I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0316

CONFIDENCE INTERVALS FOR DEA-TYPE EFFICIENCY SCORES: HOW TO AVOID THE COMPUTATIONAL BURDEN OF THE BOOTSTRAP

BADIN, L. and L. SIMAR

http://www.stat.ucl.ac.be

Confidence Intervals for DEA-type Efficiency Scores: How to Avoid the Computational Burden of the Bootstrap

Luiza Bădin*
Department of Mathematics
Academy of Economic Studies
Piaţa Romană nr. 6
Bucharest, Romania

Léopold Simar†
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium

July 11, 2003

Abstract

One important issue in statistical inference is to provide confidence intervals for the parameters of interest. Once the statistical properties of the estimators have been established, the corresponding asymptotic results can be used for constructing confidence intervals. However, in nonparametric efficiency estimation, the asymptotic properties of DEA estimators are only available for the bivariate case (Gijbels *et al.*, 1999). An appealing alternative is the bootstrap method and a general methodology for applying bootstrap in nonparametric frontier estimation is provided by Simar and Wilson (1998, 2000b). Nevertheless, all the procedures involving bootstrap method are based on a large number of data replications, and in frontier estimation this approach also implies performing DEA (*i.e.* solving linear programs) a large number of times. Hence, a more simple and less computing intensive technique is always welcome.

In this paper we propose a simple procedure for constructing confidence intervals for the efficiency scores. We consider some classical confidence intervals for an endpoint of a distribution and we show how these results can be adapted to the problem of frontier estimation. We provide an algorithm for constructing similar confidence intervals for the efficiency scores. Then some Monte Carlo experiments estimate the coverage probabilities of the obtained intervals. The results are quite satisfactory even for small samples. We then illustrate the approach with a real data set when analyzing the efficiency of 36 Air Controllers in Europe.

Keywords: Nonparametric Frontier Models, Efficiency, Bootstrap, Random Subsampling, Confidence Intervals.

*Research support from the EC project "Inco-Copernicus", IC15-CT98-1007 is acknowledged.

†Research support from "Projet d'Actions de Recherche Concertées" (No. 98/03–217) and from the "Interuniversity Attraction Pole", Phase V (No. P5/24) from the Belgian Government are also acknowledged.

1 Introduction

The efficiency scores of economic producers are usually evaluated by measuring the radial distance, either in the input space or in the output space, from each producer to an estimated production frontier. The nonparametric approach known as Data Envelopment Analysis (DEA) is based on the idea of enveloping the data, under specific assumptions on the technology such as free disposability, convexity, scale restrictions, without imposing any uncertain parametric structure. However, the method gives only point estimates for the true efficiency scores which are unknown, since the true frontier is unknown.

It turns out that DEA measures efficiency relative to a nonparametric, maximum likelihood estimate of the frontier, conditional on observed data resulting from an underlying data-generating process (DGP). These methods have been widely applied to examine technical and allocative efficiency in a variety of industries; see Lovell (1993) and Seiford (1996, 1997) for comprehensive bibliographies of these applications. Aside from the production setting, the problem of estimating monotone concave boundaries also naturally occurs in portfolio management. In capital asset pricing models (CAPM), the objective is to analyze the performance of investment portfolios. Risk and average return on a portfolio are analogous to inputs and outputs in models of production; in CAPM, the attainable set of portfolios is naturally convex and the boundary of this set gives a benchmark relative to which the efficiency of a portfolio can be measured. These models were developed by Markovitz (1959) and others; Sengupta (1991) and Sengupta and Park (1993) provide links between CAPM and nonparametric estimation of frontiers as in DEA.

Although labeled *deterministic*, statistical properties of DEA estimators are now available (see Simar and Wilson, 2000a). Banker (1993) proved the consistency of the DEA efficiency estimator, but without giving any information on the rate of convergence. Korostelev *et al.* (1995a, 1995b) proved the consistency of DEA estimators of the attainable set and also derived the speed of convergence and Kneip *et al.* (1998) proved the consistency of DEA in the multivariate case, providing the rates of convergence as well. Gijbels *et al.* (1999) derived the asymptotic distribution of DEA estimators in the bivariate case (one input and one output).

We begin by introducing some basic concepts of nonparametric efficiency measurement in the spirit of Simar and Wilson (2000a). Suppose producers use input vector $x \in \mathbb{R}_+^p$ to produce output vector $y \in \mathbb{R}_+^q$. The production set of the feasible input-output combinations can be defined as:

$$\Psi = \left\{(x, y) \in \mathbb{R}_+^{p+q} |\ x \text{ can produce } y\right\}. \tag{1.1}$$

For any $y \in \mathbb{R}_+^q$ we denote by $X(y)$ the input requirement set, i.e. the set of all input

vectors which yield at least y :

$$X(y) = \{x \in \mathbb{R}^p_+ | \, (x,y) \in \Psi\}.$$

The input efficient frontier is a subset of $X(y)$ and is given by:

$$\partial X(y) = \{x \in \mathbb{R}^p_+ | \; x \in X(y), \; \theta x \notin X(y) \;\; \forall \theta \in (0,1)\}.$$

Now, the corresponding Farrell input-oriented measure of technical efficiency can be defined as:

$$\theta(x,y) = \inf\{\theta | \; \theta x \in X(y)\}. \tag{1.2}$$

A value $\theta(x,y) = 1$ means that producer (x,y) is input efficient, while a value $\theta(x,y) < 1$ suggests the radial reduction in all inputs that producer (x,y) should perform in order to produce the same output being input-efficient. For a given level of output and an input direction, the efficient level of input is defined by:

$$x^{\partial}(x,y) = \theta(x,y)x. \tag{1.3}$$

The basic definition of radial technical efficiency dates back to Debreu (1951) and Farrell (1957). Based on Farrell's idea, Charnes *et al.*(1978) proposed a linear programming model for evaluating the technical efficiency assuming convexity of the set of feasible input-output combinations and free disposability of inputs and outputs. This approach is known as Data Envelopment Analysis (DEA).

Suppose the sample of observed producers is $\mathcal{X} = \{(x_i,y_i), \; i = 1,...,n\}$. The idea of the DEA method is to estimate the technical efficiency of a producer (x_0,y_0), by measuring its radial distance to the boundary of the conical hull[1] of $\mathcal{X}$. The DEA estimator of Ψ is defined by:

$$\widehat{\Psi}_{DEA} = \left\{(x,y) \in \mathbb{R}^{p+q}_+ | \; y \leq \sum_{i=1}^{n} \gamma_i y_i, \; x \geq \sum_{i=1}^{n} \gamma_i x_i, \; \gamma_i \geq 0, \; i = 1,...,n\right\}. \tag{1.4}$$

Replacing Ψ with its DEA estimate yields:

$$\widehat{X}(y) = \left\{x \in \mathbb{R}^p_+ | \, (x,y) \in \widehat{\Psi}_{DEA}\right\},$$

$$\partial \widehat{X}(y) = \left\{x \in \mathbb{R}^p_+ | \; x \in \widehat{X}(y), \; \theta x \notin \widehat{X}(y) \;\; \forall \theta \in (0,1)\right\},$$

[1] We consider here a technology with constant returns to scale. The procedure below will also be valid with varying returns to scale. In this case, in equations (1.4) and (1.5), we must add the constraint $\sum_{i=1}^{n} \gamma_i = 1$.

and the corresponding DEA estimator of the input efficiency score of the producer (x_0, y_0) is obtained by solving the following linear program:

$$\widehat{\theta}_{DEA}(x_0, y_0) = \min \left\{ \theta > 0 |\ y_0 \leq \sum_{i=1}^{n} \gamma_i y_i,\ \theta x_0 \geq \sum_{i=1}^{n} \gamma_i x_i,\ \gamma_i \geq 0,\ i = 1, ..., n \right\}. \quad (1.5)$$

The efficient level of input is obtained from (1.3):

$$\widehat{x}^{\partial}(x_0, y_0) = \widehat{\theta}_{DEA}(x_0, y_0) x_0. \quad (1.6)$$

The DEA method has become very popular and widely used due to its nonparametric nature which requires very few assumptions on technology. Consistency and asymptotic distribution are now available for DEA estimators. However, the asymptotic distribution of DEA estimator is available only in the bivariate case, therefore the bootstrap remains the single tool for statistical inference in a more general setting.

Under some specific regularity assumptions on the data generating process (DGP), Simar and Wilson (1998, 2000b) propose two bootstrap procedures: one based on resampling the efficiency scores (*homogeneous* bootstrap) and the other, based on resampling the observed input-output pairs (*heterogeneous* bootstrap). Both allow estimating the bias and constructing confidence intervals for the efficiency scores.

The most critical aspect when applying bootstrap method in nonparametric frontier estimation is to correctly simulate the underlying DGP in order to produce a consistent bootstrap distribution estimator. Since the naive bootstrap is not consistent here, it is recommended to resample from a smooth, consistent estimator of either the density of the scores or, the joint density of the input-output pairs. Each procedure is very intensive in terms of computing time, since it involves many bootstrap replications with many DEA efficiency estimation as well. Hence, a more simple and less computing intensive technique for providing confidence intervals for the efficiency scores is welcome.

In this paper we propose some easy way to build confidence intervals for the DEA-type efficiency scores and present Monte Carlo evidence on the coverage probabilities of the estimated intervals. The intervals are based on asymptotic arguments but seem to perform satisfactory even for small samples. The paper is organized as follows. Section 2 reviews some classical results of extreme value theory and addresses the issue of confidence intervals estimation of an endpoint of a distribution. We focus on the problem of constructing confidence intervals for the lower bound of the support of a distribution using resampling techniques. In Section 3 we show how these intervals can be adapted to the problem of frontier estimation: we provide an algorithm for constructing confidence intervals for the input efficiency scores. Section 4 presents numerical results. First we discuss the results

from some Monte Carlo experiments (estimated coverages of the obtained intervals), then we illustrate the approach with a real data set on 36 Air Controllers in Europe. The paper ends with some conclusions. Throughout the paper, we consider only the input orientation, but the results can be easily adapted for the output-oriented case.

2 Inference for an Endpoint of a Distribution

The problem of estimating an endpoint or a truncation point of a distribution has attracted a lot of attention since many practical situations deal with variables that are restricted to a finite domain.

In this section we first review some classical results of extreme value theory and we focus afterward on inference for an endpoint using bootstrap method and random subsampling.

2.1 Some classical results of extreme value theory

Let $X_1, ...X_n$ be independent identically distributed random variables with cumulative distribution $F(x)$ whose support is bounded below and suppose we are interested in estimating ϕ, the lower endpoint of the support of F. If $X_{(1)} \leq ... \leq X_{(n)}$ are the order statistics, the most intuitive estimator of ϕ is $X_{(1)}$, the sample minimum, which is the closest observation to ϕ. However, since $X_{(1)}$ is always larger than the lower bound of the support, it is biased.

As far as confidence intervals will be of concern, asymptotic distributions will be useful. One of the most remarkable results of extreme value theory is due to Gnedenko (1943). According to Gnedenko (1943), a distribution $F(x)$ belongs to the domain of attraction of a $Weibull$-type limiting law if and only if:

(i) the support is bounded below by a real number $\phi = \inf\{x : F(x) > 0\} \in \mathbb{R}$, $F(\phi) = 0$;

(ii) for each $c > 0$, and for some $\delta > 0$,

$$\lim_{x \to 0+} \frac{F(cx + \phi)}{F(x + \phi)} = c^{\delta}. \tag{2.1}$$

The parameter δ in (2.1) is called *the index of variation* and it also represents the shape parameter of the limiting Weibull distribution. The $Weibull$-type limiting law can be introduced as follows: assuming there are sequences of real numbers a_n and b_n such that the limit distribution of the normalized minimum $Y_n = \left(X_{(1)} - b_n\right)/a_n$ is non-degenerate, then under (i) and (ii), this limit distribution is:

$$\Lambda(y) = \begin{cases} 1 - \exp\left(-y^{\delta}\right) & y > 0 \\ 1 & y \leq 0 \end{cases} \qquad Weibull$$

Under assumptions (i) and (ii), one can provide asymptotically valid confidence intervals for ϕ, as well as tests of hypothesis about ϕ, based on the quantiles of the Weibull distribution. Note that in many applications (and in particular for the frontier framework analyzed below), the value of δ will be known. Indeed, as pointed by Loh (1984), for all distribution F such that the density at the frontier point ϕ is finite and > 0, $\delta = 1$.

For $\delta = 1$, Robson and Whitlock (1964) proposed the one sided confidence interval for ϕ:

$$\left(X_{(1)} - \frac{1-\alpha}{\alpha} \left(X_{(2)} - X_{(1)} \right), \quad X_{(1)} \right) \tag{2.2}$$

with asymptotic coverage $1-\alpha$. This result was generalized by Cooke (1979), who obtained under (2.1) the interval:

$$\left(X_{(1)} - \frac{(1-\alpha)^{1/\delta}}{1-(1-\alpha)^{1/\delta}} \left(X_{(2)} - X_{(1)} \right), \quad X_{(1)} \right), \tag{2.3}$$

which also has asymptotic coverage $1-\alpha$.

2.2 Inference for an endpoint with resampling methods

Another approach to the estimation of an endpoint is to use resampling methods to estimate the distribution of the normalized extreme involved and to approximate numerically this distribution and the quantiles required for constructing confidence intervals.

Bickel and Freedman (1981) noticed the failure of bootstrap method when attempting to estimate the uniform distribution by bootstrapping the normalized maximum. Addressing a similar problem, Loh (1984) pointed out that even if bootstrap distribution is not consistent, the bootstrap confidence intervals may still be consistent in some situations. The problem of endpoint estimation with resampling methods was further investigated by Swanepoel (1986) who proved that weak consistency can be achieved by choosing an appropriate size $m = m(n)$ for the bootstrap samples, such that $m \to \infty$ and $m/n \to 0$. Deheuvels *et al.* (1993) refined Swanepoel's result by deriving as well the range of rates for $m(n)$ which ensure the strong consistency of bootstrap distribution. They defined the bootstrap distribution using an unified form of extreme value distributions. On the contrary, Athreya and Fukuchi (1997) defined consistent bootstrap distributions with associated resample sizes for each of the three types of domain of attraction. They also constructed confidence intervals for the lower bound based on bootstrap versions of Weissman's (1982) statistics.

We focus in the following on inference for the lower bound ϕ using bootstrap method and Hartigan's (1969) random subsampling. We shall describe in more detail the method proposed by Loh (1984) in order to apply the same idea in frontier framework.

Suppose $\{X_1, ..., X_n\}$ is a random sample from a distribution F which satisfies (i) and (ii) and let $X_{(1)} \leq ... \leq X_{(n)}$ be the order statistics. Let $\{X_1^*, ..., X_n^*\}$ be a bootstrap sample with the corresponding order statistics $X_{(1)}^* \leq ... \leq X_{(n)}^*$ and denote by P_* the associate resampling probability. Consider also the natural estimator of ϕ, $\widehat{\phi} = X_{(1)}$ and its bootstrap correspondent $\widehat{\phi}^* = X_{(1)}^*$. If the bootstrap sample is obtained by sampling with replacement from the original data, then P_* simply corresponds to the empirical distribution and we have:

$$P_*\left(\widehat{\phi}^* < X_{(j+1)} | X_1, ..., X_n\right) = 1 - \left(1 - \frac{j}{n}\right)^n \to 1 - e^{-j}, \tag{2.4}$$

for $j = 1, ..., n-1$. Now, if we consider a random subsample, that is any arbitrary nonempty subset of random size from the original sample $\{X_1, ..., X_n\}$, then the random subsampling probability P_{RS} gives:

$$P_{RS}\left(\widehat{\phi}^* < X_{(j+1)} | X_1, ..., X_n\right) = 1 - \frac{2^{n-j} - 1}{2^n - 1} \to 1 - 2^{-j}, \tag{2.5}$$

for $j = 1, ..., n-1$.

Loh (1984) proves that the conditional distribution of $n^{1/\delta}\left(\widehat{\phi}^* - \widehat{\phi}\right)$ under (2.4) or (2.5) does not have a weak limit, therefore it can not approximate the true distribution of $n^{1/\delta}\left(\widehat{\phi} - \phi\right)$ not even in the limit. Consequently, confidence intervals based on the percentiles of these distributions will not be consistent.

Still, Loh (1984) proposes another method, initially criticized by Efron (1979) and derives consistent confidence intervals based on bootstrap and random subsampling. Let $t_{1-\alpha}$ be the $100\,(1-\alpha)$ percentile of the bootstrap distribution of $\widehat{\phi}^*$. We have:

$$\begin{aligned} P_*\left(\widehat{\phi}^* < t_{1-\alpha}\right) &= P_*\left(\widehat{\phi}^* - \widehat{\phi} < t_{1-\alpha} - \widehat{\phi}\right) \\ &= P_*\left(\widehat{\phi} - t_{1-\alpha} < \widehat{\phi} - \widehat{\phi}^*\right) = 1 - \alpha. \end{aligned}$$

According to the bootstrap principle, the distribution of $\widehat{\phi} - \widehat{\phi}^*$ could be an approximation of the true distribution of $\phi - \widehat{\phi}$, therefore:

$$P\left(\widehat{\phi} - t_{1-\alpha} < \phi - \widehat{\phi}\right) \simeq 1 - \alpha.$$

The latter approximation is equivalent to:

$$P\left(\phi > 2\widehat{\phi} - t_{1-\alpha}\right) \simeq 1 - \alpha,$$

which gives the following approximate $1 - \alpha$ confidence interval for ϕ:

$$\left(2\widehat{\phi} - t_{1-\alpha},\ \widehat{\phi}\right). \tag{2.6}$$

However, it turns out that the asymptotic validity of the intervals and the coverage probability as well, depend on the value of δ.

Weissman (1981) considered the following pivotal quantity:

$$R_{k,n} = \frac{X_{(1)} - \phi}{X_{(k)} - X_{(1)}},\ k \geq 2$$

and derived its asymptotic distribution

$$P\left(R_{k,n} \leq x\right) \to 1 - \left(1 - \left(\frac{x}{1+x}\right)^{\delta}\right)^{k-1},\ x \geq 0. \tag{2.7}$$

From (2.7) with $x = 1$, for each fixed $1 \leq j \leq n-1$, as $n \to \infty$ we have:

$$P\left(\frac{\widehat{\phi} - \phi}{X_{(1+j)} - X_{(1)}} < 1\right) \to 1 - \left(1 - 2^{-\delta}\right)^{j}$$

and equivalently

$$P\left(\phi > 2X_{(1)} - X_{(1+j)}\right) \to 1 - \left(1 - 2^{-\delta}\right)^{j}.$$

Since an attempt to bootstrap Weissman's pivotal quantity could produce an undefined quantity, one may choose to bootstrap only the numerator, keeping the denominator unchanged. Then the bootstrap and random subsampling distributions of the resulting statistic give respectively:

$$P_{*}\left(\frac{\widehat{\phi}^{*} - \widehat{\phi}}{X_{(1+j)} - X_{(1)}} < 1\right) = P_{*}\left(\widehat{\phi}^{*} < X_{(1+j)}\right) \to 1 - e^{-j}, \tag{2.8}$$

$$P_{RS}\left(\frac{\widehat{\phi}^{*} - \widehat{\phi}}{X_{(1+j)} - X_{(1)}} < 1\right) = P_{RS}\left(\widehat{\phi}^{*} < X_{(1+j)}\right) \to 1 - 2^{-j}. \tag{2.9}$$

It can be proved (Loh, 1984) that if $1 - \alpha = 1 - e^{-j}$, for some j $(1 \leq j \leq n-1)$, then $\left(2\widehat{\phi} - X_{(1+j)},\ \widehat{\phi}\right)$ is an approximate $1 - \alpha$ bootstrap confidence interval for ϕ if and only if $\delta = \ln\left(1 - e^{-1}\right) / \ln 2^{-1}$, which is obtained by solving:

$$1 - \left(1 - 2^{-\delta}\right)^{j} = 1 - e^{-j}.$$

Similarly, if $1 - \alpha = 1 - 2^{-j}$, then $\left(2\widehat{\phi} - X_{(1+j)},\ \widehat{\phi}\right)$ is an approximate $1 - \alpha$ random subsampling confidence interval for ϕ if and only if $\delta = 1$, i.e. the solution of:

$$1 - \left(1 - 2^{-\delta}\right)^{j} = 1 - 2^{-j}.$$

Therefore, $\left(2X_{(1)} - X_{(1+j)},\ X_{(1)}\right)$ is an asymptotically valid confidence interval for ϕ, with associated confidence coefficient $1 - e^{-j}$, if and only if $\delta = \ln\left(1 - e^{-1}\right) / \ln 2^{-1}$, and $1 - 2^{-j}$ if and only if $\delta = 1$.

The appealing thing about this interval is that, although resampling arguments are used to derive its asymptotic confidence coefficient, there is in fact no need to resample. Moreover, assuming F satisfies (i) and (ii) with δ known, using a "generalized bootstrap" defined by resampling so that:

$$P_*\left(\widehat{\phi}^* < X_{(j+1)}\right) = 1 - \left(1 - 2^{-\delta}\right)^j, \tag{2.10}$$

the interval

$$\left(2X_{(1)} - X_{(1+j)},\ X_{(1)}\right) \tag{2.11}$$

has asymptotic confidence coefficient $1 - \alpha = 1 - \left(1 - 2^{-\delta}\right)^j$ (Loh, 1984). When δ is unknown, the results for the generalized bootstrap can be made adaptive by replacing δ with a consistent estimate. But remember that in most applications, and in particular in frontier estimation, $\delta = 1$.

So, in the next section, we show how the two main results provided so far, the Cooke (2.3) and the Loh (2.11) confidence intervals for boundary points, can be adapted to the problem of finding confidence intervals for efficiency scores.

3 Confidence Intervals for Efficiency Scores

3.1 A "toy" problem: a univariate frontier

Let us come to the problem of frontier estimation and connect it with the issues discussed previously. Consider the one-dimensional frontier problem and suppose each firm produces one unit of output using input X, where X is a random variable with values in (ϕ, ∞), $\phi > 0$. We are interested in estimating ϕ, the lower boundary of the support of X, based on a sample $\mathcal{X} = \{(x_i, 1), i = 1, ..., n\}$. In terms of frontiers, ϕ can be interpreted as the lowest level of input required to produce one unit of output and can be defined as:

$$\phi = \inf\left\{x | F_X(x) > 0\right\},$$

where $F_X(x) = P(X \leq x)$. According to (1.3), for an arbitrary fixed point $(x_0, 1) \in \mathcal{X}$ we have:

$$x^{\partial}(x_0, 1) = \theta x_0 = \phi. \tag{3.1}$$

This is also the input efficient frontier which is unknown and has to be estimated.

Assuming further that $F_X(\cdot)$ satisfies (ii), we are just in the situation described in the previous section and we can provide asymptotically valid confidence intervals for the efficient level of input. We shall consider the interval based on asymptotic theory defined in (2.3) and the interval (2.11) based on resampling methods. Let $x_{(1)} \leq ... \leq x_{(n)}$ be the ordered values of the inputs. The point estimator for the efficient level of input is $\widehat{\phi} = x_{(1)}$ and the intervals:

$$\mathcal{I}_1 = \left(x_{(1)} - (1-\alpha)^{1/\delta} \left(1 - (1-\alpha)^{1/\delta}\right)^{-1} \left(x_{(2)} - x_{(1)}\right), \quad x_{(1)} \right) \tag{3.2}$$

$$\mathcal{I}_2 = \left(2x_{(1)} - x_{(1+j)}, \ x_{(1)}\right) \tag{3.3}$$

are approximate $1-\alpha$ and $1 - \left(1 - 2^{-\delta}\right)^j$ confidence intervals for ϕ.

From (3.1), replacing ϕ with $\widehat{\phi}$, we obtain an estimator for the input efficiency score of producer $(x_0, 1)$:

$$\widehat{\theta}_0 = \frac{x_{(1)}}{x_0} \tag{3.4}$$

which is simply the DEA input efficiency score. Consequently, dividing the endpoints of the intervals (3.2) and (3.3) by x_0, we obtain:

$$\left(\frac{x_{(1)}}{x_0} - (1-\alpha)^{1/\delta} \left(1 - (1-\alpha)^{1/\delta}\right)^{-1} \left(\frac{x_{(2)}}{x_0} - \frac{x_{(1)}}{x_0}\right), \quad \frac{x_{(1)}}{x_0} \right) \tag{3.5}$$

$$\left(2\frac{x_{(1)}}{x_0} - \frac{x_{(1+j)}}{x_0}, \ \frac{x_{(1)}}{x_0}\right). \tag{3.6}$$

If $\widehat{\theta}_{(1)} \leq ... \leq \widehat{\theta}_{(n)} = 1$ are the ordered values of the estimated efficiency scores, then $\widehat{\theta}_{(n-j)} = x_{(1)}/x_{(1+j)}$, $j = 0, ..., n-1$ and the intervals (3.5) and (3.6) become:

$$\mathcal{J}_1 = \left(\widehat{\theta}_0 - (1-\alpha)^{1/\delta} \left(1 - (1-\alpha)^{1/\delta}\right)^{-1} \left(\widehat{\theta}_0/\widehat{\theta}_{(n-1)} - \widehat{\theta}_0\right), \quad \widehat{\theta}_0 \right) \tag{3.7}$$

$$\mathcal{J}_2 = \left(2\widehat{\theta}_0 - \widehat{\theta}_0/\widehat{\theta}_{(n-j)}, \ \widehat{\theta}_0\right). \tag{3.8}$$

The intervals $\mathcal{J}_1$ and $\mathcal{J}_2$ are approximate $1-\alpha$ and $1 - \left(1 - 2^{-\delta}\right)^j$ confidence intervals for the efficiency score θ_0.

3.2 Multivariate situation

Suppose we observe a sample of n producers $\mathcal{X} = \{(x_i, y_i), i = 1, ..., n\}$, where $x_i \in \mathbb{R}^p$ is the input vector and $y_i \in \mathbb{R}^q$ is the output vector of producer i and the production possibilities set is Ψ, defined in (1.1). The nonparametric approach to efficiency measurement, so-called deterministic, assumes that all the observations are attainable, that is $\text{Prob}((x_i, y_i) \in \Psi) = 1$ for $i = 1, ..., n$. In order to define a statistical model in this multivariate setting, we need to make some economic assumptions and, also, appropriate assumptions on the DGP. We adopt here the statistical model defined in Simar and Wilson (2000a).

ASSUMPTION 1 *Ψ is convex and inputs and outputs in Ψ are freely disposable i.e. if $(x, y) \in \Psi$, then $(\tilde{x}, \tilde{y}) \in \Psi$ for all $(\tilde{x}, \tilde{y})$ such that $\tilde{x} \geq x$ and $\tilde{y} \leq y$.*

ASSUMPTION 2 *The sample observations in $\mathcal{X}$ are realizations of iid random variables on Ψ with pdf $f(x, y)$ and the structure of inefficiency is homogeneous.*

Since we are dealing with radial distances, it is useful to describe the position of (x, y) in terms of the polar coordinates of x:

$$\begin{array}{ll} \textit{modulus}: & \omega = \omega(x) \in \mathbb{R}_+ \\ \textit{angle}: & \eta = \eta(x) \in [0, \frac{\pi}{2}]^{p-1} \end{array}$$

and consequently, using cylindrical coordinates we have $(x, y) \Leftrightarrow (\omega, \eta, y)$, which allows us to decompose the density $f(x, y)$ as follows:

$$f(x, y) = f(\omega|\eta, y) f(\eta|y) f(y), \tag{3.9}$$

assuming that all the conditional densities exist.

For a given (η, y), the efficient level of input has modulus

$$\omega\left(x^{\partial}(x, y)\right) = \inf\{\omega \in \mathbb{R}_+ | \ f(\omega|\eta, y)\},$$

so that we have $\theta(x, y) = \omega\left(x^{\partial}(x, y)\right) / \omega(x)$.

The conditional density $f(\omega|\eta, y)$ on $\left[\omega\left(x^{\partial}(x, y)\right), \infty\right)$ induces a density $f(\theta|y, \eta)$ on $[0, 1]$. The homogeneity assumption on the structure of inefficiency implies

$$f(\theta|\eta, y) = f(\theta). \tag{3.10}$$

ASSUMPTION 3 *There is a probability mass in a neighborhood of the true frontier, that is, for all $y \geq 0$ and $\eta \in [0, \frac{\pi}{2}]^{p-1}$, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $\forall$ $\omega \in [\omega(x^{\partial}(x, y)), \omega(x^{\partial}(x, y)) + \epsilon_2]$, $f(\omega|\eta, y) \geq \epsilon_1$.*

We know, (Kneip *et al.* (1998), that under Assumptions 1–3, the DEA efficiency score $\widehat{\theta}(x_0, y_0)$ is a consistent estimator of $\theta(x_0, y_0)$, where (x_0, y_0) is the producer of interest. In this multivariate setup, we can come back to the one-dimensional "toy" problem described above by focusing on the modulus of the frontier point.

Consider the true pseudo-data generated on the ray passing through the point of interest (x_0, y_0) as follows. Due to the homogeneity assumption in (3.10), we can consider the set of pseudo-data defined for $i = 1, \ldots, n$ as

$$\begin{aligned} \tilde{x}_i &= \theta_0 x_0/\theta_i \\ y_i &= y_0, \end{aligned} \tag{3.11}$$

where $\theta_i = \theta(x_i, y_i)$ are the true but unknown efficiency scores of the data points (x_i, y_i). These pseudo-true data are n pseudo observations generated on the ray of interest and having the same efficiency distribution as the original data cosidered in the DGP defined above.

The frontier point on this ray has modulus $\omega(x^{\partial}(x_0, y_0)) = \omega(x_0)\theta_0$, therefore, in terms of the modulus, the pseudo data defined in (3.11) can be written:

$$\widetilde{\omega}_i = \omega(x^{\partial}(\widetilde{x}_i, y_0)) = \frac{\theta_0 \omega(x_0)}{\theta_i}, \text{ for } i = 1, ..., n. \tag{3.12}$$

Since θ_i are unknown, a naive approach would be to plug the (consistent) estimated values $\widehat{\theta}_0$ and $\widehat{\theta}_i$ in (3.12) to obtain

$$\omega_i^* = \frac{\widehat{\theta}_0 \omega(x_0)}{\widehat{\theta}_i}, \text{ for } i = 1, ..., n. \tag{3.13}$$

These ω_i^* can play the same role as the observations x_i in the univariate "toy" problem above, yielding the order statistics $\omega_{(i)}^* = \widehat{\theta}_0 x_0/\widehat{\theta}_{(n-i+1)}$ for $i = 1, ...n$, where $\widehat{\theta}_{(1)} \leq ... \leq \widehat{\theta}_{(n)}$ are the ordered DEA scores.[2]

This would allow one to find, as above, confidence intervals for $\omega(x_0^{\partial})$ as those obtained in (3.2) and (3.3) for ϕ, then dividing the endpoints by $\omega(x_0)$, this would provide the same confidence intervals for θ_0 as given in (3.7) and (3.8). Namely, the confidence intervals for $\omega(x_0^{\partial})$ with asymptotic level $1 - \alpha$ and $1 - \left(1 - 2^{-\delta}\right)^j$, would be:

$$\mathcal{I}_1 = \left(\omega_{(1)}^* - (1-\alpha)^{1/\delta} \left(1 - (1-\alpha)^{1/\delta}\right)^{-1} \left(\omega_{(2)}^* - \omega_{(1)}^*\right), \ \ \omega_{(1)}^* \right) \tag{3.14}$$

$$\mathcal{I}_2 = \left(2\omega_{(1)}^* - \omega_{(1+j)}^*, \ \omega_{(1)}^*\right) \tag{3.15}$$

[2]Remember that $\widehat{\theta}_{(n)} = 1$, therefore $\omega_{(1)}^* = \widehat{\theta}_0 \omega(x_0) = \omega\left(\widehat{x}_0^{\partial}\right)$.

Dividing the endpoints of the intervals (3.14) and (3.15) by $\omega(x_0)$ we would obtain the corresponding confidence intervals for θ_0:

$$\mathcal{J}_1 = \left(\widehat{\theta}_0 - \frac{(1-\alpha)^{1/\delta}}{\left(1-(1-\alpha)^{1/\delta}\right)} \left(\widehat{\theta}_0 / \widehat{\theta}_{(n-1)} - \widehat{\theta}_0 \right), \quad \widehat{\theta}_0 \right) \tag{3.16}$$

$$\mathcal{J}_2 = \left(2\widehat{\theta}_0 - \widehat{\theta}_0 / \widehat{\theta}_{(n-j)}, \ \widehat{\theta}_0 \right). \tag{3.17}$$

The latter procedure can certainly be used in the very particular case of $p = q = 1$ under CRS (constant returns to scale), because the frontier is a straight line and its estimator has only one support point, so that only one $\hat{\theta}_i$ will be equal to 1 (with probability 1). The finite sample performances of the procedure, in this very particular case, will be investigated in the Monte-Carlo experiment in Section 4.1. The procedure is quite satisfactory even in small samples.

Unfortunately, in the multivariate case, the DEA estimator always produces more than one efficient unit and the number of such units increases with $p+q$ and also with the sample size. It will also increase under the VRS (varying returns to scale) hypothesis. Consequently, the estimated efficiency scores involved in the lower bound of our confidence intervals may be equal to 1, even for, say, $j = 7$, constraining the intervals to be empty. This comes from the spurious discretization at one of the DEA estimators, whereas, the density of θ is continuous.

To overcome this drawback, instead of providing our pseudo-true data by plugging the DEA efficiency estimates of θ in (3.11), we will simulate pseudo-data by generating values of θ from a smooth estimate of the continuous density of θ and keeping in mind that by definition, $\widehat{\theta}_{(n)} = 1$. This can be viewed as adapting the *smoothed bootstrap*, as proposed in Simar and Wilson (1998), but used here in a different context.

To achieve this, we first eliminate from our pseudo-sample all the spurious efficiency scores equal to 1, then we estimate $f(\widehat{\theta})$ from the remaining $\widehat{\theta}$ and generate B samples of size $n-1$, $\left\{ \widehat{\theta}_1^{*b}, ..., \widehat{\theta}_{n-1}^{*b} \right\}_{b=1}^{B}$ from $f(\widehat{\theta})$. Here all the generated values are $< \widehat{\theta}_{(n)} = 1$. Then we approximate $\theta_{(n-j)}$ for some j $(1 \leq j \leq n-1)$ by the average of $\widehat{\theta}_{(n-j)}^{*b}$ over the B simulations:

$$\widetilde{\theta}_{(n-j)} = \frac{1}{B} \sum_{b=1}^{B} \widehat{\theta}_{(n-j)}^{*b}. \tag{3.18}$$

Thus we avoid the discretization problem and we attempt to provide a better approximation for $\theta_{(n-j)}, j \geq 1$ in this multivariate setup. We know from the literature that the non-parametric estimation of a density near a boundary point is a difficult problem and so, we

can think that the performances of our estimators $\widetilde{\theta}_{(n-j)}$ could be poor for small values of j, as $j = 1, 2$. However, the Monte-Carlo experiments below will show a quite satisfactory behavior.

A kernel density estimator for $f(z)$, based on observations $z_1, ..., z_n$ is defined by:

$$\widehat{f}(z) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{z - z_i}{h}\right) \tag{3.19}$$

where $K(\cdot)$ is the *kernel function* and h is the *bandwidth parameter*. Usually, the kernel function is a probability density function, symmetric around zero. To obtain a good estimate in terms of mean integrated square error and of course a consistent estimator, the bandwidth parameter has to be chosen appropriately. If the distribution of the data is approximately normal, then K could be the standard normal density function and h could be chosen according to the normal reference rule, or a more robust choise could be

$$h = 0.79Rn^{-1/5}, \tag{3.20}$$

where R is the interquartile range of the data (Silverman, 1986). Another approach is to use a data driven bandwidth h. For instance, the least squares cross-validation method determines the value of h that minimizes an approximation to mean integrated square error.

In order to estimate $f(\widehat{\theta})$ taking into account the boundary condition $\widehat{\theta} < 1$, we employ the *reflection method* (Silverman, 1986), using as kernel the standard normal density function $\phi(\cdot)$. We follow the same idea as in Simar and Wilson (1998).

Suppose we have m inefficient producers and denote $\mathcal{S}_m = \left\{\widehat{\theta}_1, ..., \widehat{\theta}_m\right\}$. Let each point $\widehat{\theta}_j < 1$ be reflected in the boundary by $2 - \widehat{\theta}_j > 1$ and consider the $2m$ points $\left\{\widehat{\theta}_1, ..., \widehat{\theta}_m, 2 - \widehat{\theta}_1, ..., 2 - \widehat{\theta}_m\right\}$ If $\widehat{g}_h(z)$ is a kernel density estimate constructed from the $2m$ points

$$\widehat{g}_h(z) = \frac{1}{2mh} \sum_{j=1}^{m} \left[\phi\left(\frac{z - \widehat{\theta}_j}{h}\right) + \phi\left(\frac{z - \left(2 - \widehat{\theta}_j\right)}{h}\right)\right],$$

then a consistent estimator of $f(\widehat{\theta})$ is:

$$\widehat{f}_h(z) = \begin{cases} 2\widehat{g}_h(z) & \text{if } z < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since the Gaussian kernel has unbounded support, the truncation of $\widehat{g}_h(z)$ on the right at unity provides a density estimate $\widehat{f}_h(z)$ with support on $(-\infty, 1)$. The input efficiency estimates are also bounded below by 0, but since there will be no values near 0, we ignore the effect of the left boundary and we focus on the upper bound of the support.

Now sampling from $\widehat{f}_h(z)$ is very simple and does not require the calculation of $\widehat{f}_h(z)$. Let $\left\{\beta_1^*, ..., \beta_{n-1}^*\right\}$ be a bootstrap sample, obtained by sampling with replacement from $\mathcal{S}_m$ and $\left\{\epsilon_1^*, ..., \epsilon_{n-1}^*\right\}$ a random sample of standard normal deviates. By the convolution formula, we have:

$$\widetilde{\theta}_i^* = \beta_i^* + h\epsilon_i^* \sim \frac{1}{m}\sum_{j=1}^{m}\frac{1}{h}\phi\left(\frac{z-\widehat{\theta}_{(j)}}{h}\right),$$

for $i = 1, ..., n-1$. Using the same convolution formula we get also:

$$2 - \widetilde{\theta}_i^* = 2 - (\beta_i^* + h\epsilon_i^*) \sim \frac{1}{m}\sum_{j=1}^{m}\frac{1}{h}\phi\left(\frac{z-\left(2-\widehat{\theta}_{(j)}\right)}{h}\right).$$

Define now for $i = 1, ...n-1$ the bootstrap data:

$$\theta_i^* = \begin{cases} \widetilde{\theta}_i^* & \text{if } \widetilde{\theta}_i^* < 1, \\ 2 - \widetilde{\theta}_i^* & \text{otherwise.} \end{cases} \tag{3.21}$$

It is easy to prove that θ_i^* defined in (3.21) are random variables distributed according to $\widehat{f}_h(z)$.

To rescale the bootstrap data so that the variance is approximately the sample variance of $\widehat{\theta}_i$, we employ the following transform:

$$\widehat{\theta}_i^* = \overline{\widehat{\theta}} + \left(1 + h^2/\widehat{\sigma}^2\right)^{-1/2}\left(\theta_i^* - \overline{\widehat{\theta}}\right), \tag{3.22}$$

where $\overline{\widehat{\theta}} = \frac{1}{n}\sum_{j=1}^{n}\widehat{\theta}_j$ and $\widehat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{\theta}_j - \overline{\widehat{\theta}}\right)^2$.

The method described above is synthesized in the following algorithm.

Algorithm:

[1] For each observed producer $(x_i, y_i) \in \mathcal{X}_n$, compute the DEA estimator of the efficiency score $\widehat{\theta}_i = \widehat{\theta}_{DEA}(x_i, y_i)$, $i = 1, ..., n$.

[2] If $(x_0, y_0) \notin \mathcal{X}_n$ repeat step [1] for (x_0, y_0) to obtain $\widehat{\theta}_0 = \widehat{\theta}_{DEA}(x_0, y_0)$.

[3] Define $\mathcal{S}_m = \left\{\widehat{\theta}_1, ..., \widehat{\theta}_m\right\}$ where $m = \#\left\{\widehat{\theta}_i < 1\right\}_{1\le i\le n}$, i.e. the number of inefficient producers.

[5] Determine the bandwidth h based on $\mathcal{S}_m$, via (3.20) or least-squares cross-validation method.

[6] Draw $n-1$ bootstrap values $\widehat{\theta}_i^*$, $i=1,...,n-1$ from the kernel density estimate of $f(\widehat{\theta})$ and sort in ascending order: $\widehat{\theta}_{(1)}^* \leq ... \leq \widehat{\theta}_{(n-1)}^*$.

[7] Repeat step [6] B times, to obtain a set of B bootstrap estimates $\left\{\widehat{\theta}_{(n-j)}^{*b}\right\}_{b=1}^{B}$, for some $1 \leq j \leq n-1$.

[8] Compute $\widetilde{\theta}_{(n-j)}$ by bagging the B estimates obtained on step [7]:

$$\widetilde{\theta}_{(n-j)} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\theta}_{(n-j)}^{*b}.$$

[9] The confidence intervals for θ_0 are:

$$\mathcal{J}_1 = \left(\widehat{\theta}_0 - \frac{(1-\alpha)^{1/\delta}}{\left(1-(1-\alpha)^{1/\delta}\right)} \left(\widehat{\theta}_0/\widetilde{\theta}_{(n-1)} - \widehat{\theta}_0 \right), \quad \widehat{\theta}_0 \right) \tag{3.23}$$

$$\mathcal{J}_2 = \left(2\widehat{\theta}_0 - \widehat{\theta}_0/\widetilde{\theta}_{(n-j)}, \ \widehat{\theta}_0 \right). \tag{3.24}$$

One possible shortcoming of both intervals $\mathcal{J}_1$ and $\mathcal{J}_2$ is that the scale factors involved in the lower bound of the intervals are the same whatever be (x_0, y_0), the point of interest, but this is a consequence of the homogeneity assumption regarding the structure of inefficiency.

Note that the computational burden of our *smooth bootstrap* is much less than the full homogeneous bootstrap as proposed in Simar and Wilson (1998), because we only compute once the DEA scores at the first step of the algorithm. In the full bootstrap, a DEA linear program has to be performed for each bootstrap replication, and as pointed in Simar and Wilson (1998), the procedure has to be repeated a large number of times (say, above 2000).

Note that here, the widths of the two confidence intervals can be written as:

$$\text{width}(\mathcal{J}_1) = \frac{(1-\alpha)^{1/\delta}}{\left(1-(1-\alpha)^{1/\delta}\right)} \times \widehat{\theta}_0 \left(\frac{1}{\widetilde{\theta}_{(n-1)}} - 1 \right)$$

$$\text{width}(\mathcal{J}_2) = \widehat{\theta}_0 \left(\frac{1}{\widetilde{\theta}_{(n-j)}} - 1 \right).$$

We do not have a proof that the asymptotic coverage probabilities obtained by the two intervals are the nominal ones, but in the next section, we investigate by a Monte-Carlo experiment how they behave in small to moderate sample sizes. As pointed above the non-parametric estimation of θ_{n-j} is quite difficult for small j. This is confirmed by the Monte-Carlo experiments: the confidence intervals based on the Cooke (3.23) are not

reliable because the estimates $\widetilde{\theta}_{n-1}$ turned out to be too small, giving too large intervals with estimated coverage generally near to one. However, as shown below, the confidence intervals adapted from the Loh idea (3.24) behave quite well for j larger than 3, which is typically the range of interest in many applications (nominal confidence levels above 0.8750). So we will only focus on the Loh approach when smoothing is required (VRS and/or $p+q>2$).

4 Numerical Illustrations

4.1 Monte-Carlo evidence

• Model 1

We performed Monte Carlo experiments to analyze the coverage probabilities of the estimated confidence intervals. We considered a true technology which exhibits constant returns to scale and in the case $p=q=1$ we simulated data according to the model:

$$y = xe^{-|v|}, \tag{4.1}$$

with $x \sim Uniform\,(1,9)$ and the inefficiency term $v \sim Exp\,(1)$ and $v \sim N\,(0,1)$. The *producer* under evaluation was chosen $(x_0, y_0) = (5, 2)$. The true technical efficiency score according to (4.1) is $\theta_0 = y_0/x_0 = 0.4$.

Table 1 and Table 2 present the results of our Monte Carlo experiments through 2000 trials with $v \sim Exp\,(1)$ and $v \sim N\,(0,1)$ respectively. The confidence levels considered as benchmarks are $1-\alpha = 1-\left(1-2^{-\delta}\right)^{j}$, with $j = 3,4,5,6,7$ and δ is obviously equal to 1.

Since we are in a CRS scenario with $p=q=1$ we can use the simple confidence intervals provided in (3.16) and (3.17). We see that the procedure provides intervals with estimated coverage close to the desired confidence level even with only 25 observations. For $n \geq 100$ the estimated coverages are almost exact. The performances of our procedure are slightly better than the reported performances, with similar scenarios, in Simar and Wilson (2000a) and (2003). As pointed there, the bootstrap procedure lead generally to underestimated coverage probabilities and so, too narrow confidence intervals.

Figure 1 depicts a sample of 100 observations generated from model (4.1) with $v \sim Exp(1)$. The true frontier is $y = x$ and is represented by the dashed line. The estimated DEA frontier is an inward-biased estimator of the true frontier and is represented by the continuous line. The fixed point is (x_0, y_0) and the dots $(x_i^*, 2)$, $i = 1, ..., 100$ define the pseudo-data, with x_i^* given by (3.11).

• **Model 2**

The second model considered is also a CRS technology with two inputs and one output:

$$y = x_1^{1/3} x_2^{2/3} e^{-|v|} \tag{4.2}$$

where $v \sim N(0, (1/2)^2)$, and x_1, x_2 are independently generated from $Uniform(1, 9)$. The fixed point considered here is $(x_{01}, x_{02}, y_0) = \left(5, 8/\sqrt{5}, 3\right)$ with a true technical efficiency score $\theta_0 = 0.75$. Here again, $\delta = 1$. We performed simulations only for estimating the coverages of $\mathcal{J}_2$ and we report the results for three confidence levels obtained from $1 - \alpha = 1 - (1 - 2^{-j})$, for $j = 5, 6, 7$. For step [5] of Algorithm 2 we used the bandwidth given by (3.20) and for step [7] we performed $B = 200$ bootstrap replications.

The results are presented in Table 3. Again, in this case, the results are quite reliable even with moderate sample sizes. They are less good than in the preceding model, since we have a slower rate of the DEA estimators in this setup and we use the smoothing technique to estimate the order statistics of the efficiency scores. However, the performances of our procedure overcome slightly those of the bootstrap as reported in similar setup in Simar and Wilson (2003). Note also that we have, as expected, better performances for higher j (higher confidence levels). A look to the column average widths shed some lights on the precision of the obtained confidence intervals and also on the role of the sample size in this precision.

4.2 An Empirical illustration

We illustrate the approach with real data coming from the efficiency analysis of Air Controllers in Europe (Mouchart and Simar, 2002). We have data on activity of 36 european air controllers in the year 2000. The activity of each controller can be described by one input (an aggregate factor of different kind of labor) and one output (an aggregate factor of the activity produced, based on the number of air movements controlled, the number flight hours controlled, ...).

The observations and the VRS-DEA nonparametric frontier estimates appear in Figure 2. The efficiency scores are naturally input oriented (the activity of an Air Controler is typically exogeneous): they are provided in Table 4, second column. So, for instance, the unit #1 should reduce its labor by a factor of 0.6458 for being considered as efficient, comparing to the others units.

- The efficient units are units #7, #9, #14 and #26. They define the efficient frontier in Figure 2.
- The "'Loh" confidence intervals are computed according the procedure described above (see (3.24)), with a bandwidth selected as in (3.20).

- The bootstrap confidence intervals are the one-sided confidence intervals obtained by the Simar and Wilson (1998) procedure (homogeneous bootstrap).

We can see that the results are very similar: we have the same order of magnitude of the intervals but the bootstrap intervals are slighly narrower. This might be an advantage of our procedure since we know from reported Monte-Carlo experiments in the literature (see above) that the bootstrap confidence intervals have a tendency of being too narrow.

The computing time for the Loh method (with B = 100), was only 0.14 seconds (Pentium III, 450 Mghz processor), whereas, for the bootstrap (with B = 2000 loops, necessary for such confidence intervals), it increases to 19 minutes: so we have a factor of more than 8000 in favor of the procedure developped in this paper. This would be particularly useful when working with larger data sets and/or when performing Monte-Carlo experiments in this framework.

5 Conclusions

We proposed in this paper a simple and fast way of constructing confidence intervals for the efficiency scores. The Monte-Carlo evidence confirms that the method provides satisfactory results even for small samples. The alternative approach is the full bootstrap approach, which is much more computer intensive. Note that our attained coverage probabilities reported in our Monte-Carlo experiments, are at least as good as those reported in similar Monte-Carlo experiments for the full bootstrap approach (see Simar and Wilson , 2000a and 2003).

Our approach focuses on confidence intervals. The main advantage of the bootstrap over our approach is that symmetric confidence intervals of any level can be produced. The widths of the intervals seem also to be narrower, but at a cost of computing time which turned out to be a factor 10000 in our empirical illustration. For giving an idea, the last Monte-Carlo scenario of Table 3 took 5 hours on a Pentium III, 450 Mghz machine: an equivalent bootstrap experiment would be impossible on such a machine when applying a factor still greater than 8000 (here $p = 2, q = 1$).

Note also that the bootstrap remains the only available approach if one wants to estimate the bias and/or the standard deviation of the estimators of the efficiency scores. Also the approach here relies strongly on the homogeneous assumption. In a heteregenous case, the only alternative is the heterogeneous bootstrap of Simar and Wilson (2000b). Similarly, in testing issues, estimation of p-values needs also the use of the full bootstrap as in Simar and Wilson (1999), (2001) and (2002).

References

[1] Athreya, K. B., Fukuchi, J. (1997), Confidence intervals for endpoints of a c.d.f. via bootstrap, *Journal of Statistical Planning and Inference*, 58, 299-320.

[2] Banker, R.D. (1993), Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39(10), 1265-1273.

[3] Bickel, P.J., Freedman, D.A. (1981), Some asymptotic theory for the bootstrap, *Annals of Statistics* 9, 1196-1217.

[4] Charnes, A., Cooper, W.W., Rhodes, E. (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2(6), 429-444.

[5] Cooke, P.J. (1979), Statistical inference for bounds of random variables, *Biometrika* 66, 367-374.

[6] Debreu, G. (1951), The coefficient of resource utilization, *Econometrica* 19(3), 273–292.

[7] Deheuvels, P., Mason, D., Shorack, G. (1993), Some results on the influence of extremes on the bootstrap, *Annals of Institute Henri Poincaré*, 29, 83-103.

[8] Deprins, D., Simar, L., Tulkens, H. (1984), Measuring Labor Inefficiency in Post Offices, in *The Performance of the Public Enterprises: Concepts and Measurements*, eds. Marchand, M., Pestieau, P. and Tulkens H., Amsterdam, North Holland, 243-267.

[9] Efron, B. (1979), Bootstrap methods: Another look at the Jackknife, *Annals of Statistics* 7, 1-26.

[10] Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society, Series A,* 120, 253-281.

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

[12] Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics* 44, 423-454.

[13] Hartigan, J. (1969), Using subsample values as typical values, *Journal of the American Statistical Association* 64, 1303-1317.

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

[15] Korostelev, A., Simar, L., Tsybakov, A.B. (1995a), Efficient estimation of monotone boundaries, *Annals of Statistics* 23, 476-489.

[16] Korostelev, A., Simar, L., Tsybakov, A.B. (1995b), On estimation of monotone and convex boundaries, *Publications de l 'Institut de Statistique de l 'Université de Paris XXXIX* 1, 3-18.

[17] Loh, W. (1984), Estimating an endpoint of a distribution with resampling methods, *Annals of Statistics* 12, 1543-1550.

[18] Lovell, C.A.K. (1993), Production Frontiers and Productive Efficiency, in: H. Fried, C.A.K. Lovell, S. Schmidt (Eds.) *The Measurement of Productive Efficiency: Techniques And Applications*, Oxford, Oxford University Press, 3-67.

[19] Markovitz, H. M. (1959), *Portfolio Selection: Efficient Diversification of Investments.* John Wiley and Sons, Inc., New York.

[20] Mouchart,M. and L. Simar (2002), Efficiency analysis of air navigation services provision: first insights, Consulting report, Institut de Statistique, Louvain-la-Neuve, Belgium.

[21] Park, B., Simar, L., Weiner, C. (2000), The FDH estimator for productivity efficiency scores: asymptotic properties, *Econometric Theory* 16, 855-877.

[22] Robson, D.S., Whitlock, J.H. (1964), Estimation of a truncation point, *Biometrika* 51, 33-39.

[23] Seiford, L.M. (1996), Data envelopment analysis: The evolution of the state-of-the-art (1978–1995),*Journal of Productivity Analysis*, 7, 2/3, 99–138.

[24] Seiford, L. M. (1997), A bibliography for data envelopment analysis (1978–1996), *Annals of Operations Research* 73, 393–438.

[25] Sengupta J.K. (1991). Maximum probability dominance and portfolio theory, *Journal of Optimization Theory and Applications*, 71, 341-357.

[26] Sengupta J.K. and Park H.S. (1993). Portfolio efficiency tests based on stochastic-dominance and cointegration. *International Journal of Systems Science*, 24, 2135-2158.

[27] Silverman, B.W. (1986), *Density estimation for statistics and data analysis,* Chapman and Hall, London.

[28] Simar, L., Wilson, P.W., (1998), Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science* 44(1), 49-61.

[29] Simar, L. and P. Wilson (1999), Estimating and Bootstrapping Malmquist Indices, *European Journal of Operational Research*, 115, 459–471.

[30] Simar, L., Wilson, P.W. (2000a), Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis* 13, 49-78.

[31] Simar, L., Wilson, P.W. (2000b), A general methodology for bootstrapping in non-parametric frontier models, *Journal of Applied Statistics*, Vol. 27, No. 6, 2000, 779-802.

[32] Simar L. and P. Wilson (2001), Testing Restrictions in Nonparametric Frontier Models, *Communications in Statistics, simulation and computation*, 30 (1), 161-186.

[33] Simar L. and P. Wilson (2002), Nonparametric Test of Return to Scale, *European Journal of Operational Research*, 139, 115–132.

[34] Simar, L., Wilson, P.W. (2003), Performance of the bootstrap for DEA estimators and iterating the principle, forthcoming in *Handbook on Data Envelopment Analysis*, W.W. Cooper and J. Zhu, editors, Kluwer.

[35] Swanepoel, J.W.H. (1986), A note on proving that the (modified) bootstrap works, *Comm. Statist. Theory Method* 15 (11), 3193-3203.

[36] Weissman, I. (1981), Confidence intervals for the threshold parameter, *Communications in Statistics, Theory and Method* 10(6), 549-557.

[37] Weissman, I. (1982), Confidence intervals for the threshold parameter II: unknown shape parameter, *Communications in Statistics, Theory and Method* 11(21), 2451-2474.

TABLE 1

Monte Carlo estimates of confidence intervals coverages
One Input, One Output ($p = q = 1$), $v \sim Exp(1)$, CRS

n	Significance level	Estimated coverage $\mathcal{J}_1$	Estimated coverage $\mathcal{J}_2$
10	0.8750	0.8895	0.9450
	0.9375	0.9505	0.9865
	0.9688	0.9695	0.9960
	0.9844	0.9860	0.9990
	0.9922	0.9930	0.9995
25	0.8750	0.8745	0.8930
	0.9375	0.9380	0.9555
	0.9688	0.9685	0.9820
	0.9844	0.9845	0.9940
	0.9922	0.9925	0.9980
50	0.8750	0.8835	0.8865
	0.9375	0.9455	0.9565
	0.9688	0.9745	0.9760
	0.9844	0.9860	0.9905
	0.9922	0.9935	0.9965
100	0.8750	0.8815	0.8795
	0.9375	0.9410	0.9395
	0.9688	0.9740	0.9745
	0.9844	0.9860	0.9905
	0.9922	0.9935	0.9945
200	0.8750	0.8740	0.8630
	0.9375	0.9375	0.9295
	0.9688	0.9730	0.9705
	0.9844	0.9860	0.9850
	0.9922	0.9940	0.9955
400	0.8750	0.8670	0.8735
	0.9375	0.9360	0.9370
	0.9688	0.9685	0.9725
	0.9844	0.9810	0.9865
	0.9922	0.9905	0.9940

TABLE 2

Monte Carlo estimates of confidence intervals coverages
One Input, One Output ($p = q = 1$), $v \sim N(0, 1)$, CRS

n	Significance level	Estimated coverage $\mathcal{J}_1$	Estimated coverage $\mathcal{J}_2$
10	0.8750	0.8940	0.9435
	0.9375	0.9480	0.9805
	0.9688	0.9745	0.9930
	0.9844	0.9870	0.9990
	0.9922	0.9950	1.0000
25	0.8750	0.8810	0.8970
	0.9375	0.9450	0.9595
	0.9688	0.9700	0.9830
	0.9844	0.9850	0.9945
	0.9922	0.9910	0.9985
50	0.8750	0.8865	0.8890
	0.9375	0.9435	0.9580
	0.9688	0.9725	0.9795
	0.9844	0.9890	0.9915
	0.9922	0.9935	0.9955
100	0.8750	0.8755	0.8835
	0.9375	0.9360	0.9400
	0.9688	0.9735	0.9695
	0.9844	0.9845	0.9865
	0.9922	0.9935	0.9935
200	0.8750	0.8800	0.8750
	0.9375	0.9365	0.9355
	0.9688	0.9650	0.9695
	0.9844	0.9825	0.9890
	0.9922	0.9930	0.9925
400	0.8750	0.8815	0.8755
	0.9375	0.9395	0.9420
	0.9688	0.9695	0.9715
	0.9844	0.9810	0.9855
	0.9922	0.9915	0.9920

TABLE 3

Monte Carlo estimates of confidence intervals coverages
Two Inputs, One Output ($p = 2$, $q = 1$), $B = 200$, CRS

n	Significance level	Estimated coverage	Average widths
10	0.8750	0.7515	0.1845
	0.9375	0.8695	0.2401
	0.9688	0.9320	0.3030
	0.9844	0.9710	0.3791
	0.9922	0.9895	0.4808
25	0.8750	0.7620	0.0826
	0.9375	0.8570	0.1018
	0.9688	0.9230	0.1215
	0.9844	0.9550	0.1418
	0.9922	0.9725	0.1627
50	0.8750	0.7840	0.0503
	0.9375	0.8600	0.0593
	0.9688	0.9200	0.0684
	0.9844	0.9555	0.0776
	0.9922	0.9740	0.0870
100	0.8750	0.8280	0.0337
	0.9375	0.8810	0.0381
	0.9688	0.9295	0.0426
	0.9844	0.9570	0.0472
	0.9922	0.9750	0.0518
200	0.8750	0.8870	0.0235
	0.9375	0.9195	0.0257
	0.9688	0.9465	0.0279
	0.9844	0.9690	0.0302
	0.9922	0.9805	0.0324
400	0.8750	0.9445	0.0169
	0.9375	0.9640	0.0180
	0.9688	0.9755	0.0191
	0.9844	0.9840	0.0203
	0.9922	0.9875	0.0214

TABLE 4

The Air Controlers efficiency levels and confidence intervals of level 0.9688. The bandwidth for the Loh approach is 0.0996. For the bootstrap approach, the homogeneous bootstrap from Simar and Wilson (1998) has been used (2000 replications).

Units	Eff. Scores	Loh Lower	Loh Upper	Boot. Lower	Boot. Upper
1	0.6458	0.5389	0.6458	0.5689	0.6458
2	0.4862	0.4057	0.4862	0.4286	0.4862
3	0.5333	0.4450	0.5333	0.4682	0.5333
4	0.4694	0.3917	0.4694	0.4130	0.4694
5	0.5340	0.4456	0.5340	0.4707	0.5340
6	0.7174	0.5986	0.7174	0.6328	0.7174
7	1.0000	0.8344	1.0000	0.8803	1.0000
8	0.9366	0.7815	0.9366	0.8265	0.9366
9	1.0000	0.8344	1.0000	0.8701	1.0000
10	0.7249	0.6049	0.7249	0.6369	0.7249
11	0.6137	0.5121	0.6137	0.5409	0.6137
12	0.7594	0.6337	0.7594	0.6661	0.7594
13	0.7658	0.6390	0.7658	0.6736	0.7658
14	1.0000	0.8344	1.0000	0.8677	1.0000
15	0.6633	0.5535	0.6633	0.5818	0.6633
16	0.7936	0.6622	0.7936	0.6972	0.7936
17	0.6573	0.5485	0.6573	0.5791	0.6573
18	0.3553	0.2965	0.3553	0.3127	0.3553
19	0.4344	0.3625	0.4344	0.3822	0.4344
20	0.4213	0.3515	0.4213	0.3715	0.4213
21	0.6838	0.5706	0.6838	0.6008	0.6838
22	0.3029	0.2527	0.3029	0.2663	0.3029
23	0.5218	0.4354	0.5218	0.4601	0.5218
24	0.4355	0.3634	0.4355	0.3832	0.4355
25	0.9375	0.7823	0.9375	0.8258	0.9375
26	1.0000	0.8344	1.0000	0.8696	1.0000
27	0.6592	0.5501	0.6592	0.5798	0.6592
28	0.8472	0.7069	0.8472	0.7472	0.8472
29	0.6439	0.5373	0.6439	0.5669	0.6439
30	0.6832	0.5701	0.6832	0.5998	0.6832
31	0.5750	0.4798	0.5750	0.5070	0.5750
32	0.8725	0.7280	0.8725	0.7671	0.8725
33	0.8196	0.6839	0.8196	0.7229	0.8196
34	0.6632	0.5534	0.6632	0.5852	0.6632
35	0.9253	0.7721	0.9253	0.8152	0.9253
36	0.8309	0.6933	0.8309	0.7317	0.8309

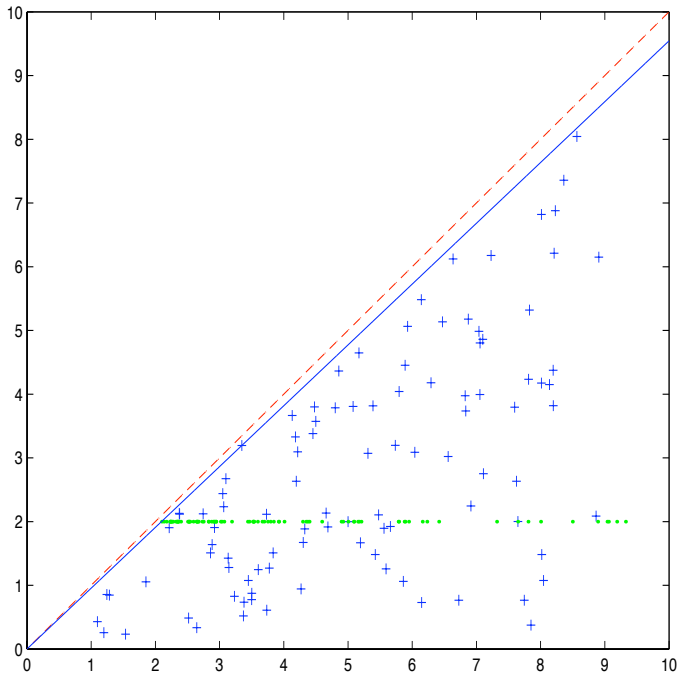

Figure 1: Sample of 100 observations generated from model (4.1) and their projections according to (3.11) for $(x_0, y_0) = (5, 2)$. Dashed line - true frontier, solid line - DEA frontier.

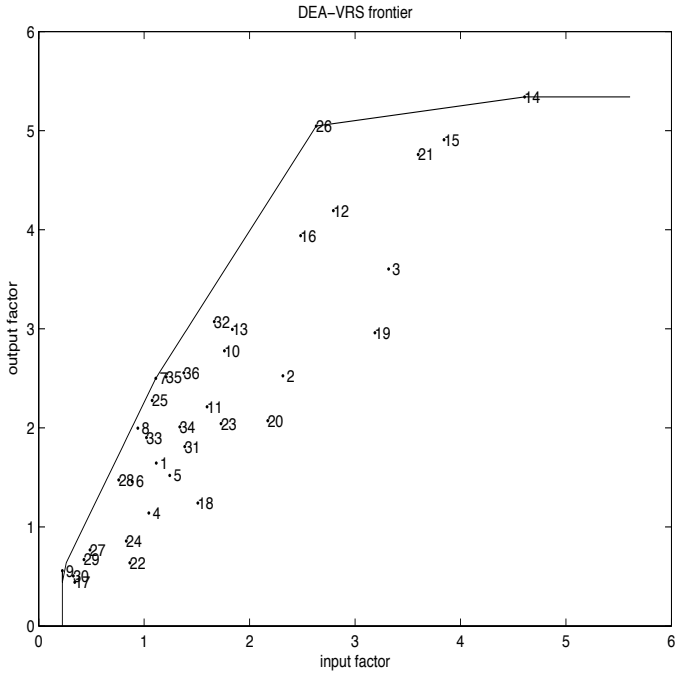

Figure 2: Data on the Air Controlers: the solid line is the VRS-DEA frontier.