UCL Université catholique de Louvain

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

ISBA

DISCUSSION PAPER

2018/20

Improving Finite Sample Approximation by Central Limit Theorems for DEA and FDH efficiency scores

Simar, L. and V. Zelenyuk

Improving Finite Sample Approximation by Central Limit Theorems for DEA and FDH efficiency scores

Léopold Simar* Valentin Zelenyuk§

leopold.simar@uclouvain.be v.zelenyuk@uq.edu.au

August 20, 2018

Abstract

We propose an improvement of the finite sample approximation of the central limit theorems (CLTs) that were recently derived for statistics involving production efficiency scores estimated via Data Envelopment Analysis (DEA) or Free Disposal Hull (FDH) approaches. The improvement is very easy to implement since it involves a simple correction of the already employed statistics without any additional computational burden and preserves the original asymptotic results such as consistency and asymptotic normality. The proposed approach persistently showed improvement in all the scenarios that we tried in various Monte-Carlo experiments, especially for relatively small samples or relatively large dimensions (measured by total number of inputs and outputs) of the underlying production model. This approach therefore is expected to be valuable (and at almost no additional computational costs) for practitioners wishing to perform statistical inference about production efficiency using DEA or FDH approaches.

Key words: Data Envelopment Analysis, DEA; Free Disposal Hull, FDH; Statistical Inference; Production Efficiency; Productivity.

*Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium.

§School of Economics and Centre for Efficiency and Productivity Analysis (CEPA), University of Queensland, Colin Clark Building (39), St Lucia, Brisbane, Qld 4072, Australia. The author acknowledges the financial support from ARC Future Fellowship grant (FT170100401).

1 Introduction

Nonparametric envelopment estimators are used by applied researchers in efficiency and productivity analysis to provide measures of producers' performances. These estimators are obtained by enveloping the clouds of points obtained by an observed sample of input and output levels of firms. The most popular estimators of this type are known as data envelopment analysis (DEA) if we assume the convexity of the set of attainable combinations of inputs and outputs (see Farrell, 1957 and Charnes, Cooper and Rhodes, 1978) and the free disposal hull (FDH) estimators that does not impose convexity, but only a free disposability assumption (see Deprins, Simar and Tulkens, 1984). These estimators found their use in many areas of research in production economics, operations research and management science, from both theoretical and applied angles and published in the leading journals of these fields, including this journal (e.g., see Gorman and Ruggiero (2008) and Chowdhury et al. (2014) for some recent examples and Olesen (1995) and Førsund and Sarafoglou (2005) for comprehensive reviews and references therein).

A large literature has been devoted to the statistical properties of DEA and FDH estimators of efficiency for a fixed point; see Simar and Wilson (2013, 2015) for comprehensive and recent surveys and the references therein. It is there explained that inference for a single measure of efficiency of a particular firm is available by using appropriate bootstrap techniques; see Kneip, Simar and Wilson (2008, 2011) and Simar and Wilson (2011) for the DEA case and Jeong and Simar (2006) for the FDH case. But in many situations, the applied researcher is willing to compare means of group of firms, or he might be interested in testing hypothesis on the attainable set (convexity, returns to scale, etc.) or in dynamic settings, he might be willing to analyze changes in productivity over time, etc. In all these cases, the inference will be based on statistics which are appropriate functions of the efficiency scores over the sample.

The pioneering work of Kneip, Simar and Wilson (2015) (hereafter KSW) has shown that inference on simple sample means of estimated scores is problematic unless the dimension of the problem (the total number of inputs and outputs) is very small (e.g. as small as 1 for the FDH case and 2 for the most common DEA estimates). The basic problem comes from the fact that the DEA/FDH efficiency estimates have a bias that does not vanish fast enough when the sample size increases. However, KSW provide central limit theorems (CLT) based on a bias correction method (using a generalized jackknife) and, if not enough (depending on the chosen estimator and on the dimension of the problem), using means of the efficiency estimates on a subsample of points. The important result is that, when using the appropriate statistics, regular quantiles of the standard normal can be used to derive confidence intervals or to compute p-values.

These basic results have been used in KSW for providing confidence intervals on mean

of efficiency scores in a group, and adapted in various testing situations (testing equality of means between two groups of firms, testing convexity and testing returns to scale) in Kneip, Simar and Wilson (2016). It was then extended to inference for aggregate efficiency in Simar and Zelenyuk (2018), and to conditional measures of efficiency, with a test of the separability condition in Daraio, Simar and Wilson (2018). Recently similar CLT results have been derived for Malmquist indices (Kneip, Simar and Wilson, 2018) and for cost and allocative efficiency in Simar and Wilson (2018).

In all these papers, the authors show with intensive Monte-Carlo experiments that the CLT can be applied in many practical situations and that the results behave as expected: improving the accuracy of the normal approximations (size and power of the tests, or coverage of the resulting confidence intervals) when the sample size increases, but the accuracy may reduce when the dimension of the problem increases. In some situations, these results might be disappointing for the practitioner.

In this paper we suggest an "easy to compute" correction of the used statistics, that does not require any additional computational burden and which improves in all the situations the accuracy of the normal approximation provided by the corresponding CLT. We will describe how the correction is motivated and derived in the simplest case of an average of efficiency scores (which cover all the extensions mentioned above) and in the case of aggregate efficiency, where the correction requires more care. We illustrate how this work in practice in these two cases, with several Monte-Carlo experiments (different dimensions and sample sizes). In some cases, the improvements may be substantial.

The paper is organized as follows. The next section introduces the basic notions and notations and summarizes the background (i.e. the results from KSW). The Section 3 gives the basic idea of our corrected statistics in the simplest case of a simple average of efficiency scores then Section 4 shows how to adapt this basic idea to other situations, with some emphasis on the case of aggregate efficiency. Section 5 indicates through Monte-Carlo experiments how this works in practice.

2 Theoretical Background

Assume that each firm uses p inputs $x \in \mathbb{R}^p_+$ to produce a vector of q outputs, denoted by $y \in \mathbb{R}^q_+$ and that the technology set, Ψ is defined as

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q}_+ \mid x \text{ can produce } y\}, \tag{2.1}$$

which gives the set of combinations of inputs and outputs that are technologically feasible. The *technology*, or *efficient frontier* of Ψ, is given by

$$\Psi^{\partial} = \left\{(x,y) \in \Psi \mid \left(\gamma^{-1}x, \gamma y\right) \notin \Psi \text{ for all } \gamma > 1\right\}. \tag{2.2}$$

We assume the usual regularity conditions on technology in production theory (e.g., see Färe and Primont, 1995). Specifically, the key assumptions we need are:

Assumption 2.1 *Ψ is closed and Ψ^{∂} exists.*

Assumption 2.2 *All outputs and all inputs are strongly disposable, i.e., $(x_o, y_o) \in \Psi \Rightarrow (x,y) \in \Psi \ \forall x \geqq x_o, y \leqq y_o$.*

For the sake of brevity we will focus on the case when efficiency is measured via the so-called Farrell-Debreu output oriented measure of technical efficiency (similar results can be derived for other orientations). For a firm with an input-output allocation (x,y), the output efficiency is defined as

$$\lambda(x,y) = \sup\{\theta > 0 \mid (x, \theta y) \in \Psi\}. \tag{2.3}$$

Many methods have been developed in the literature to estimate the technology frontier Ψ^{∂} and the related values of efficiency scores, $\lambda(x,y)$ from a sample of n i.i.d. observations on inputs and outputs $\mathcal{X}_n = \{(X_i, Y_i) \mid i = 1, \ldots, n\}$. Here we will focus on the DEA and FDH nonparametric and consistent estimators that envelop the data under certain constraints on technology. Specifically, if one wishes to assume Constant Returns to Scale (CRS) for Ψ^{∂}, then the CRS-DEA estimator is given by

$$\widehat{\lambda}(x,y \mid \mathcal{X}_n) \equiv \max_{\xi_1,\ldots,\xi_n,\theta} \{\theta \ : \ \sum_{k=1}^{n} \xi_k Y_k \geqq \theta\, y, \ \sum_{k=1}^{n} \xi_k X_k \leqq x, \ \theta \geq 0, \xi_k \geq 0\}, \tag{2.4}$$

inspired by Farrell (1957) and popularized as a linear program (LP) by Charnes, Cooper and Rhodes (1978).[1] Alternatively, if one wishes to assume Variable Returns to Scale (VRS) for Ψ^{∂}, then the VRS-DEA estimator is given by

$$\begin{aligned}\widehat{\lambda}(x,y \mid \mathcal{X}_n) \equiv \max_{\xi_1,\ldots,\xi_n,\theta} \{\theta \ : \ & \sum_{k=1}^{n} \xi_k Y_k \geqq \theta\, y, \ \sum_{k=1}^{n} \xi_k X_k \leqq x, \\ & \theta \geq 0, \ \xi_k \geq 0, \ \sum_{k=1}^{n} \xi_k = 1\}.\end{aligned} \tag{2.5}$$

[1] To be precise, both papers focused on the input orientation, which here (due to CRS) is the reciprocal of the output orientation problem (2.4).

i.e., constraint $\sum_{k=1}^{n} \xi_k = 1$ is added to the CRS-DEA specification (2.4).

If convexity of Ψ is not assumed, but only strong disposability of inputs and outputs, then one can use the FDH estimator, introduced by Deprins, Simar and Tulkens (1984):

$$\widehat{\lambda}(x, y \mid \mathcal{X}_n) \equiv \max_{\theta} \{\theta \ : \ Y_k \geqq \theta\, y, \ X_k \leqq x, \ k = 1, \ldots, n, \ \theta \geq 0\}. \tag{2.6}$$

As pointed out above, the asymptotic statistical properties of DEA and FDH estimators for $\lambda(x, y)$ have been established when evaluated at fixed points (x, y), and bootstrap techniques have to be used for practical inference due to the complex nature of the asymptotic distribution of the estimators (see Simar and Wilson, 2015, for details and references).

When confidence intervals for means or testing issues about the shape of Ψ^{∂} are concerned, the practitioner needs the statistical properties of some statistics built from the estimators of efficiency scores at the random points. The simplest statistic one may consider is the mean of the efficiency scores to draw inference on μ_λ, the average efficiency score in the economy

$$\mu_\lambda = \mathrm{E}\big(\lambda(X, Y)\big), \tag{2.7}$$

where the expectation is over the random variables (X, Y). Below we will also need to estimate the variance σ^2_λ of the efficiency scores in the population:

$$\sigma^2_\lambda = \mathrm{Var}\big(\lambda(X, Y)\big). \tag{2.8}$$

Inference on μ_λ could be obtained from the sample mean of the estimated efficiency scores:

$$\overline{\lambda}_n = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n), \tag{2.9}$$

where one of the nonparametric estimators above would be evaluated at the sample points (X_i, Y_i). However, the basic results in KSW indicate that this is not so easy and that simple CLT is not directly available as we now briefly explain.

The basic results from KSW can be summarized as follows: for the FDH and DEA estimators, under some regularity conditions and as $n \to \infty$,

$$\mathrm{E}\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \lambda(X_i, Y_i)\right) = C n^{-\kappa} + R_{n,\kappa} \tag{2.10}$$

$$\mathrm{E}\left(\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \lambda(X_i, Y_i)\right)^2\right) = o\big(n^{-\kappa}\big), \tag{2.11}$$

$$\left|\mathrm{COV}\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \lambda(X_i, Y_i), \ \widehat{\lambda}(X_j, Y_j \mid \mathcal{X}_n) - \lambda(X_j, Y_j)\right)\right| = o\big(n^{-1}\big) \tag{2.12}$$

for some finite constant C and for all $i, j \in \{1, \ldots, n\}$, $i \neq j$. Note that the values of the constant C, the rate κ, and the remainder term $R_{n,\kappa}$ depends on which estimator is used. Specifically, when we assume CRS, the CRS-DEA estimator (2.4) achieve the rate $\kappa = 2/(p+q)$. When we assume VRS, the VRS-DEA estimator (2.5) achieves the rate $\kappa = 2/(p+q+1)$. When we do not assume convexity and maintain only the strong disposability assumption, the appropriate estimator is the FDH estimator (2.6). It achieves the rate $\kappa = 1/(p+q)$. Specific rates are available for the remainder term $R_{n,k}$ but for our purpose here it is sufficient to note that in all the case it is a term of order $o\left(n^{-\kappa}\right)$.

Due to these results KSW showed that under certain regularity assumptions, as $n \to \infty$,

$$\sqrt{n}\big(\overline{\lambda}_n - \mu_\lambda - Cn^{-\kappa} + o(n^{-\kappa})\big) \xrightarrow{\mathcal{L}} N(0, \sigma_\lambda^2), \tag{2.13}$$

from which we understand why we have a problem when $\kappa \leq 1/2$, e.g., for CRS-DEA it happens already when $p+q > 3$ for VRS-DEA when $p+q > 2$ and for FDH when $p+q > 1$!

The solution suggested by KSW is to correct for the leading term in the inherent bias of DEA and FDH, $Cn^{-\kappa}$, via a consistent estimator defined as

$$\widehat{B}_{n,\kappa} = (2^\kappa - 1)^{-1}(\overline{\lambda}^*_{n/2} - \overline{\lambda}_n), \tag{2.14}$$

where $\overline{\lambda}^*_{n/2}$ is a generalized jackknife analog of $\overline{\lambda}_n$. It is obtained as

$$\overline{\lambda}^*_{n/2} = \big(\overline{\lambda}^{(1)}_{n/2} + \overline{\lambda}^{(2)}_{n/2}\big)/2, \tag{2.15}$$

where $\overline{\lambda}^{(1)}_{n/2}$ is a version of $\overline{\lambda}_n$ based on a random subset of $\mathcal{X}_n$ of size $n/2$ and $\overline{\lambda}^{(2)}_{n/2}$ is the analog based on the remaining part of the sample (for simplicity assume n is even). Formally for $\ell = 1, 2$

$$\overline{\lambda}^{(\ell)}_{n/2} = 2n^{-1} \sum_{\{i|(X_i,Y_i)\in\mathcal{X}^{(\ell)}_{n/2}\}} \widehat{\lambda}(X_i, Y_i \mid \mathcal{X}^{(\ell)}_{n/2}),$$

where $\mathcal{X}^{(1)}_{n/2} \cap \mathcal{X}^{(2)}_{n/2} = \emptyset$ and $\mathcal{X}^{(1)}_{n/2} \cup \mathcal{X}^{(2)}_{n/2} = \mathcal{X}_n$.[2] It is shown in KSW that

$$\widehat{B}_{n,\kappa} = Cn^{-\kappa} + R'_{n,\kappa} + o_p\big(n^{-1/2}\big), \tag{2.16}$$

where $R'_{n,\kappa}$ has the same order as the original $R_{n,\kappa}$ introduced above.

With these results KSW obtain the following results

[2]As suggested in Kneip et al. (2016), this operation of random split in two parts is repeated a large number, L, of times by shuffling the observations in $\mathcal{X}_n$ before each split providing $\widehat{B}^{(l)}_{n,\kappa}$ for $l = 1, \ldots, L$. Then to reduce the variance of the bias estimator we use $\widehat{B}_{n,\kappa} = L^{-1}\sum_{l=1}^{L} \widehat{B}^{(l)}_{n,\kappa}$.

Theorem 2.1 *Under assumptions for Theorem 3.1, 3.2 or 3.3 of Kneip et al. (2015), for $p+q \leq 5$ if CRS-DEA estimator is used and Ψ^{∂} is globally CRS and convex, for $p+q \leq 4$ if VRS-DEA estimator is used and Ψ^{∂} is convex, for $p+q \leq 3$ if FDH estimator is used and Ψ^{∂} satisfies free disposability of inputs and outputs, when $n \to \infty$ we have*

$$\sqrt{n}\left(\overline{\lambda}_n - \widehat{B}_{n,\kappa} - \mu_\lambda + R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\lambda^2), \tag{2.17}$$

When $\kappa < 1/2$ then, as $n \to \infty$ we have

$$\sqrt{n_\kappa}\left(\overline{\lambda}_{n_\kappa} - \widehat{B}_{n,\kappa} - \mu_\lambda + R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\lambda^2), \tag{2.18}$$

with $\overline{\lambda}_{n_\kappa}$ being a subsample version of $\overline{\lambda}_n$, where the averaging is taken over a random subsample $\mathcal{X}_{n_\kappa}^* \subset \mathcal{X}_n$ of size $n_\kappa = \lfloor n^{2\kappa} \rfloor < n$. Formally

$$\overline{\lambda}_{n_\kappa} = n_\kappa^{-1} \sum_{\{j|(X_j,Y_j)\in\mathcal{X}_{n_\kappa}^*\}} \widehat{\lambda}(X_j, Y_j \mid \mathcal{X}_n). \tag{2.19}$$

Note that the bounds on $p+q$ given for being able to apply (2.17) are due to the particular definitions of $R_{n,\kappa}$ for each estimator (see KSW for details). We remark that in all the cases, since $R_{n,\kappa} = o(n^{-\kappa})$, the remainder term can be neglected. It should be noticed that for $p+q=5$ in the CRS case, $p+q=4$ in the VRS case and $p+q=3$ for the FDH case, both versions of the CLT are applicable. As commented in KSW, looking to the peculiar form of the remainder, the second version of the CLT in these limiting cases is preferable, because the neglected terms are of small order $O(\cdot)$. This is confirmed by the Monte-Carlo experiments (see KSW and our simulation results below for details).

For practical inference σ_λ^2 can be replaced by a consistent estimator. KSW suggest the following one

$$\widehat{\sigma}_{\lambda,n}^2 = \frac{1}{n}\sum_{j=1}^{n}(\widehat{\lambda}(X_j, Y_j|\mathcal{X}_n) - \overline{\lambda}_n)^2, \tag{2.20}$$

thus providing, when (2.17) can be applied, an asymptotically correct $(1-\alpha)$ confidence interval for μ_λ given by

$$\left[\overline{\lambda}_n - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\widehat{\sigma}_{\lambda,n}/\sqrt{n}\right], \tag{2.21}$$

where $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution. When the dimension $p+q$ increases and $\kappa < 1/2$, an asymptotically correct $(1-\alpha)$ confidence interval for μ_λ is given by

$$\left[\overline{\lambda}_{n_\kappa} - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\widehat{\sigma}_{\lambda,n}/\sqrt{n_\kappa}\right]. \tag{2.22}$$

KSW also present some MC evidence, supporting the theory and also suggesting that the

application of these new theorems require fairly large samples, even for modest dimensions of the production model. For instance, when $p+q=4$ with CRS-DEA (i.e., the fastest) estimator, the coverage of true values by 90%-confidence intervals estimated based on this theorem was only about 0.828 even when $n=1000$. For the FDH (with a different DGP but same dimension $p+q=4$) the coverage is about 0.576. So our goal is to try to improve the approximations provided by this important theorem for finite samples, which in this context also includes fairly large samples like $n=1000$.

3 Improving Approximations

As pointed above, for practical inference, one needs a consistent estimator of σ_λ^2, the variance of the efficiencies in the population. We know that $\sigma_\lambda^2 = \mathrm{E}(\lambda(X,Y)-\mu_\lambda)^2$ and KSW have shown that $\widehat{\sigma}_{\lambda,n}^2$ is such estimator.

Looking to this choice, it appears to be an empirical version of $\mathrm{Var}\big(\widehat{\lambda}(X,Y \mid \mathcal{X}_n)\big)$, but by construction, we know that this estimator underestimates $\sigma_\lambda^2 = \mathrm{Var}\big(\lambda(X,Y)\big)$. Indeed $\lambda(X,Y) = \sup\{\theta \mid (X,\theta Y) \in \Psi\}$ whereas $\widehat{\lambda}(X,Y \mid \mathcal{X}_n) = \sup\{\theta \mid (X,\theta Y) \in \widehat{\Psi}\}$, where, according to the chosen estimator, $\widehat{\Psi}$ is the conical convex hull (CRS-DEA case), or the convex hull (VRS-DEA case) or the free disposal hull (FDH case) of the clouds of points $\mathcal{X}_n$, implicitly defined in the expressions (2.4) to (2.6) above. But in all the cases, we have

$$\widehat{\Psi} \subset \Psi$$

implying, with probability one, that

$$1 \leq \widehat{\lambda}(X,Y \mid \mathcal{X}_n) \leq \lambda(X,Y),$$

providing random variables with lower variance.[3] Note also that in practice, many $\widehat{\lambda}(X_i,Y_i \mid \mathcal{X}_n) = 1$, still reducing the variability of the estimators. This is a part of the source of the lower coverages than the nominal one, that we observed in all the Monte-Carlo experiments.

Our goal therefore is to provide and explore another consistent estimator of σ_λ^2 that should at least partially correct for this disappointing property. The motivation for the new variance estimator goes along the following lines.

A natural estimator of σ_λ^2 should be its empirical version (if available), i.e.

$$\widetilde{\sigma}_\lambda^2 := \frac{1}{n}\sum_{i=1}^{n}(\lambda(X_i,Y_i) - \widetilde{\lambda}_n)^2$$

[3]Similar inequalities hold for the input efficiency scores, where $0 < \theta(X,Y) \leq \widehat{\theta}(X,Y \mid \mathcal{X}_n) \leq 1$.

where

$$\widetilde{\lambda}_n := \frac{1}{n}\sum_{i=1}^{n}\lambda(X_i, Y_i)$$

but the problem of course is that $\lambda(X_i, Y_i)$ is unavailable and so $\widetilde{\lambda}_n$ and $\widetilde{\sigma}^2_\lambda$ are also unavailable–we only know from existing asymptotic theory (see KSW), that

$$\widetilde{\lambda}_n = \mu_\lambda + O_p\big(n^{-1/2}\big).$$

A natural approach (taken by KSW) is to replace the unavailable $\lambda(X_i, Y_i)$ by their consistent estimates $\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n)$ and the sample mean $\widetilde{\lambda}_n$ by $\overline{\lambda}_n$, but we know (as a consequence of Lemma 4.1 in KSW) that

$$\overline{\lambda}_n = \mu_\lambda + B_{n,\kappa} + o(n^{-\kappa})$$

where the leading term for the bias is given by

$$B_{n,\kappa} = Cn^{-\kappa}$$

as described above. Now, by (2.16) we have a consistent estimator of this bias term which is already computed for deriving the CLT. So, we suggest to replace the unavailable $\widetilde{\lambda}_n$ by its bias corrected version, defined as

$$\widetilde{\lambda}^{bc}_n := \overline{\lambda}_n - \widehat{B}_{n,\kappa}$$

(instead of $\overline{\lambda}_n$ as in KSW).

So at the end, the new consistent estimator of σ^2_λ we propose is given by

$$\begin{aligned}\widehat{\widetilde{\sigma}}^2_{\lambda,n} &= \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \widetilde{\lambda}^{bc}_n\big)^2. \\ &= \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \overline{\lambda}_n + \widehat{B}_{n,\kappa}\big)^2\end{aligned} \tag{3.1}$$

It is easy to check, using (2.16), that

$$\widetilde{\lambda}^{bc}_n = \mu_\lambda + o(n^{-\kappa}) + o_p(n^{-1/2}),$$

meaning that this correction will preserve the existing asymptotic properties (consistency and asymptotic normality) derived by KSW, when $\widehat{\widetilde{\sigma}}^2_{\lambda,n}$ is used in place of unknown σ^2_λ, and has a chance of improving its approximation in practice, i.e., for finite samples (simulated or

real data), relative to cases when $\widehat{\sigma}^2_{\lambda,n}$ is used.

The resulting estimated $(1-\alpha)$ confidence intervals for μ_λ will therefore be

$$\left[\overline{\lambda}_n - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\widehat{\widehat{\sigma}}_{\lambda,n}/\sqrt{n}\right], \tag{3.2}$$

when (2.17) can be applied, and

$$\left[\overline{\lambda}_{n_\kappa} - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\widehat{\widehat{\sigma}}_{\lambda,n}/\sqrt{n_\kappa}\right] \tag{3.3}$$

when (2.18) can be applied, and where as before, $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution.

Also note that after some elementary manipulations we can simplify the expression for the new variance estimator to be

$$\widehat{\widehat{\sigma}}^2_{\lambda,n} = \widehat{\sigma}^2_{\lambda,n} + \widehat{B}^2_{n,\kappa}, \tag{3.4}$$

and so one can immediately see that, as expected, we increase slightly the original KSW estimator ($\widehat{\sigma}^2_{\lambda,n}$) with a positive term, $\widehat{B}^2_{n,\kappa}$, which is of small order, $O_p\big(n^{-2\kappa}\big)$. So, as $n \to \infty$, this correction does not play any role asymptotically, but in finite samples the bias may be so big that the difference is significant (in particular when κ decreases, i.e. when the dimension $p+q$ increases). We will investigate the practical consequences in some Monte-Carlo experiments, confirming the usefulness of this correction, even with sample sizes as large as $n = 1000$.

4 Implications to Other Results

4.1 Obvious Extensions

The basic results of KSW have been extended in various situations, Kneip et al. (2016) indicate how to use these CLTs for testing the equality of the means of two groups of firms, for testing the convexity of the attainable set and for testing returns to scale of the technology. All these tests are built on a statistic which is the difference of two sample means of efficiency scores computed on two independent random samples. Then the above CLTs can be used for both sample means, corrected for the bias term and rescaled by an estimator of the variance of the efficiency scores. It is clear that the accuracy of the procedures will be improved by using in each case the corrected estimator defined in (3.4), adapted to each particular case.

The same applies to the inference suggested in Kneip et al. (2018), where CLTs are derived for Malmquist indices, for the CLTs for conditional efficiency scores derived in Daraio et al. (2018) (and for their test of the separability condition), and all the CLTs derived in

Simar and Wilson (2018) for various cost, revenue and allocative efficiencies. All the CLTs there have a similar shape than the ones derived in KSW (and summarized above) with a bias correction factor and a rescaling by an estimator of the variance of the efficiency scores.

To save space we do not give the details for all these extensions because they are rather obvious to derive. The derivations are a bit more tedious for aggregate efficiencies as explained in the next section.

4.2 Aggregate Efficiency

Simar and Zelenyuk (2018) (hereafter SZ) adapt and generalize the results from KSW to the case of CLT for aggregate or weighted efficiency (e.g., such as industry efficiency), which can be defined in terms of a ratio of two true means. They focus their presentation on the output aggregate technical efficiencies of a group of firms derived in Färe and Zelenyuk (2003) which rest on certain assumptions on the aggregate technology and the law of one price $w \in \mathbb{R}^q_{++}$ on output markets. The output aggregate technical efficiency over a group of n firms, can be written as

$$\tau_n = \frac{n^{-1}\sum_{i=1}^n w^T Y_i^\partial}{n^{-1}\sum_{i=1}^n w^T Y_i}, \tag{4.1}$$

where Y_i^∂ is the counterfactual output vector for the firm i obtained by its radial projection on the efficient frontier Ψ^∂, i.e. $Y_i^\partial = \lambda(X_i, Y_i)Y_i$. Let us denote $Z_i = w^T Y_i$ and $Z_i^\partial = \lambda(X_i, Y_i)Z_i$.

SZ consider the ratio of the true means, $\tau = \mu_1/\mu_2$, as the parameter of interest, where $\mu_1 = \mathrm{E}\big(Z_i^\partial\big)$ and $\mu_2 = \mathrm{E}\big(Z_i\big)$, and then by standard arguments show that

$$\sqrt{n}\big(\tau_n - \tau\big) \xrightarrow{\mathcal{L}} N(0, \sigma_\tau^2), \tag{4.2}$$

where

$$\sigma_\tau^2 = \tau^2\left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} - 2\frac{\sigma_{12}}{\mu_1\mu_2}\right), \tag{4.3}$$

and where $\sigma_1^2 = \mathrm{Var}(Z_i^\partial)$, $\sigma_2^2 = \mathrm{Var}(Z_i)$ and $\sigma_{12} = \mathrm{COV}(Z_i^\partial, Z_i)$. Since τ_n is not available, we will estimate τ by using our nonparametric frontier estimators to obtain

$$\widehat{\tau}_n = \frac{\widehat{\mu}_{1,n}}{\widehat{\mu}_{2,n}} = \frac{(1/n)\sum_{i=1}^n \widehat{Z}_i^\partial}{(1/n)\sum_{i=1}^n Z_i}, \tag{4.4}$$

where $\widehat{Z}_i^\partial = Z_i \times \widehat{\lambda}(X_i, Y_i|\mathcal{X}_n)$. Extending the theory of KSW, they derive an estimator of the bias of $\widehat{\tau}_n$ and a consistent estimator of its variance providing the CLT which is summarized below.

The bias estimator deserves some details, the leading part of the bias is governed by the

inherent bias of the numerator due to the nonparametric estimators of $\lambda(X_i, Y_i)$. As in KSW, the generalized jackknife is used for the estimation of the bias of the numerator only, i.e. $B_{\mu_{1,n},\kappa}$. By repeating the random splits in two parts L times (see above (2.14) and Footnote 2) we end up with the bias correction for $\widehat{\tau}_n$ given by

$$\widehat{B}_{\tau_n,\kappa} = \frac{\widehat{B}_{\mu_{1,n},\kappa}}{\widehat{\mu}_{2,n}} = \frac{L^{-1}\sum_{\ell=1}^{L}(2^\kappa - 1)^{-1}\big(\widehat{\mu}^*_{1,n,\ell} - \widehat{\mu}_{1,n}\big)}{\widehat{\mu}_{2,n}}, \tag{4.5}$$

where $\widehat{\mu}^*_{1,n,\ell}$ is the analog of $\overline{\lambda}^*_{n/2}$, as defined in (2.14) and (2.15), and where the use of $\widehat{\mu}_{2,n}$ at the denominator does not change the order of the approximation (see SZ for details). They obtain the following CLT.

Theorem 4.1 *Under the appropriate set of Assumptions described in Theorem 3.1, 3.2 or 3.3 of Kneip et al. (2015), for $p+q \leq 5$ if CRS-DEA estimator is used and Ψ^∂ is globally CRS and convex, for $p+q \leq 4$ if VRS-DEA estimator is used and Ψ^∂ is convex, for $p+q \leq 3$ if FDH estimator is used and Ψ^∂ satisfies free disposability of inputs and outputs, when $n \to \infty$ we have*

$$\sqrt{n}\ \Big(\widehat{\tau}_n - \widehat{B}_{\tau_n,\kappa} - \tau + R_{n,\kappa}\Big) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\tau), \tag{4.6}$$

where $R_{n,\kappa} = o\big(n^{-\kappa}\big)$.

For $\kappa < 1/2$, then as $n \to \infty$ we have

$$\sqrt{n_\kappa}\ \Big(\widehat{\tau}_{n_\kappa} - \widehat{B}_{\tau_n,\kappa} - \tau + R_{n,\kappa}\Big) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\tau), \tag{4.7}$$

with $\widehat{\tau}_{n_\kappa}$ being a subsample version of $\widehat{\tau}_n$, in the sense that the averages are taken over a random subsample $\mathcal{X}^*_{n_\kappa} \subset \mathcal{X}_n$ of size $n_\kappa = \lfloor n^{2\kappa} \rfloor < n$. Formally

$$\widehat{\tau}_{n_\kappa} = \frac{n_\kappa^{-1}\sum_{\{j|(X_j,Y_j)\in\mathcal{X}^*_{n_\kappa}\}}\widehat{Z}^\partial_j}{n_\kappa^{-1}\sum_{\{j|(X_j,Y_j)\in\mathcal{X}^*_{n_\kappa}\}} Z_j}, \tag{4.8}$$

where $\widehat{Z}^\partial_j = \widehat{\lambda}(X_j, Y_j|\mathcal{X}_n)Z_j$.

The remark after Theorem 2.1 above is still valid here. For practical inference, SZ propose to estimate σ^2_τ by plugging estimators of its components in equation (4.3), i.e.

$$\widehat{\sigma}^2_{\tau,n} = \widehat{\tau}^2_n\left[\frac{\widehat{\sigma}^2_{1,n}}{\widehat{\mu}^2_{1,n}} + \frac{\widehat{\sigma}^2_{2,n}}{\widehat{\mu}^2_{2,n}} - 2\,\frac{\widehat{\sigma}_{12,n}}{\widehat{\mu}_{1,n}\widehat{\mu}_{2,n}}\right] \tag{4.9}$$

where $\widehat{\sigma}^2_{1,n}$ and $\widehat{\sigma}^2_{2,n}$ are the empirical variances of the $\widehat{Z}^\partial_i$ and Z_i, respectively, and $\widehat{\sigma}_{12,n}$ is the empirical covariance between $\widehat{Z}^\partial_i$ and Z_i. While they showed that this estimator is

consistent, i.e., $\widehat{\sigma}_\tau^2 \xrightarrow{p} \sigma_\tau^2$, they also showed that in the MC experiments, the estimated confidence intervals based on the CLT with this estimator of σ_τ^2 were significantly undercovering the true values (similarly as in KSW), even for samples like $n = 1000$.

Using similar logic as in the section above, an alternative estimator for σ_τ^2 can be suggested by correcting for the bias of $\widehat{\mu}_{1,n}$ in the estimator of the variance of the nonparametric estimators, i.e. $\widehat{\sigma}_{1,n}^2$ (note that this correction has no influence on $\widehat{\sigma}_{12,n}$). Following the same arguments as above we should rather use

$$\widehat{\widehat{\sigma}}_{1,n}^2 = \widehat{\sigma}_{1,n}^2 + \widehat{B}_{\mu_{1,n},\kappa}^2 \tag{4.10}$$

where $\widehat{B}_{\mu_{1,n},\kappa}^2$ is defined above as the numerator of (4.5). So we have

$$\widehat{\widehat{\sigma}}_{\tau,n}^2 = \widehat{\tau}_n^2 \left[\frac{\widehat{\widehat{\sigma}}_{1,n}^2}{\widehat{\mu}_{1,n}^2} + \frac{\widehat{\sigma}_{2,n}^2}{\widehat{\mu}_{2,n}^2} - 2\,\frac{\widehat{\sigma}_{12,n}}{\widehat{\mu}_{1,n}\widehat{\mu}_{2,n}} \right], \tag{4.11}$$

giving an alternative consistent estimator of the variance σ_τ^2 with larger values as expected.

The resulting estimated $(1-\alpha)$ confidence intervals for τ will therefore be

$$\left[\widehat{\tau}_n - \widehat{B}_{\tau_n,\kappa} \pm z_{1-\alpha/2} \widehat{\widehat{\sigma}}_{\tau,n} / \sqrt{n} \right], \tag{4.12}$$

when (4.6) can be applied, and

$$\left[\widehat{\tau}_{n_\kappa} - \widehat{B}_{\tau_n,\kappa} \pm z_{1-\alpha/2} \widehat{\widehat{\sigma}}_{\tau,n} / \sqrt{n_\kappa} \right] \tag{4.13}$$

when (4.7) can be applied.

Again, the proposed correction is of small order, vanishes asymptotically fast enough to preserve the asymptotic results from SZ, and might be significant in practice, especially for moderate sample sizes and large dimensions, as illustrated in our Monte-Carlo experiments below.

5 Monte-Carlo Evidence

In this section we summarize results from many Monte-Carlo (MC) experiments we conducted to analyze the performance of the newly proposed estimators for the variances described above, which aim to improve the approximation of the CLTs in finite samples.

As in the related works, to evaluate the performance we will estimate and present the coverage of confidence intervals for different approaches in various scenarios.[4] We focus on

[4]Recall that this coverage is defined as percentage of times an estimated confidence interval with selected

the usual choices for the nominal confidence levels: $0.90, 0.95, 0.99$ and for various sample sizes, for 1000 replications. Meanwhile, for the procedure of the bias estimation described above, we choose $L = 20$ reshuffles.[5]

For all the scenarios, the technology set is assumed to be characterized by

$$\Psi = \{(x, y) \in \mathbb{R}^p_+ \times \mathbb{R}^1_+ \mid y \leq \psi(x)\},$$

where for $\psi(x)$ we use Cobb-Douglas production function, i.e., $\psi(x) = y^\partial(x) = a_0 \Pi_{r=1}^p x_r^{\beta_r}$, which appears to be the most popular in empirical works and in simulations.[6] Following common practice and without much loss of generality, each input is generated from the standard uniform distribution, the inefficiency scores are generated as $\lambda_i \sim |N(0, \sigma_\lambda^2)| + 1$ for various choices of σ_λ^2.[7] The 'observed outputs' are then generated by projecting the frontier points $y^\partial(x)$ inside the technology set Ψ, i.e., $Y_i = \psi(X_i)/\lambda_i$. As in SZ, the true values of $\tau = \mu_1/\mu_2$ were computed via a prior simulation of $n = 2,000,000$ realizations of $(Y_i, Y_i^\partial = \psi(X_i))$ and then using (4.4). While we will limit our focus onto DEA-VRS estimator, which appears to be the most popular in practice, analogous conclusions apply to the DEA-CRS and FDH estimators.

In the tables below, note that the shortcut 'CLT1' refers to relevant part of CLT when $\kappa \geq 1/2$ (i.e., (2.17) for the sample mean and (4.6) for the aggregate efficiency) and 'CLT2' refers to relevant part of CLT when $\kappa < 1/2$ (i.e., (2.18) for the sample mean and (4.7) for the aggregate efficiency) and the asterisk indicates that the improved estimator for the variance was used when applying these theoretical results.

We first consider the case where $p = 1$, $q = 1$, i.e., where the bias is very small and, in fact, converges to zero fast enough so that even the standard CLT applies, although KSW and SZ showed that their approaches (based on CLT1, i.e., (2.17) for the sample mean and (4.6) for the aggregate efficiency) perform substantially better in the finite samples. Table 1 presents the results for such a case, when $\beta_1 = 0.4$ (other choices show similar results). As expected, because the bias is very small and converges to zero very fast, the improvement due to the proposed variance estimator is very small, yet persistent at virtually all sample sizes and nominal levels we considered, whether for the sample mean of for the aggregate efficiency.

Similar results are obtained for $p = 2$, $q = 1$, and Table 2 presents the results for such a

nominal confidence level covers the true value, out of all MC replications.

[5]Other choices of L we tried gave similar results.

[6]Because what matters the most from a theoretical perspective (e.g., for the rate of convergence) is the total number of inputs and outputs $(p + q)$, here we focus on a single-output case, i.e., $q = 1$, and consider different number of inputs, changing p.

[7]We tried many choices of σ_λ^2 and results were similar, with generally the same conclusion. The tables present results for $\sigma_\lambda^2 = 1$.

Table 1: Coverage for the Sample Mean and Aggregate Efficiency, for $p = 1$, $q = 1$

	CI level = 0.90		CI level = 0.95		CI level = 0.99	
	CLT1	CLT1*	CLT1	CLT1*	CLT1	CLT1*
	CLT for the Sample Mean					
10	0.500	0.557	0.573	0.635	0.685	0.736
20	0.602	0.653	0.676	0.711	0.795	0.824
50	0.691	0.717	0.775	0.790	0.892	0.898
100	0.771	0.784	0.834	0.842	0.918	0.921
200	0.826	0.830	0.880	0.883	0.967	0.968
300	0.846	0.846	0.912	0.913	0.960	0.962
500	0.847	0.848	0.907	0.907	0.979	0.979
1000	0.867	0.867	0.914	0.914	0.979	0.979
	CLT for the Aggregate Efficiency					
10	0.529	0.580	0.587	0.657	0.706	0.757
20	0.611	0.646	0.675	0.716	0.825	0.845
50	0.717	0.738	0.797	0.812	0.903	0.918
100	0.769	0.775	0.846	0.856	0.933	0.935
200	0.816	0.821	0.892	0.896	0.973	0.975
300	0.844	0.844	0.914	0.916	0.976	0.976
500	0.843	0.843	0.912	0.912	0.981	0.981
1000	0.881	0.882	0.922	0.923	0.973	0.973

case, where $\beta_1 = 0.4$, $\beta_2 = 0.2$. From this table, one can see that a slightly more pronounced improvement from the proposed variance estimator, as expected, since the estimation bias generally increases with increase in dimension, ceteris paribus, yet still not so large here, for 2-input-1-output case.

The next case we consider is when $p = 3$, $q = 1$ and so both CLT1 and CLT2 apply (in the case of DEA-VRS). Table 3 presents the results for $\beta_1 = 0.4$, $\beta_2 = 0.2$, $\beta_3 = 0.1$ (other choices show similar results). One can see that the improvement from using the proposed variance estimator is persistent and larger than what we observed for smaller dimensions, and both for CLT1 and CLT2. It is especially substantial at relatively small samples. As expected, the larger the sample the smaller the improvement (both in relative and in absolute terms), it is still observed even for samples like 1000, e.g., for nominal coverage set at 0.95, the original approach of KSW (when CLT2 is used) provides coverage of 0.918, while the improved version provides coverage of 0.934. Similar conclusion is also obtained when observing results for the aggregate efficiency. The MC results also suggest that all the estimated coverages, even with the improved variance estimator, are still underestimating the nominal levels, especially for relatively small samples, however recall that they are much better than the coverages based on the standard CLT (which were around 0 for these nominal levels and most samples).

Table 2: Coverage for the Sample Mean and Aggregate Efficiency, for $p = 2$, $q = 1$

	CI level = 0.90		CI level = 0.95		CI level = 0.99	
	CLT1	CLT1*	CLT1	CLT1*	CLT1	CLT1*
	CLT for the Sample Mean					
10	0.282	0.352	0.321	0.405	0.407	0.484
20	0.395	0.445	0.437	0.505	0.549	0.636
50	0.503	0.563	0.591	0.638	0.705	0.752
100	0.611	0.650	0.708	0.735	0.836	0.859
200	0.693	0.712	0.770	0.784	0.880	0.892
300	0.733	0.748	0.815	0.829	0.923	0.926
500	0.789	0.793	0.864	0.868	0.945	0.948
1000	0.825	0.826	0.887	0.893	0.979	0.980
	CLT for the Aggregate Efficiency					
10	0.260	0.348	0.320	0.409	0.434	0.524
20	0.431	0.490	0.486	0.556	0.610	0.681
50	0.567	0.617	0.653	0.703	0.785	0.830
100	0.679	0.712	0.769	0.797	0.878	0.896
200	0.730	0.742	0.804	0.819	0.915	0.924
300	0.766	0.775	0.848	0.859	0.939	0.946
500	0.807	0.816	0.876	0.886	0.957	0.957
1000	0.839	0.842	0.897	0.901	0.970	0.973

Next we consider the case when $p = 4$, $q = 1$ and so $\kappa < 1/2$, meaning that only CLT2 case applies here. Table 4 presents the results for $\beta_1 = 0.4$, $\beta_2 = 0.2$, $\beta_3 = 0.1$, $\beta_4 = 0.15$ (other choices show similar results). The same conclusions as reached in the previous table apply here as well, except that the improvement is even more pronounced (and applicable only for CLT2). Specifically, one can see from the Table 4 that the improvement from using the proposed variance estimator is also persistent here, more substantial (as expected, due to larger bias). It is especially improving the coverage at relatively small samples, and there is also an improvement even when $n = 1000$, where e.g., for 0.95 nominal coverage, the improved version provides coverage of 0.937, while the original approach of KSW provides coverage of 0.912.

Table 3: Coverage for the Sample Mean and Aggregate Efficiency, for $p = 3$, $q = 1$.

	CI level = 0.90				CI level = 0.95				CI level = 0.99			
	CLT1	CLT1*	CLT2	CLT2*	CLT1	CLT1*	CLT2	CLT2*	CLT1	CLT1*	CLT2	CLT2*
	CLT for the Sample Mean											
10	0.136	0.189	0.172	0.251	0.162	0.227	0.216	0.290	0.211	0.299	0.267	0.380
20	0.197	0.263	0.283	0.368	0.237	0.308	0.338	0.429	0.314	0.405	0.426	0.538
50	0.338	0.397	0.487	0.576	0.389	0.477	0.569	0.647	0.511	0.588	0.680	0.778
100	0.454	0.509	0.608	0.674	0.521	0.573	0.703	0.771	0.642	0.710	0.842	0.891
200	0.475	0.524	0.705	0.748	0.568	0.623	0.784	0.835	0.732	0.771	0.906	0.937
300	0.595	0.634	0.770	0.809	0.671	0.721	0.857	0.882	0.805	0.845	0.937	0.959
500	0.636	0.661	0.793	0.827	0.728	0.755	0.880	0.894	0.852	0.876	0.956	0.967
1000	0.762	0.776	0.859	0.870	0.830	0.841	0.918	0.934	0.934	0.945	0.975	0.983
	CLT for the Aggregate Efficiency											
10	0.146	0.212	0.191	0.289	0.172	0.249	0.243	0.323	0.230	0.330	0.305	0.417
20	0.210	0.283	0.310	0.407	0.251	0.346	0.365	0.474	0.342	0.450	0.481	0.609
50	0.413	0.497	0.567	0.656	0.486	0.576	0.654	0.736	0.615	0.703	0.778	0.847
100	0.534	0.593	0.695	0.752	0.612	0.674	0.781	0.842	0.750	0.799	0.896	0.933
200	0.604	0.644	0.766	0.804	0.684	0.722	0.844	0.877	0.821	0.862	0.942	0.963
300	0.667	0.703	0.815	0.846	0.756	0.789	0.876	0.892	0.876	0.898	0.961	0.967
500	0.709	0.732	0.842	0.861	0.785	0.813	0.902	0.917	0.907	0.927	0.967	0.971
1000	0.786	0.796	0.846	0.866	0.862	0.869	0.923	0.931	0.941	0.948	0.979	0.981

Table 4: Coverage for the Sample Mean and Aggregate Efficiency, for $p = 4$, $q = 1$.

	CI level = 0.90		CI level = 0.95		CI level = 0.99	
	CLT2	CLT2*	CLT2	CLT2*	CLT2	CLT2*
	CLT for the Sample Mean					
10	0.101	0.145	0.113	0.179	0.153	0.246
20	0.140	0.228	0.173	0.271	0.255	0.370
50	0.280	0.404	0.353	0.499	0.503	0.675
100	0.453	0.580	0.549	0.685	0.723	0.835
200	0.608	0.712	0.717	0.813	0.864	0.923
300	0.727	0.803	0.820	0.889	0.928	0.961
500	0.802	0.866	0.882	0.923	0.972	0.989
1000	0.849	0.872	0.912	0.937	0.976	0.984
	CLT for the Aggregate Efficiency					
10	0.108	0.172	0.129	0.206	0.180	0.276
20	0.174	0.271	0.212	0.334	0.302	0.443
50	0.386	0.524	0.475	0.597	0.619	0.760
100	0.562	0.684	0.662	0.785	0.834	0.914
200	0.725	0.806	0.820	0.889	0.926	0.964
300	0.811	0.867	0.888	0.929	0.957	0.979
500	0.856	0.891	0.921	0.944	0.983	0.992
1000	0.860	0.883	0.926	0.941	0.980	0.988

Similar conclusions are obtained for cases with larger dimensions, for the sake of space we only present two more cases: $p = 5$, $q = 1$, in Table 5 where $\beta_1 = 0.4$, $\beta_2 = 0.2$, $\beta_3 = 0.1$, $\beta_4 = 0.15$ and $p = 7$, $q = 1$ in Table 6, where $\beta_1 = 0.05$, $\beta_2 = 0.1$, $\beta_3 = 0.15$, $\beta_4 = 0.2$, $\beta_5 = 0.125$, $\beta_6 = 0.075$, $\beta_7 = 0.025$ (other values of parameters gave similar conclusions).

As expected, the improvement is much more substantial here, and almost everywhere. For example, for $p = 5$, $q = 1$, for the nominal coverage set at 0.95: for $n = 500$ the improved approach provides coverage of 0.92, while the original approach of KSW provides coverage of 0.846 and even for $n = 1000$, the original approach of KSW provides coverage of 0.92, while the improved approach provides coverage of 0.954, i.e., almost exactly recovering the nominal coverage.

Meanwhile, for $p = 7$, $q = 1$ and the nominal coverage is 0.95, one can see that for $n = 500$ the improved approach provides coverage of 0.941, while the original approach of KSW provides coverage of only 0.778, and when $n = 1000$ the improved approach provides coverage of 0.973 (i.e., slightly overshooting, by 0.023), while the original approach of KSW provides coverage of 0.895 (undershooting by 0.055). The results for the aggregate efficiency are similar, with a bit more overshooting observed.

Overall, by comparing all the tables (presented and those we omitted for the sake of space), an expected general conclusion is confirmed: in majority of cases, the larger the

Table 5: Coverage for the Sample Mean and Aggregate Efficiency, for $p = 5$, $q = 1$

	CI level = 0.90		CI level = 0.95		CI level = 0.99	
	CLT1	CLT1*	CLT1	CLT1*	CLT1	CLT1*
	CLT for the Sample Mean					
50	0.166	0.293	0.221	0.381	0.344	0.534
100	0.330	0.491	0.425	0.609	0.597	0.804
200	0.569	0.715	0.681	0.812	0.832	0.935
300	0.648	0.770	0.750	0.855	0.892	0.969
500	0.775	0.850	0.846	0.920	0.957	0.988
1000	0.855	0.905	0.920	0.954	0.976	0.991
	CLT for the Aggregate Efficiency					
50	0.235	0.385	0.304	0.488	0.452	0.664
100	0.457	0.612	0.563	0.726	0.730	0.883
200	0.689	0.797	0.781	0.880	0.901	0.966
300	0.750	0.841	0.836	0.914	0.948	0.990
500	0.835	0.898	0.904	0.954	0.979	0.992
1000	0.867	0.915	0.935	0.966	0.986	0.996

dimension the larger the improvement tends to be due to the new variance estimator we proposed in this paper.

6 Concluding Remarks

In this paper we propose a simple yet valuable improvement of the finite sample approximation of the recently derived central limit theorems for the statistics involving production efficiency estimates from DEA or FDH. This improvement is indeed very easy to implement because it amounts to a simple correction of the already developed statistics and involves no extra computational burden, while also preserves the existing asymptotic theory, including consistency and asymptotic normality.

The evidence form our Monte-Carlo experiments suggest that the proposed approach persistently gave improvements in most of the cases that we tried, especially for relatively small samples or relatively large dimensions of the assumed production model.

Given the importance of the central limit theorems in practical application of statistical testing and construction of confidence intervals, this approach is expected to be very useful (and at almost no additional computational costs) for practitioners analyzing production efficiency using DEA or FDH approaches.

Table 6: Coverage for the Sample Mean and Aggregate Efficiency, for $p = 7$, $q = 1$

	CI level = 0.90		CI level = 0.95		CI level = 0.99	
	CLT1	CLT1*	CLT1	CLT1*	CLT1	CLT1*
	CLT for the Sample Mean					
50	0.087	0.198	0.116	0.249	0.197	0.397
100	0.155	0.323	0.219	0.433	0.355	0.652
200	0.379	0.589	0.466	0.708	0.643	0.901
300	0.481	0.708	0.589	0.830	0.796	0.961
500	0.690	0.849	0.778	0.941	0.928	0.992
1000	0.801	0.923	0.895	0.973	0.976	1.000
	CLT for the Aggregate Efficiency					
50	0.109	0.226	0.141	0.299	0.220	0.461
100	0.219	0.391	0.276	0.523	0.419	0.741
200	0.429	0.682	0.547	0.793	0.733	0.952
300	0.573	0.801	0.685	0.890	0.860	0.983
500	0.750	0.903	0.840	0.966	0.960	0.996
1000	0.858	0.945	0.924	0.984	0.987	1.000

References

[1] Charnes, A., Cooper, W.W., and E. Rhodes (1978), Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2, 429–444.

[2] Chowdhury, H., Zelenyuk, V., Laporte, A. and W.P. Wodchis (2014) Analysis of productivity, efficiency and technological changes in hospital services in Ontario: How does case-mix matter? *International Journal of Production Economics*, 150, 74-82.

[3] Daraio, C., Simar, L. and P.W. Wilson (2018), Central Limit Theorems for Conditional Efficiency Measures and Tests of the "Separability" Condition in Nonparametric, Two-Stage Models of Production, Discussion paper 2016/27, ISBA, UCL, in press *The Econometrics Journal*, doi: 10.1111/ectj.12103

[4] Deprins, D., Simar, L. and H. Tulkens (1984), Measuring labor inefficiency in post offices. In *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243–267.

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

[6] Färe, R, Primont D (1995) *Multi-Output Production and Duality: Theory and Applications*, Boston: Kluwer Academic Publishers.

[7] Färe, R. and V. Zelenyuk (2003), On Aggregate Farrell Efficiencies, *European Journal of Operational Research* 146, 3, 615–621.

[8] Førsund, F. and N Sarafoglou (2005) The tale of two research communities: The diffusion of research on productive efficiency, *International Journal of Production Economics*, 98:1, 17-40.

[9] Gorman, M. and J. Ruggiero (2008) Evaluating US state police performance using data envelopment analysis, *International Journal of Production Economics*, 113:2, 1031-1037.

[10] Jeong, S.O. and L. Simar (2006), Linearly interpolated FDH efficiency score for nonconvex frontiers, *Journal of Multivariate Analysis*, 97, 2141–2161.

[11] Kneip, A, L. Simar and P.W. Wilson (2008), Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory*, 24, 1663–1697.

[12] Kneip, A., Simar, L. and P.W. Wilson (2011), A Computational Efficient, Consistent Bootstrap for Inference with Non-parametric DEA Estimators. *Computational Economics.* , 38,483–515.

[13] Kneip, A., Simar, L. and P.W. Wilson (2015), When bias kills the variance: Central Limit Theorems for DEA and FDH efficiency scores, *Econometric Theory*, 31, 394–422.

[14] Kneip, A., Simar, L. and P.W. Wilson (2016), Testing Hypothesis in Nonparametric Models of Production, *Journal of Business and Economic Statistics*, 34:3, 435–456.

[15] Kneip, A., Simar, L. and P.W. Wilson (2018), Inference in Dynamic, Nonparametric Models of Production: Central Limit Theorems for Malmquist Indices, Discussion paper 2018/10, ISBA, UCL.

[16] Olesen, O.B. (1995), Some unsolved problems in data envelopment analysis: A survey, *International Journal of Production Economics*, 39:1–2, 5-36.

[17] Simar, L. and P.W. Wilson (2011), Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36, 33–53.

[18] Simar, L. and P.W. Wilson (2018), Technical, Allocative and Overall Efficiency: Inference and Hypothesis Testing. Discussion paper 2018/18, ISBA, UCL.

[19] Simar, L. and V. Zelenyuk (2018), Central Limit Theorems for Aggregate Efficiency, *Operations Research*, 66:1, 137–149.