0.0000
Residuals	2.5836	35821	0.0000721		
Total	21.1193	35839			

Table 9
ANOVA for ADEA critical values at $\alpha = 5\%$.

Source of variation	Sum of squares	df	Mean square	F	p-value
Number of inputs	1476.21	6	246.04	114548.47	0.0000
Number of DMUs	28.01	3	9.34	4347.09	0.0000
p_{rel}	183.40	3	61.13	28463.01	0.0000
f_{form}	0.05	1	0.05	25.52	0.0000
F_{shape}	16.64	1	16.64	7748.49	0.0000
F_{spread}	16.13	1	16.13	7511.33	0.0000
ϕ	0.29	1	0.29	135.64	0.0000
H_{shape}	0.08	1	0.08	36.82	0.0000
H_{mean}	0.38	1	0.38	178.74	0.0000
Residuals	76.94	35821	0.00215		
Total	1798.13	35839			

Table 10
Coefficients of linear models for ACE and ADEA critical values at $\alpha = 5\%$.

Inputs	DMUs	*ACE method*				*ADEA method*				
		Intercept	F_{shape}	F_{spread}	R^2_{adj}	Intercept	F_{shape}	F_{spread}	p_{rel}	R^2_{adj}
2	20	0.0555***	0.0071***	0.0176***	0.89	0.295***	0.002***	0.002***	−0.003**	0.18
	30	0.0639***	0.0083***	0.0201***	0.89	0.264***	0.001***	0.002***	−0.001*	0.17
	40	0.0692***	0.0090***	0.0217***	0.89	0.250***	0.001***	0.002***	−0.002***	0.24
	50	0.0735***	0.0098***	0.0231***	0.89	0.239***	0.001***	0.001***	−0.001**	0.26
3	20	0.0213***	0.0032***	0.0070***	0.88	0.230***	0.011***	0.005***	0.299***	0.90
	30	0.0284***	0.0036***	0.0087***	0.91	0.223***	0.008***	0.003***	0.271***	0.87
	40	0.0331***	0.0037***	0.0098***	0.92	0.272***	0.009***	0.008***	0.195***	0.77
	50	0.0364***	0.0037***	0.0106***	0.93	0.347***	0.010***	0.012***	0.088***	0.65
4	20	0.0077***	0.0015***	0.0024***	0.80	0.308***	0.024***	0.022***	0.388***	0.94
	30	0.0137***	0.0023***	0.0042***	0.87	0.272***	0.021***	0.019***	0.367***	0.94
	40	0.0181***	0.0026***	0.0054***	0.90	0.267***	0.018***	0.015***	0.326***	0.92
	50	0.0214***	0.0028***	0.0062***	0.91	0.272***	0.017***	0.012***	0.288***	0.91
5	20	0.0022***	0.0004***	0.0004***	0.60	0.370***	0.031***	0.030***	0.472***	0.95
	30	0.0062***	0.0011***	0.0017***	0.81	0.314***	0.029***	0.027***	0.482***	0.96
	40	0.0095***	0.0016***	0.0026***	0.85	0.290***	0.028***	0.024***	0.468***	0.96
	50	0.0123***	0.0019***	0.0035***	0.88	0.279***	0.026***	0.022***	0.448***	0.95
6	20	0.0002***	0***	−0.0001***	0.34	0.442***	0.032***	0.036***	0.502***	0.96
	30	0.0025***	0.0003***	0.0003***	0.53	0.377***	0.032***	0.033***	0.539***	0.97
	40	0.0047***	0.0007***	0.0009***	0.75	0.339***	0.033***	0.031***	0.548***	0.97
	50	0.0068***	0.0010***	0.0016***	0.81	0.316***	0.033***	0.029***	0.546***	0.97
7	20	0***	0	0***	0.05	0.522***	0.028***	0.036***	0.483***	0.95
	30	0.00077***	0.00002*	−0.00010***	0.13	0.453***	0.031***	0.034***	0.541***	0.97
	40	0.00221***	0.00021***	0.00016***	0.25	0.407***	0.033***	0.033***	0.570***	0.97
	50	0.00368***	0.00043***	0.00055***	0.58	0.373***	0.034***	0.033***	0.588***	0.97
8	20	0	0	0	0.00	0.611***	0.023***	0.031***	0.418***	0.92
	30	0.00016***	−0.00001***	−0.00011***	0.58	0.539***	0.026***	0.031***	0.495***	0.95
	40	0.00091***	−0.00003***	−0.00017***	0.25	0.488***	0.029***	0.031***	0.541***	0.96
	50	0.00186***	0.00007***	0	0.02	0.449***	0.032***	0.031***	0.574***	0.97

*, ** and *** denote significance at the level of 0.05, 0.01 and 0.001, respectively.

(vi) Regarding the factors F_{shape} and F_{spread}, the ratios between their estimated coefficients and the intercept parameter were smaller for ADEA than for ACE in all the experiments. This means that ADEA critical values are more robust against different input distribution shapes and spreads. In contrast, ADEA critical values are affected by the probability of an input being relevant; this is not true for ACE critical values.
(vii) The coefficients of F_{shape}, F_{spread}, and p_{rel} were positive in most of the experiments. Hence, input distributions with less excess kurtosis (i.e., taller tails), greater coefficient of variation, and higher probabilities of an input being relevant are generally associated with greater critical values.

We shall now suggest default critical values for ACE and ADEA, as shown in Table 11, by selecting a level for each factor in the linear models, as detailed below. DEA practitioners may nevertheless select other levels for the factors to obtain better critical values based on the characteristics of the data in their particular DEA problem.

(i) $F_{\text{shape}} = -1$ (lognormal), based on the assumption that the inputs will more frequently have a lognormal or similar distribution than a uniform distribution.
(ii) $F_{\text{spread}} = +1$ (51%), since the inputs will commonly have a high coefficient of variation in many DEA problems, potentially exceeding 100%.
(iii) $p_{\text{rel}} = 0.6$ to be conservative regarding the increase in ADEA critical values, since they increase with p_{rel} and we do not know the actual probability that an input is relevant.

5.2. Performance analysis of ACE and ADEA methods

In this subsection, we detail and compare the performance of ACE and ADEA methods when using critical values generated by our methodology. In particular, we summarize and illustrate only the results of the performance metrics associated with nominal size $\alpha = 5\%$. (Numerical results for both $\alpha = 5\%$ and $\alpha = 10\%$ are given in Appendix A.) Fig. 1 shows that the estimated type I er-

Table 11
Suggested default ACE and ADEA critical values at $\alpha = 5\%$.

Number of inputs	*ACE method* Number of DMUs				*ADEA method* Number of DMUs			
	20	30	40	50	20	30	40	50
2	0.0660	0.0757	0.0819	0.0868	0.294	0.264	0.250	0.238
3	0.0251	0.0335	0.0392	0.0433	0.403	0.381	0.388	0.402
4	0.0086	0.0156	0.0209	0.0248	0.539	0.489	0.460	0.441
5	0.0022	0.0068	0.0105	0.0139	0.652	0.602	0.567	0.544
6	0.0001	0.0025	0.0049	0.0074	0.747	0.701	0.666	0.640
7	0	0.00065	0.00216	0.00380	0.820	0.781	0.749	0.725
8	0	0.00006	0.00077	0.00179	0.870	0.841	0.815	0.792

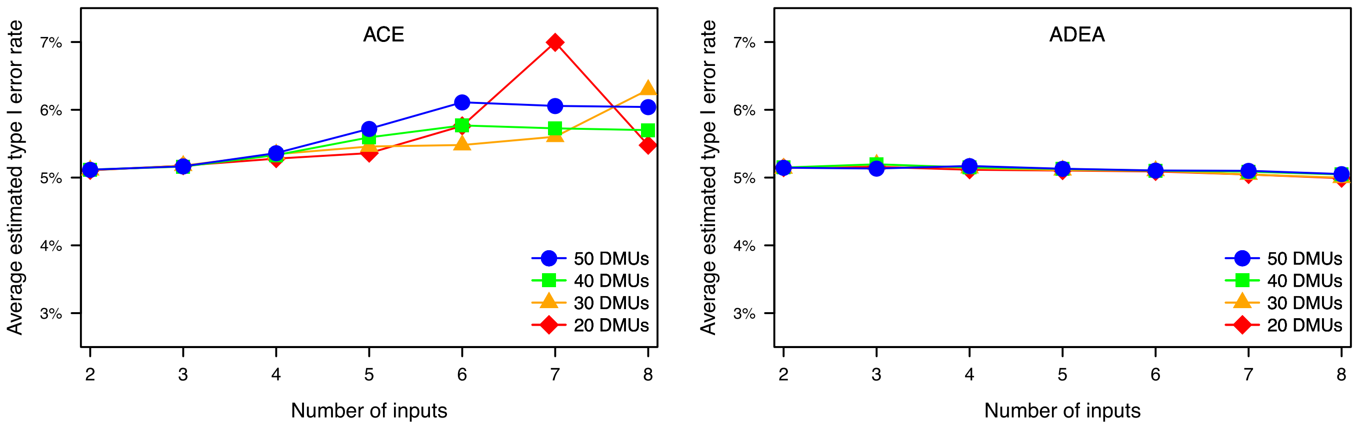

Fig. 1. ACE (left) and ADEA (right) average estimated Type I error rate at $\alpha = 5\%$ in our experiments.

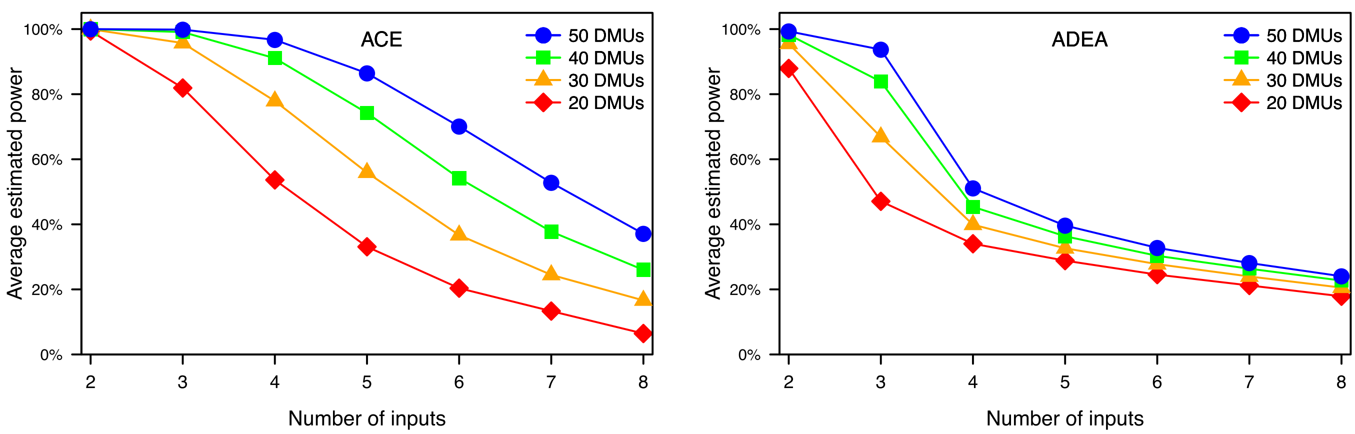

Fig. 2. ACE (left) and ADEA (right) average estimated power at $\alpha = 5\%$ in our experiments.

ror rate (A) was very close to the nominal size of 5% for ADEA, whereas these rates steadily increased from 5% to 6% as the number of inputs in the DEA model increased from 2 to 8.

Fig. 2 shows that the average estimated power of both methods increased with more DMUs in the model and decreased with more inputs, as expected. In comparison, ACE had a higher average estimated power than ADEA in all model specifications, except for those with 6–8 inputs and 20 DMUs, where ADEA was more powerful than ACE. Notice that ACE critical values were close to zero (i.e., about 0.001 and below) in the latter model specifications, indicating that ACE power decays to low levels as its critical values get very close to zero, which is associated with a high number of inputs in the model. It is worth highlighting the following characteristics of ADEA power curves: (i) A steep descent when the number of inputs in the model increased from 3 to 4, and (ii) an asymptotic behavior starting at six inputs, which is an indication that ADEA power slowly decreases beyond that number of inputs.

Fig. 3 illustrates that ACE and ADEA average strong type I error rates (B) were low in all model specifications. Specifically, we can see that the average B rates were below 1.2%, indicating that more than 3.8% of the 5% estimated type I error rate was associated with the elimination of inputs with the minimum relevant parameter value, i.e., $\lambda_{\text{rel}} = 0.1$. The B rates of both methods decreased as the number of DMUs increased, which is desirable. Notably, ACE average B rates were below 0.2% in all model specifications with 30 or more DMUs. In comparison, ACE showed smaller B rates than ADEA in the experiments with three or more inputs, whereas the opposite held for the experiments with two inputs.

Fig. 4 shows that ADEA average tie-related type I error rates (W) were very close to the nominal size $\alpha = 5\%$ in the model specifications with four or more inputs. This is a consequence of how ADEA

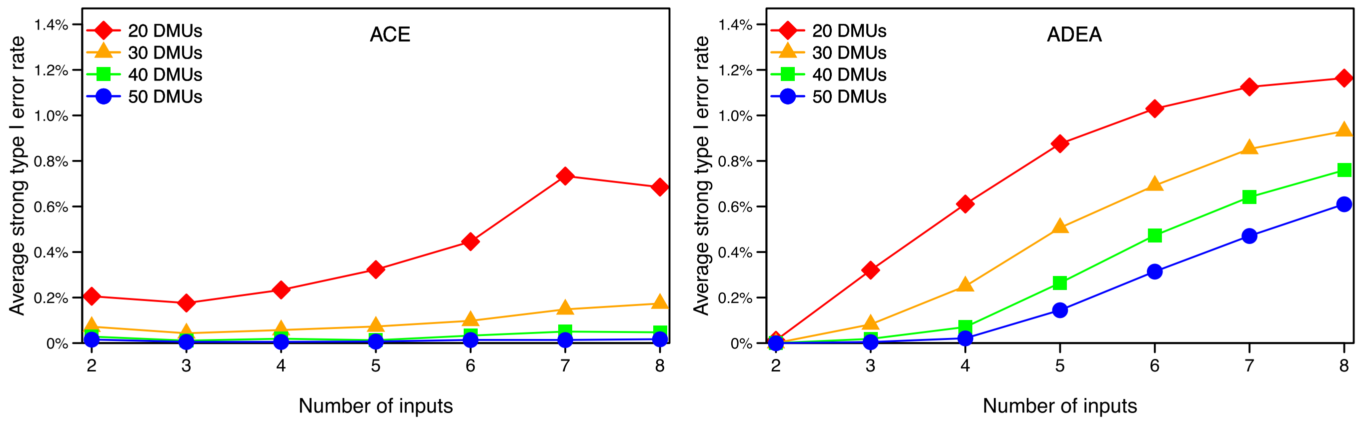

Fig. 3. ACE (left) and ADEA (right) average strong Type I error rate at $\alpha = 5\%$ in our experiments.

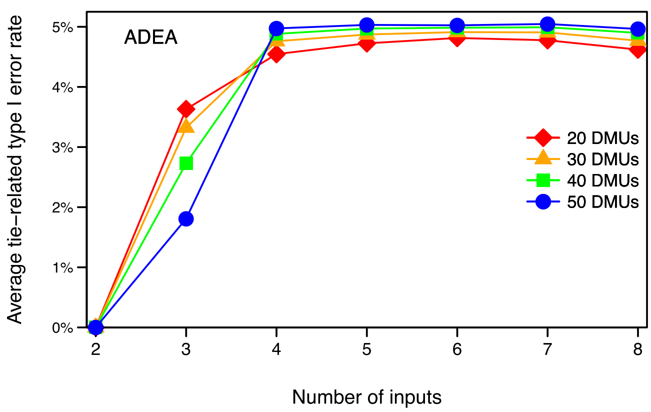

Fig. 4. ADEA average tie-related Type I error rate at $\alpha = 5\%$ in our experiments.

calculates the relevance measures; ties are more likely to occur as the number of variables in the model increases. Further, this behavior causes ADEA power to decrease significantly when the number of inputs increases from 3 to 4. In contrast, the W rates for ACE were practically 0% for all model specifications.

Finally, we can see in Fig. 5 that the average type I error impact (Δ) of both methods was small, being between −0.06 and 0 for most model specifications. Therefore, type I error did not significantly decrease the correlation between the theoretical and estimated efficiencies of the DMUs. It is also clear that the magnitude of this impact was smaller with more inputs in the model and slightly larger when the number of DMUs increased. Moreover, ACE displayed a better performance than ADEA regarding the D rates in all the experiments.

6. Conclusion

In this paper, we proposed a methodology for calculating the critical values of relevance measures in variable selection methods in DEA. In addition, we defined a set of performance metrics to evaluate these methods when using the critical values generated by our methodology. We then conducted an extensive simulation study in which we applied the proposed methodology to ACE and ADEA methods in 28 single-output model specifications (i.e., using 2–8 inputs and 20, 30, 40, and 50 DMUs) for several different scenarios by varying the DGP factors. The main findings from these simulations can be summarized as follows.

The magnitude and behavior of the calculated critical values across all the model specifications were reasonable in relation to the formulation of each method. The estimated type I error rate when using the calculated critical values was very close to the nominal size, and the estimated power of both methods increased with more DMUs and decreased with more inputs in the DEA model. Notably, in most model specifications, the critical values of ACE could be estimated using a linear model with a high adjusted R^2, using the shape and spread of the input probability distribution as independent variables, whereas the linear models for ADEA critical values also included the probability of an input being relevant in the DEA model as an independent variable; thus, these models could be used to estimate critical values for ACE and ADEA in

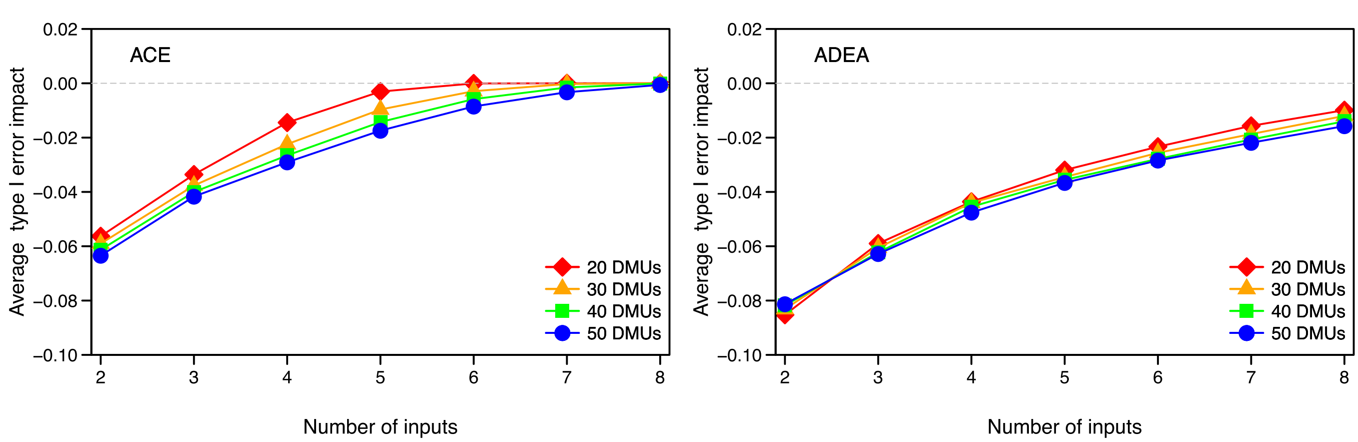

Fig. 5. ACE (left) and ADEA (right) average Type I error impact at $\alpha = 5\%$ in our experiments.

any single-output DEA model with 2–8 inputs and 20–50 DMUs, as explained in Section 5.1. Moreover, the proportion of efficient DMUs in the sample, the shape and mean of the efficiency distribution of non-efficient DMUs, and the functional form of the production function did not have a significant impact on the calculated critical values. Therefore, the proposed methodology suitable for generating critical values that are robust against the common assumptions in the literature regarding these four factors.

In terms of performance, we observed that ACE usually achieved higher power estimates than ADEA, except for the model specifications with a high number of inputs and a low number of DMUs (i.e., 6–8 inputs with 20 DMUs) where ADEA was more powerful than ACE. Notably, ADEA power curves show an asymptotic behavior starting at six inputs, i.e., the power slowly decreases beyond that number of inputs. We identified ACE's main drawback to be the convergence of its critical values to zero as the number of inputs increases because its power decreases given that the average change in the estimated efficiencies associated with the elimination of relevant inputs approaches zero, and such a change is similar to that associated with irrelevant inputs. In comparison, we identified ADEA's main shortcoming to be the high occurrence of ties where relevant and irrelevant inputs have the same relevance measure (particularly when there are four or more inputs in the model) since such ties have a notable negative impact on its power. Furthermore, for both methods, type I error was generally associated with the elimination of inputs with the minimum relevant parameter value used among all production functions. Additionally, the impact of type I errors on the correlation between the theoretical and DEA estimated efficiencies of the DMUs was small for both methods.

Based on these findings, we conclude that our methodology generates consistent and rational results, and can be applied in any variable selection method based on a single relevance measure to find its critical values in a given single-output DEA model. Specifically, the inputs can be generated using the probability distribution that best fits the shape and average coefficient of variation of the actual inputs in the dataset, and the probability of an input being relevant can be adjusted based on the DEA user' assessment of how well-selected the inputs are. It is safe to assume that the levels of the other four factors in the DGP can be set to one of the options proposed in this paper since they do not have a significant impact on the critical values, as noted above. It is also possible to incorporate additional components into the data generation process, such as input generation from a multivariate distribution with a correlation matrix, or limiting the theoretical relevance of the inputs in the production function.

Furthermore, our methodology addresses the following issues: (i) the lack of guidelines for choosing appropriate relevance measures thresholds based on observable characteristics of the dataset, in order to apply variable selection methods based on such measures (Nataraja & Johnson, 2011), and (ii) the fact that the required assumptions for performing statistical tests based on relevance measures are not necessarily satisfied (Sirvent et al., 2005).

Suggestions for future research include: (i) implementing the proposed methodology in the multi-output case; (ii) applying the proposed methodology to other variable selection methods and improving the ACE and ADEA methods; (iii) analyzing the impact of input-input and input-output correlations and the returns to scale of the production function on the critical values; (iv) developing a variable selection approach by combining the results obtained using the critical values of two or more variable selection methods.

Acknowledgments

This work was supported by Universidad Simón Bolívar, Universidad de Cádiz and Asociación Universitaria Iberoamericana de Postgrado. We would like to thank the editor and the three anonymous referees for their constructive and valuable suggestions to improve this paper.

Appendix A. Additional simulation study results

Table A1
ANOVA of ACE critical values at $\alpha = 10\%$.

Source of variation	Sum of squares	df	Mean square	F	p-value
Number of inputs	24.0896	6	4.0149	91908.67	0.0000
Number of DMUs	0.4831	3	0.1610	1680.74	0.0000
p_{rel}	0.0021	3	0.0007	7.45	0.0001
f_{form}	0.0017	1	0.0017	17.85	0.0000
F_{shape}	0.2502	1	0.2502	2612.10	0.0000
F_{spread}	1.3721	1	1.3721	14322.12	0.0000
ϕ	0.0134	1	0.0134	139.89	0.0000
H_{shape}	0.0004	1	0.0004	3.75	0.0528
H_{mean}	0.0095	1	0.0096	100.20	0.0000
Residuals	3.4317	35,821	0.0000096		
Total	29.6539	35,839			

Table A2
ANOVA for ADEA critical values at $\alpha = 10\%$.

Source of variation	Sum of squares	df	Mean square	F	p-value
Number of inputs	1528.64	6	254.77	140383.99	0.0000
Number of DMUs	34.90	3	11.63	6410.58	0.0000
p_{rel}	149.09	3	49.70	27384.59	0.0000
f_{form}	0.05	1	0.05	25.54	0.0000
F_{shape}	13.31	1	13.31	7331.90	0.0000
F_{spread}	15.92	1	15.92	8771.55	0.0000
ϕ	0.22	1	0.22	118.59	0.0000
H_{shape}	0.06	1	0.06	34.10	0.0000
H_{mean}	0.31	1	0.31	172.00	0.0000
Residuals	65.01	35,821	0.00181		
Total	1807.51	35,839			

Table A3
Coefficients of linear models for ACE and ADEA critical values at $\alpha = 10\%$.

Inputs	DMUs	ACE method				ADEA method				
		Intercept	F_{shape}	F_{spread}	R^2_{adj}	Intercept	F_{shape}	F_{spread}	p_{rel}	R^2_{adj}
2	20	0.0698***	0.0088***	0.0217***	0.90	0.335***	0	−0.004***	−0.004*	0.10
	30	0.0774***	0.0098***	0.0240***	0.90	0.293***	0.001*	−0.004***	−0.003	0.13
	40	0.0823***	0.0104***	0.0257***	0.90	0.273***	0.001***	−0.003***	−0.004**	0.14
	50	0.0860***	0.0108***	0.0268***	0.90	0.258***	0.001***	−0.002***	−0.001	0.11
3	20	0.0265***	0.0035***	0.0083***	0.91	0.300***	0.011***	0.009***	0.297***	0.89
	30	0.0333***	0.0039***	0.0098***	0.93	0.302***	0.009***	0.009***	0.246***	0.82
	40	0.0376***	0.0039***	0.0108***	0.93	0.362***	0.011***	0.015***	0.141***	0.72
	50	0.0407***	0.0040***	0.0116***	0.94	0.414***	0.012***	0.017***	0.052***	0.73
4	20	0.0103***	0.0017***	0.0030***	0.83	0.386***	0.022***	0.025***	0.374***	0.95
	30	0.0165***	0.0025***	0.0049***	0.89	0.338***	0.020***	0.022***	0.363***	0.95
	40	0.0208***	0.0028***	0.0060***	0.92	0.314***	0.018***	0.018***	0.344***	0.94
	50	0.0240***	0.0029***	0.0068***	0.93	0.299***	0.017***	0.014***	0.328***	0.93
5	20	0.0035***	0.0006***	0.0007***	0.68	0.461***	0.028***	0.032***	0.436***	0.96
	30	0.0078***	0.0012***	0.0020***	0.83	0.392***	0.027***	0.029***	0.461***	0.96
	40	0.0112***	0.0017***	0.0030***	0.87	0.357***	0.027***	0.027***	0.458***	0.96
	50	0.0141***	0.0020***	0.0039***	0.90	0.336***	0.026***	0.025***	0.449***	0.96
6	20	0.0007***	0***	−0.0001***	0.17	0.543***	0.027***	0.035***	0.439***	0.96
	30	0.0034***	0.0004***	0.0005***	0.61	0.470***	0.029***	0.033***	0.487***	0.97
	40	0.0058***	0.0008***	0.0012***	0.76	0.423***	0.030***	0.032***	0.509***	0.97
	50	0.0079***	0.0011***	0.0018***	0.82	0.391***	0.031***	0.031***	0.519***	0.97
7	20	0.00004***	0.00001***	−0.00003***	0.45	0.635***	0.021***	0.031***	0.386***	0.93
	30	0.00131***	0.00006***	−0.00007***	0.06	0.558***	0.025***	0.031***	0.458***	0.96
	40	0.00291***	0.00028***	0.00026***	0.36	0.504***	0.028***	0.032***	0.502***	0.97
	50	0.00444***	0.00051***	0.00067***	0.61	0.463***	0.030***	0.032***	0.531***	0.98
8	20	0***	0	0***	0.01	0.721***	0.017***	0.025***	0.311***	0.87
	30	0.00037***	−0.00004***	−0.00019***	0.59	0.649***	0.020***	0.026***	0.390***	0.92
	40	0.00136***	−0.00001	−0.00015***	0.15	0.594***	0.023***	0.027***	0.446***	0.94
	50	0.00240***	0.00011***	0.00005***	0.06	0.550***	0.026***	0.028***	0.488***	0.95

*, ** and *** denote significance at the level of 0.05, 0.01 and 0.001, respectively.

Table A4
Suggested default ACE and ADEA critical values at $\alpha = 10\%$.

Number of inputs	ACE method				ADEA method			
	Number of DMUs				Number of DMUs			
	20	30	40	50	20	30	40	50
2	0.0827	0.0916	0.0976	0.1020	0.329	0.286	0.267	0.254
3	0.0313	0.0392	0.0445	0.0483	0.476	0.450	0.451	0.450
4	0.0116	0.0189	0.0240	0.0279	0.613	0.558	0.520	0.493
5	0.0026	0.0086	0.0125	0.0160	0.727	0.671	0.632	0.604
6	0.0006	0.0035	0.0062	0.0086	0.814	0.766	0.730	0.702
7	0	0.00118	0.00289	0.00460	0.877	0.839	0.809	0.784
8	0	0.00022	0.00122	0.00234	0.916	0.889	0.866	0.845

Table A5
ACE and ADEA average performance metrics at $\alpha = 5\%$ in our experiments.

Inputs	DMUs	ACE					ADEA				
		$A_\alpha(\%)$	$P_\alpha(\%)$	$B_\alpha(\%)$	$W_\alpha(\%)$	Δ_α	$A_\alpha(\%)$	$P_\alpha(\%)$	$B_\alpha(\%)$	$W_\alpha(\%)$	Δ_α
2	20	5.11	99.41	0.21	0	−0.056	5.15	87.91	0.01	0	−0.085
	30	5.11	99.98	0.07	0	−0.059	5.14	95.51	0	0.00	−0.083
	40	5.12	100.00	0.03	0	−0.061	5.15	98.28	0	0.00	−0.082
	50	5.12	100.00	0.02	0	−0.063	5.15	99.32	0	0.00	−0.081
3	20	5.18	81.90	0.18	0	−0.034	5.16	47.07	0.32	3.63	−0.059
	30	5.18	95.72	0.04	0	−0.038	5.19	66.85	0.08	3.32	−0.061
	40	5.16	99.09	0.01	0	−0.040	5.20	83.87	0.02	2.73	−0.062
	50	5.17	99.85	0.01	0	−0.042	5.13	93.71	0	1.81	−0.063
4	20	5.28	53.64	0.23	0	−0.014	5.12	34.02	0.61	4.55	−0.044
	30	5.35	77.89	0.06	0	−0.022	5.14	39.88	0.25	4.76	−0.044
	40	5.33	91.05	0.02	0	−0.027	5.15	45.37	0.07	4.88	−0.045
	50	5.36	96.71	0.01	0	−0.029	5.17	51.07	0.02	4.97	−0.048
5	20	5.36	33.09	0.32	0	−0.003	5.10	28.84	0.88	4.73	−0.032
	30	5.46	55.90	0.07	0	−0.010	5.11	32.64	0.51	4.87	−0.035
	40	5.59	74.22	0.01	0	−0.014	5.13	36.33	0.26	4.97	−0.036
	50	5.72	86.43	0.01	0	−0.017	5.13	39.62	0.14	5.03	−0.037
6	20	5.76	20.37	0.45	0	0	5.09	24.54	1.03	4.81	−0.023
	30	5.48	36.68	0.10	0	−0.003	5.10	27.76	0.69	4.91	−0.026
	40	5.77	54.17	0.03	0	−0.006	5.10	30.31	0.47	4.98	−0.028
	50	6.11	70.08	0.01	0	−0.009	5.10	32.76	0.31	5.03	−0.028
7	20	6.99	13.33	0.73	0.01	0	5.05	21.20	1.13	4.78	−0.016
	30	5.60	24.46	0.15	0	0	5.05	23.97	0.85	4.91	−0.019
	40	5.73	37.77	0.05	0	−0.002	5.09	26.35	0.64	4.99	−0.021
	50	6.06	52.79	0.01	0	−0.003	5.10	28.12	0.47	5.05	−0.022
8	20	5.48	6.43	0.69	0.03	0	4.99	17.89	1.16	4.62	−0.010
	30	6.30	16.61	0.17	0	0	5.00	20.55	0.93	4.77	−0.012
	40	5.70	25.99	0.05	0	0	5.05	22.72	0.76	4.90	−0.014
	50	6.04	37.10	0.02	0	−0.001	5.05	24.04	0.61	4.96	−0.016

Table A6
ACE and ADEA average performance metrics at $\alpha = 10\%$ in our experiments.

Inputs	DMUs	ACE					ADEA				
		$A_\alpha(\%)$	$P_\alpha(\%)$	$B_\alpha(\%)$	$W_\alpha(\%)$	Δ_α	$A_\alpha(\%)$	$P_\alpha(\%)$	$B_\alpha(\%)$	$W_\alpha(\%)$	Δ_α
2	20	10.11	99.85	0.74	0	−0.060	10.16	93.80	0.08	0	−0.083
	30	10.12	100.00	0.32	0	−0.062	10.16	97.74	0.01	0	−0.081
	40	10.13	100.00	0.16	0	−0.064	10.16	99.14	0	0.00	−0.080
	50	10.14	100.00	0.08	0	−0.066	10.14	99.67	0	0.00	−0.080
3	20	10.18	89.09	0.46	0	−0.037	10.14	62.71	1.09	5.88	−0.063
	30	10.16	97.87	0.13	0	−0.039	10.15	80.42	0.34	4.98	−0.061
	40	10.18	99.61	0.04	0	−0.041	10.13	91.85	0.09	3.56	−0.061
	50	10.17	99.95	0.01	0	−0.043	10.11	96.81	0.02	2.03	−0.061
4	20	10.27	65.71	0.54	0	−0.018	10.10	47.06	1.72	8.35	−0.046
	30	10.28	85.65	0.14	0	−0.024	10.12	53.55	0.85	8.69	−0.045
	40	10.35	94.76	0.04	0	−0.028	10.14	58.87	0.32	8.90	−0.045
	50	10.31	98.25	0.02	0	−0.030	10.16	64.20	0.13	9.00	−0.045
5	20	10.28	45.66	0.66	0	−0.005	10.09	41.28	2.23	8.92	−0.034
	30	10.40	67.29	0.17	0	−0.012	10.09	45.47	1.42	9.20	−0.036
	40	10.54	82.56	0.04	0	−0.016	10.10	49.43	0.86	9.42	−0.036
	50	10.60	91.56	0.02	0	−0.019	10.13	52.40	0.49	9.58	−0.036
6	20	10.41	31.06	0.84	0	0	10.02	35.94	2.47	9.15	−0.024
	30	10.46	49.42	0.22	0	−0.004	10.08	39.91	1.78	9.40	−0.027
	40	10.61	65.98	0.08	0	−0.007	10.08	42.68	1.31	9.57	−0.028
	50	10.92	79.22	0.03	0	−0.010	10.10	45.35	0.93	9.68	−0.029
7	20	11.58	21.43	1.15	0.01	0	9.93	31.33	2.64	9.03	−0.016
	30	10.44	36.21	0.28	0	−0.001	10.01	35.10	2.08	9.39	−0.019
	40	10.64	50.13	0.12	0	−0.002	10.03	37.84	1.61	9.59	−0.021
	50	10.87	64.29	0.03	0	−0.004	10.03	40.00	1.24	9.71	−0.022
8	20	11.03	13.67	1.29	0.04	0	9.89	26.97	2.67	8.85	−0.010
	30	10.97	26.46	0.33	0	0	9.98	30.38	2.23	9.20	−0.012
	40	10.58	37.72	0.10	0	0	9.98	32.94	1.84	9.39	−0.014
	50	10.86	49.40	0.03	0	−0.001	9.97	34.82	1.53	9.52	−0.015

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2020.08.021.

References

Adler, N., & Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western europe. *European Journal of Operational Research, 132*(2), 18–31. https://doi.org/10.1016/S0377-2217(00)00150-8.

Adler, N., & Golany, B. (2002). Including principal component weights to improve discrimination in data envelopment analysis. *Journal of the Operational Research Society, 53*(9), 985–991. https://doi.org/10.1057/palgrave.jors.2601400.

Adler, N., & Yazhemsky, E. (2010). Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. *European Journal of Operational Research, 202*(1), 273–284. https://doi.org/10.1016/j.ejor.2009.03.050.

Amirteimoori, A., Despotis, D., & Kordrostami, S. (2012). Variable reduction in data envelopment analysis. *Optimization, 63*(5), 735–745. https://doi.org/10.1080/02331934.2012.684354.

Banker, R. D. (1996). Hypothesis tests using data envelopment analysis. *The Journal of Productivity Analysis, 7*(2/3), 139–159. https://doi.org/10.1007/BF00157038.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8.

Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis: methodology and applications* (1st). Springer Science + Business Media.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research, 132*, 245–259. https://doi.org/10.1016/S0377-2217(00)00149-1.

Eskelinen, J. (2017). Comparison of variable selection techniques for data envelopment analysis in a retail bank. *European Journal of Operational Research, 259*(2), 778–788. https://doi.org/10.1016/j.ejor.2016.11.009.

Fanchon, P. (2003). Variable selection for dynamic measures of efficiency in the computer industry. *International Advances in Economic Research, 9*(3), 175–188. https://doi.org/10.1007/BF02295441.

Fernandez-Palacin, F., Lopez-Sanchez, M. A., & Munoz-Marquez, M. (2017). Stepwise selection of variables in DEA using contribution loads. *Pesquisa Operacional, 38*(1), 31–52. https://doi.org/10.1590/0101-7438.2018.038.01.0031.

Holland, D., & Lee, S. (2002). Impacts of random noise and specification on estimates of capacity derived from data envelopment analysis. *European Journal of Operational Research, 137*(1), 10–21. https://doi.org/10.1016/S0377-2217(01)00087-X.

Jenkins, L., & Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research, 147*(1), 51–61. https://doi.org/10.1016/S0377-2217(02)00243-6.

Jitthavech, J. (2016). Variable elimination in nested DEA models: a statistical approach. *International Journal of Operational Research, 27*(3), 389–410. https://doi.org/10.1504/IJOR.2016.078945.

Kao, L.-J., Lu, C.-J., & Chiu, C. C. (2011). Efficiency measurement using independent component analysis and data envelopment analysis. *European Journal of Operational Research, 210*(2), 310–317. https://doi.org/10.1016/j.ejor.2010.09.016.

Li, Y., & Liang, L. (2010). A shapley value index on the importance of variables in DEA models. *Expert Systems with Applications, 37*(9), 6287–6292. https://doi.org/10.1016/j.eswa.2010.02.093.

Li, Y., Shi, X., Yang, M., & Liang, L. (2017). Variable selection in data envelopment analysis via akaike's information criteria. *Annals of Operations Research, 253*(1), 453–476. https://doi.org/10.1007/s10479-016-2382-2.

Limleamthong, P., & Guillén-Gosálbez, G. (2018). Mixed-integer programming approach for dimensionality reduction in data envelopment analysis: Application to the sustainability assessment of technologies and solvents. *Industrial & Engineering Chemistry Research, 57*(30), 9866–9878. https://doi.org/10.1021/acs.iecr.7b05284.

Lin, T.-Y., & Chiu, S. H. (2013). Using independent component analysis and network DEA to improve bank performance evaluation. *Economic Modelling, 32*(1), 608–616. https://doi.org/10.1016/j.econmod.2013.03.003.

Liu, J. S., Lu, L. Y., & Lu, W. M. (2016). Research fronts in data envelopment analysis. *Omega, 58*, 33–45. https://doi.org/10.1016/j.omega.2015.04.004.

Madhanagopal, R., & Chandrasekaran, R. (2014). Selecting appropriate variables for dea using genetic algorithm (ga) search procedure. *International Journal of Data Envelopment Analysis and *Operations Research*, 1*(2), 28–33. https://doi.org/10.12691/ijdeaor-1-2-3.

Morita, H., & Avkiran, N. K. (2009). Selecting inputs and outputs in data envelopment analysis by designing statistical experiments. *Journal of the Operations Research Society of Japan, 52*(2), 163–173. https://doi.org/10.15807/jorsj.52.163.

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. https://doi.org/10.1016/j.ejor.2011.06.045.

Pastor, J. T., Ruiz, J. L., & Sirvent, I. (2002). A statistical test for nested radial DEA models. *Inmaculada Operations Research, 50*(4), 728–735. https://doi.org/10.1287/opre.50.4.728.2866.

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. https://doi.org/10.1016/j.ejor.2008.11.013.

R Core Team (2017). *R: A language and environment for statistical computing, Vienna, Austria*. https://www.R-project.org/

Ruggiero, J. (2005). Impact assessment of input omission on DEA. *International Journal of Information Technology & Decision Making, 4*(3), 359–368. https://doi.org/10.1142/S021962200500160X.

Sexton, T. R., Silkman, R. H., & Hogan, A. J. (1986). Data envelopment analysis: critique and extensions. *New Directions for Program Evaluation, 1986*(32), 73–105. https://doi.org/10.1002/ev.1441.

Sharma, M. J., & Yu, S. J. (2015). Stepwise regression data envelopment analysis for variable reduction. *Applied Mathematics and Computation, 253*, 126–134. https://doi.org/10.1016/j.amc.2014.12.050.

Simar, L., & Wilson, P. W. (2001). Testing restrictions in nonparametric efficiency models. *Communications in Statistics – Simulation and Computation, 30*(1), 159–184. https://doi.org/10.1081/SAC-100001865.

Sirvent, I., Ruiz, J. L., Borraś, F., & Pastor, J. T. (2005). A monte carlo evaluation of several tests for the selection of variables in DEA models. *International Journal of Information Technology & Decision Making, 4*(3), 325–343. https://doi.org/10.1142/S0219622005001581.

Toloo, M., & Babaee, S. (2015). On variable reductions in data envelopment analysis with an illustrative application to a gas company. *Applied Mathematics and Computation, 270*, 527–533. https://doi.org/10.1016/j.amc.2015.06.122.

Toloo, M., Barat, M., & Masoumzadeh, A. (2015). Selective measures in data envelopment analysis. *Annals of Operations Research, 226*(1), 623–642. https://doi.org/10.1007/s10479-014-1714-3.

Toloo, M., & Tichý, T. (2015). Two alternative approaches for selecting performance measures in data envelopment analysis. *Measurement, 65*, 29–40. https://doi.org/10.1016/j.measurement.2014.12.043.

Ueda, T., & Hoshiai, Y. (1997). Application of principal component analysis for parsimonious summarization of DEA inputs and/or outputs. *Journal of the Operations Research Society of Japan, 40*(4), 466–478. https://doi.org/10.15807/jorsj.40.466.

Wagner, J. M., & Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research, 180*(1), 57–67. https://doi.org/10.1016/j.ejor.2006.02.048.